

Extreme Rainfall Events in the Northeastern United States Become More Frequent with Rising Temperatures, but Their Intensity Distribution Remains Stable

HELGA KRISTIN OLAFSDOTTIR,^a HOLGER ROOTZÉN,^a AND DAVID BOLIN^{a,b}

^a *Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden*

^b *Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

(Manuscript received 6 December 2020, in final form 9 August 2021)

ABSTRACT: Both the intensities of individual extreme rainfall events and the frequency of such events are important for infrastructure planning. We develop a new statistical extreme value model, the PGEV model, which makes it possible to use high-quality annual maximum series data instead of less well-checked daily data to estimate trends in intensity and frequency separately. The method is applied to annual maximum data from Vol. 10 of *NOAA Atlas 14*, dating from approximately 1900 to 2014, showing that in the majority of 333 rain gauge stations in the northeastern United States the frequency of extreme rainfall events increases as mean temperature increases, but that there is little evidence of trends in the distribution of the intensities of individual extreme rainfall events. The median of the frequency trends corresponds to extreme rainfall becoming 83% more frequent for each 1°C of temperature increase. Naturally, increasing trends in frequency also increase the yearly or decadal risks of very extreme rainfall events. Three other large areas in the contiguous United States, the Midwest, the Southeast, and Texas, are also studied, and show similar but weaker trends than those in the Northeast.

SIGNIFICANCE STATEMENT: Climate change can increase extreme daily rainfall by making individual extreme rainfall events more frequent, more intense, or both. We develop new methods based on extreme value statistics to predict to what extent these three scenarios will occur. Our aim is to inform infrastructure planning, both for protection against high-impact catastrophes and for local planning of roads and sewers. The new methods allow using high-quality annual maximum series for prediction of frequency, instead of partial duration series. In the northeastern United States, extreme daily rainfall events are becoming more frequent with little evidence of increasing trends in the distribution of intensities of individual extreme daily rainfall events. For many stations the frequency increase exceeds 150% for each 1°C of temperature increase.

KEYWORDS: Extreme events; Precipitation; Climate change; Climate prediction; Probabilistic Quantitative Precipitation Forecasting (PQPF); Trends

1. Introduction


Extreme rainfall in the contiguous United States has been estimated to have caused more than 0.75 trillion U.S. dollars (USD) in property losses since 1980 (Kunkel et al. 2020); the American Society of Civil Engineers (ASCE) estimates that trillions of dollars will be needed by 2025 to improve infrastructure in the United States (ASCE 2016), and climate change is expected to compound the problems (see, e.g., Kunkel et al. 2020). Rainfall events with immense intensity cause extensive damage, like the 2017 Hurricane Harvey, which produced daily rainfall amounts of more than 400 mm and is estimated to have caused more than 89 deaths and cost 130 billion USD (NOAA NCEI 2020). Extreme rainfall events with lower intensities also carry very high costs for making flooding of roads less frequent by improving detention ponds and building higher berms and bigger concrete culverts, and for improving drainage systems to make inundation of cellars less common.

Assuming stationarity has been standard practice when analyzing extreme hydrological events (Kundzewicz and Kaczmarek 2000). However, with increasing awareness of a changing climate, incorporating nonstationarity in the analysis, as we do in this paper, is becoming both more common and more important (see, e.g., Katz et al. 2002; Li et al. 2005; Milly et al. 2008; Towler et al. 2010; Kysely et al. 2010; Acero et al. 2011; Rootzén and Katz 2013; De Paola et al. 2018).

Climate change can increase extreme daily rainfalls in three different ways:

- C1: individual extreme rainfall events become more frequent (i.e., exceedances of high thresholds become more frequent);
- C2: individual extreme rainfall events become more intense (i.e., the conditional distribution of sizes of excesses over high thresholds becomes stochastically larger); or
- C3: individual extreme rainfall events become both more frequent and more intense.

Clearly, the maximum of (say) 100 rainfall events is likely to be larger than the maximum of 10 rainfall events following the same distribution. Because of this, C1 will make annual maxima larger and make the very extreme events like Harvey more likely, even if C2 does not happen. However, even if C1 and C2

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Helga Kristin Olafsdottir, khelga@chalmers.se

DOI: 10.1175/JCLI-D-20-0938.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

can increase risks of very extreme rainfall events in the same way, their effects on infrastructure needs still are different. If individual events with lower, but still extreme, intensity, such as flooding of roads and cellars become, say, twice as frequent, then demands for improvement will be much more urgent. In this paper we develop methods based on extreme value statistics to predict to what extent C1, C2, or C3 will occur. Our aim is to inform infrastructure planning, both for protection against high-impact catastrophes and for local planning of roads and sewers.

To answer statistical questions such as those above, the quality of data is crucial. Analysis of station data on extreme rainfall events typically uses either annual maximum series (AMS) of daily rainfall or partial duration series (PDS) data [also called peaks over thresholds (PoT) data] consisting of sizes of excesses of a high threshold. For the United States, daily rainfall data are available in the NOAA GHCN-Daily dataset (Menne et al. 2012). With the large number of station records, manual control of the data is not possible and instead quality assurance is mainly automated. AMS data on daily rainfall are provided by *NOAA Atlas 14* (Vol. 10; Perica et al. 2015, revised 2019). The AMS data are obtained from the GHCN-Daily data but have been submitted to additional careful manual screening and quality control, and are expected to be of substantially higher quality than the PDS data. As a (perhaps unnecessary) side remark, one can of course compute yearly maxima from the GHCN-Daily data. However, these data would be of the same quality as the daily data, and would take away the point of using the *NOAA Atlas 14* AMS data, namely, higher quality.

The method we develop and apply in this work makes it possible to use the high-quality AMS data to answer questions on the occurrence of cases C1–C3. Use of PDS data has the advantage that it provides direct information about both frequencies and intensities of extreme daily rainfall events, and hence about the extent of C1–C3. Furthermore, the PDS data cover more regions and have later ending dates, but instead have the drawback that it is difficult to judge to what extent results based on analysis of PDS data might be influenced by missing days and by errors in measurement and transcription.

Our method is built on the close relation between the two main models of extreme value statistics, the generalized extreme value (GEV) distribution for annual maxima and the generalized Pareto–Poisson model for partial duration series. It is related to the Langbein formula (Langbein 1949) commonly used in hydrology.

We apply the new method to show that C1 holds so that extreme daily rainfall events in the northeastern United States are becoming more frequent, but that C2 does not occur (i.e., the distribution of intensities of individual extreme rainfall events is not changing). As a consequence of C1, the intensity of, for example, the maximum yearly or decadal rainfall is also increasing. In the analysis we use the 99.5% threshold to define extreme daily rainfall events. However, it should be noted that according to our model trends for exceedances of higher thresholds will be the same, and that this agrees well with data.

The results are validated using a high-quality manually checked set of daily rainfall measurements from 16 weather

stations in the northeastern United States, and by parametric bootstrap simulations. We also perform the same analysis for three other large areas in the contiguous United States, and conclude that changes there are similar but slower than in the Northeast.

Our results are in line with earlier empirically based studies including *NOAA Atlas 14*, Vol. 10, where it was observed, but not confirmed, that annual maximal rainfall in the northeastern United States seem to increase (Perica et al. 2015, revised 2019). Easterling et al. (2017) and Barlow et al. (2019) summarize the large and quite diverse research on precipitation changes in the United States, which often concludes that heavy rainfall events in some or most parts of the United States have increased in intensity and frequency since 1901 and projects that the increase will continue over the twenty-first century. The later summary notes a very large variability in the definition of “extreme rainfall events” and the earlier one does not seem to provide a clear definition of the meaning of “increase in intensity.” Based on data and model output for Europe and the United States, Myhre et al. (2019) find strong increases in the frequency of extreme precipitation events, but that changes in intensity are weak. The climate literature shows that general circulation models (GCMs) have more difficulties in modeling precipitation than temperature and in capturing the often more local phenomena that lead to extreme precipitation amounts (Huang et al. 2017). It also points to important differences in extreme rainfall trends between the different regions in the United States, with the largest changes predicted to occur in the U.S. Northeast, and that increases have continued also after the end dates of the *NOAA Atlas 14* data (Kirchmeier-Young and Zhang 2020; Kunkel et al. 2020; Easterling et al. 2017; Risser et al. 2019).

In the regions we study, except for the U.S. Southeast, there are clear seasonal differences in the rate of occurrence of extreme rainfall events. However, for infrastructure planning the question of when during the year events happen often are of secondary importance, and we have not tried to include seasonality in our analysis.

The structure of this article is as follows. Section 2 describes the datasets on which the analysis is based. Section 3 gives an overview of the extreme value statistics methods used for the nonstationary analysis and derives the formula that connects the PDS and AMS models. Inference and model selection is discussed in section 4. The results of the analysis for the U.S. Northeast are presented in sections 5a, 5b, and 5c. The results for the other three areas in the contiguous United States are described in section 5d. Section 6 contains a concluding discussion. Additional technical details are provided in four appendixes.

2. Station and temperature data

We use three datasets of rainfall measurements from stations in the contiguous United States. The first dataset, the *NE-PDS* data, contains daily rainfall from 20 stations from the U.S. Northeast. It is a part of the data used to develop *NOAA Atlas 14*, Volume 10, and has gone through a more thorough quality control than is usual for daily data. These 20 stations were chosen since they all suggested increases in

TABLE 1. The number of stations from the regions in the *USA* data containing at least 60 years of yearly maxima and covering the year 2010, and the years of the first- and last-observed yearly maximum in the region.

	California	Midwest	Northeast	Ohio River basin	Semi-arid	Southeast	Texas
No. of stations	5	1122	333	0	0	433	423
First year	1850	1837	1816	—	—	1842	1849
Last year	2010	2012	2014	—	—	2012	2018

extreme rainfall events. Hence they are not suitable for inference about whether such events are becoming more common. Instead we use them to study if our inference from AMS data agrees with standard PoT modeling of daily data; to check if C1, C2, or C3 hold for this special dataset; and as an aid for threshold selection.

The second dataset, the *NE* dataset, is part of *NOAA Atlas 14* (Volume 10, part 3). It contains annual maxima of daily rainfall for 685 stations in the northeastern United States. The third dataset, the *USA* dataset, consists of the *NE* data together with the same data for six other large regions (Semi-arid, Ohio River basin, Midwest, Southeast, Texas, California) in the contiguous United States and is also obtained from the *NOAA Atlas 14* time series data ([NOAA's National Weather Service 2019](#)). The stations and regions are as defined by NOAA.

The time span of the observations depends on the dates of the *NOAA Atlas 14* updates and differs between stations and regions but ranges from approximately 1900 to 2014 (see [Table 1](#)). The station positions for the datasets *NE-PDS* and *NE* are displayed below (see [Figs. 4](#) and [6](#), respectively), and the areas in the *USA* dataset are shown in [Fig. 1](#). We only use stations containing at least 60 years of data covering a time period up until at least 2010. This results in 16 stations from

NE-PDS and 333 stations from *NE*. The number of stations from the regions in the *USA* data that satisfy the requirements are given in [Table 1](#). The requirements are not met by any station in the Ohio River basin nor in the Semi-arid region, and only by five stations in the California region. Thus these regions are excluded from the analysis.

The temperature data are obtained from NOAA ([NOAA NCEI 2019](#)). They contain the yearly average Northern Hemispheric temperature. To represent the changing climate, a lowess smoothing of this data is used, shown in [Fig. 2](#). The smoothing is computed with the standard parameters of the lowess function in R ([R Core Team 2017](#)).

3. Extreme value methods

The two basic models of extreme value statistics are the block maxima model and the PoT model (see, e.g., [Coles 2001](#)). For AMS, the blocks are years, and the model assumes that yearly maxima are independent and follow a GEV distribution, with cumulative distribution (CDF) function

$$G(x) = \exp \left\{ - \left[1 + \gamma \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\gamma} \right\}. \quad (1)$$

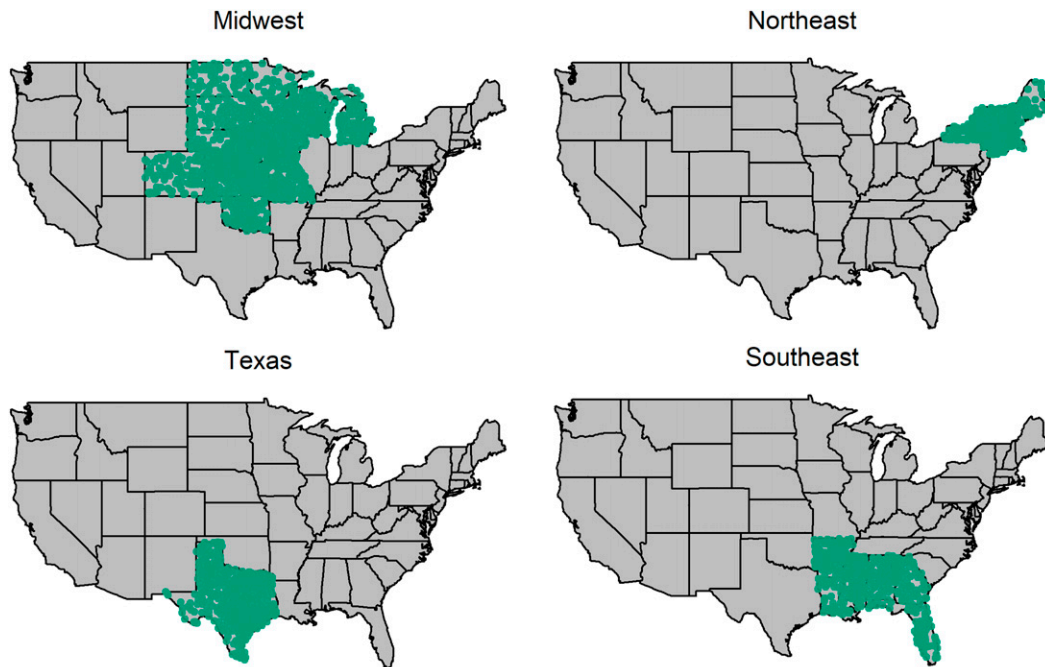


FIG. 1. Map of stations from the contiguous United States with division into regions.

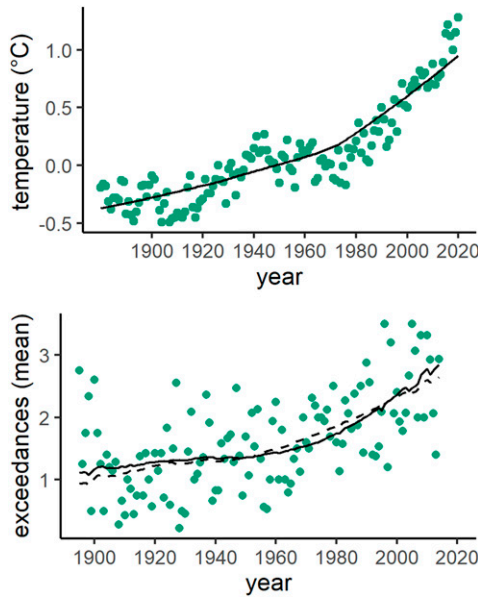


FIG. 2. (top) Yearly average temperature (°C) for the Northern Hemisphere along with loess smoothing. (bottom) Number of yearly exceedances over 99.5% threshold with regression curves $\exp(\lambda_1 t + \lambda_0)$, averaged over 16 stations from the NE-PDS dataset. Here, t represents either temperature (solid) or time (dashed).

Here μ is a real-valued location parameter, $\sigma > 0$ a scale parameter, and γ a real-valued shape parameter. For $\gamma = 0$, the distribution function simplifies to a double exponential distribution with location parameter μ and scale parameter σ .

The PoT model for partial duration series is that exceedances of a high threshold u occur as a Poisson process with rate parameter λ_u , and that sizes of excesses (observed rainfall event intensity values minus the threshold) are mutually independent, independent of the Poisson process, and follow a generalized Pareto distribution, with CDF

$$H(x) = 1 - \left(1 + \frac{\gamma x}{\sigma_u}\right)^{-1/\gamma}, \tag{2}$$

where $\sigma_u > 0$ is a scale parameter depending on the choice of the threshold u , and γ is a real-valued shape parameter. For $\gamma = 0$, the distribution simplifies to an exponential distribution with scale parameter σ . The assumption that excess sizes are mutually independent is often unrealistic. However, this can be handled in a standard way by declustering—that is, by identifying separate rainfall events and in the analysis only using the maximum daily rain amount in each rainfall event, as discussed later in section 4. These models are motivated by the basic probability theory of extremes, but as for any model, checking goodness of fit is important, and sometimes more complicated models are needed.

Recall that our goal is to answer questions about the cases C1–C3 in the introduction, which can be done using a PoT model for PDS data. It cannot be done using a standard GEV model since the (higher-quality) AMS data do not

provide direct information about frequencies of rainfall events. However, in the following section we construct a modified GEV model, the PGEV model, that will allow us to nevertheless use AMS data to study the extent of C1–C3. Our model is the extreme value statistics counterpart of the Langbein formula from hydrology.

a. Connection between the PDS and AMS models

Suppose the PoT model holds for excesses over a threshold u , with λ_u as the yearly rate of the Poisson process. Let N be the number of excesses in a year and let M be the annual maximum. Then, for $x > u$, and with H given by Eq. (2)

$$P(M < x) = \sum_{n=0}^{365} P(M < x | N = n) P(N = n) \approx \sum_{n=0}^{\infty} H(x - u)^n \frac{\lambda_u^n}{n!} e^{-\lambda_u}. \tag{3}$$

From straightforward computations (appendix A), it follows that M has a GEV distribution with location, scale, and shape parameters

$$\mu_{\text{PGEV}} = u + \frac{(\lambda_u^\gamma - 1)\sigma_u}{\gamma}, \quad \sigma_{\text{PGEV}} = \sigma_u \lambda_u^\gamma, \quad \gamma_{\text{PGEV}} = \gamma, \tag{4}$$

where λ_u , σ_u , and γ are the parameters of the PoT model, and where we write PGEV to indicate that the parameterization of the GEV model is obtained from the PoT model. If $\lambda_u = 1$, this is the Langbein formula (Langbein 1949; Takeuchi 1984) commonly used in hydrology, as further explained in appendix A.

b. Models for climate change

The standard method for incorporating a changing climate into the GEV model for the AMS series is to assume trends in the location and scale parameters as follows:

$$\mu(t) = \mu_0 + \mu_1 t, \quad \ln \sigma(t) = \sigma_0 + \sigma_{\text{GEV},1} t, \tag{5}$$

with t being a suitable covariate such as calendar time. We call this the GEV trend model.

The corresponding PoT trend model has the parameters

$$\ln \sigma_u(t) = \sigma_{u,0} + \sigma_{u,1} t, \quad \ln \lambda_u(t) = \lambda_{u,0} + \lambda_{u,1} t. \tag{6}$$

Here, $\lambda_{u,1}$ captures changes in the frequency of large rainstorms and $\sigma_{u,1}$ models changes in their intensities, and thus answers C1 and C2 of the introduction. Importantly, if the PoT model holds, for some threshold, say u_0 , it also holds for all thresholds $u > u_0$. Neither $\lambda_{u,1}$ nor $\sigma_{u,1}$ depends on the choice of the threshold $u > u_0$, see appendix B. The subscript u will hence be deleted and these parameters will be denoted σ_1 and λ_1 for the remainder of this paper. For both the AMS and the PoT models we assume that the shape parameter γ of the distributions is not affected by trends. This assumption is common when modeling nonstationary extreme precipitation (De Paola et al. 2018). For details about extreme value modeling, see Coles (2001).

To use AMS data to address cases C1–C3 of the introduction, we exploit Eq. (4) to transfer the parameters of the PoT trend model to GEV parameters to obtain the PGEV trend model for AMS data,

$$\mu_{\text{PGEV}}(t) = u + \frac{[\lambda_u(t)^\gamma - 1]\sigma_u(t)}{\gamma}, \quad \sigma_{\text{PGEV}}(t) = \sigma_u(t)\lambda_u(t)^\gamma, \tag{7}$$

with $\sigma_u(t)$ and $\lambda_u(t)$ given by Eq. (6). The shape parameter γ is the same in all three models. Trends in the frequencies of extreme rainfall events are caught by the parameter λ_1 and trends in their intensities by σ_1 . From Eq. (7) it is clear that an increase in the frequency of extreme rainfall events also will result in an increased location parameter in the GEV model.

Besides using calendar time as covariate, we also performed the analysis using a lowess smoothing of the yearly average Northern Hemispheric temperature (see Fig. 2). The results were similar for the two choices of covariates, so we only present the results obtained by using temperature as covariate, because (i) temperature directly affects the intensities and frequencies of rainfall events since the water-holding capacity of air increases with temperature, (Ganguli and Coulibaly 2017; Pendergrass 2018); (ii) general circulation models (GCMs) always include temperature change, and thus statistical models with temperature as covariate combined with a GCM can be used to predict changes in extreme rainfall events under the many different scenarios used in GCMs (Solomon et al. 2007); and (iii) temperature as covariate catches a broken trend around 1960 (see Fig. 2) and thus allows the use of simple models for the complete series of station data, while time as a covariate would require more complicated models or restricting analysis to measurements from 1960 and later. Using the Northern Hemispheric temperature and using the local temperatures for each region yielded similar results. However, trends in frequency of extreme rainfall events were stronger in the Southeast when using the Northern Hemispheric temperature. The results presented in this paper are from using the Northern Hemispheric temperature for all stations.

4. Inference and model selection

All model parameters were estimated with the maximum likelihood method. We used likelihood ratio tests with standard deviations estimated from observed information to test for the presence of trends. A reason for this is that maximum likelihood estimation makes it possible to accommodate many different kinds of observations schemes and models and in particular provides a general method to handle trends in parameters. We performed the estimation for the AMS and PoT models in R (R Core Team 2017), using the R package extRemes (Gilleland and Katz 2016).

a. Declustering

Before estimating the PoT model, one must make sure that the mutual independence assumption is reasonable. However, a single weather event can often cause more than

TABLE 2. Sign tests of differences Δ between estimates of λ_1 obtained from PGEV and PoT models for the declustered *NE-PDS* data. From threshold 0.99 and upward, we cannot reject the null hypothesis $\Delta = 0$.

Quantile	0.95	0.99	0.995	0.997
<i>p</i> value	0.004	0.077	0.210	0.454

one extreme daily rainfall event; more generally, in many situations threshold excesses come in small clusters with dependence within a cluster, but with independence between clusters. Declustering, as described in appendix C, is the standard method to remove dependence. The data from the 16 stations in *NE-PDS* indicated that rainfall events seldom were longer than 7 days. As a further check we redid the analyses below with cluster sizes 7–10. This led to quite similar results, and did not affect any of the conclusions. We hence used the declustered *NE-PDS* data with cluster size 7.

b. Threshold choice

A last step to take before obtaining a final estimate of the models is to choose the threshold, u . For PDS data, thresholds customarily are obtained as a compromise between bias and variance, with a higher threshold expected to give less bias but higher variance, with the compromise based on parameter stability and goodness-of-fit plots. Rain amounts vary widely between stations and the thresholds have to be obtained as a suitable individual quantile of the station observations. We used the *NE-PDS* data to inform the choice of quantile and then used the same quantile for all stations.

Parameter stability and goodness of fit plots showed good fit of the PoT model for the 99% and higher quantiles. Table 2 indicates that for the 99% and lower quantiles there may be a shift in the distribution of the λ estimates between the PGEV and the PoT models. We hence chose to use the threshold 99.5% throughout.

The PGEV model contains the threshold u , which cannot be directly estimated from the AMS data in the same way as from the PDS data. Instead, the value of the threshold u corresponding to the 99.5% quantile was obtained for the AMS data from Eq. (4) as

$$\hat{u} = \hat{\mu} - \frac{\hat{\sigma}(1 - 1.83^{-\hat{\gamma}})}{\hat{\gamma}}. \tag{8}$$

Here the parameters $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\gamma}$ were estimated from the stationary GEV model and the average number of yearly exceedances λ was set to 1.83 (corresponding to 0.5% of days within a year).

c. Model selection

To investigate if trends exist we made likelihood ratio tests (Coles 2001, 35–36) against the best submodel, and reported p values. Thus, for the model with trends in both λ and σ , we made the comparison with the model with a trend in only one of the parameters σ or λ that had the lowest p value when compared to the stationary model. Here, one

should note that if the trend parameter σ_1 equals 0 then by the first expression in Eq. (7) tests of the GEV model with $\mu_1 \neq 0$ and $\sigma_1 = 0$ against the model with no trends are likely to lead to rejection of the null hypothesis if λ_1 is different from 0. Similarly, even if the trend parameter σ_1 in fact equals 0 then by the second expression in Eq. (7) tests of the model with $\sigma_1 \neq 0$ and $\lambda_1 = 0$ against the model with no trends are, for the PGEV model, likely to lead to rejection of the null hypothesis if it is different from 0.

In our analysis, tests were performed for each station. Even without any trend effects, small p values are bound to occur for some of the stations and this should be accounted for. For example, if p values are computed for n stations for which the null hypothesis is true, then the expected number of stations for which the p value is less than 0.05 is $0.05n$, so that if the null hypothesis of no trend is true for all the $n = 333$ stations in the NE dataset then the expected number of p values that are smaller than 0.05 is 16.6. Standard methods for handling multiple significance include Boole's inequality and the Benjamini-Hochberg method. These methods are aimed at situations where one believes that there is a real effect in only one or a few of the tested cases and where the aim is to find those cases. For example, in gene research one often believes that only very few of the tested genes are linked to disease and the aim is to find those genes. However, for the problem we study here the situation is different. One expects that trends in extreme rainfall at the stations in an area all are driven by the same large-scale climate change processes and additionally are affected by local variability and noise. Thus one would expect that, apart from random variability, there either are no trends in the station measurements, or else that there are similar trends in many of the stations. To account for this, instead of using for instance Benjamini-Hochberg, we display results as plots of ordered p values. If there are no trends, the p -value curve should align around a 45° line through zero. A curve below the line suggests that climate change leads to trends at many stations and the farther below the 45° line the curves are, the stronger the evidence.

5. Results

In this section we present the results for the three models that can be used to detect trends in extreme rainfall events, the GEV and PoT trend models and the PGEV model. Recall that the GEV and PGEV models are fitted to high-quality AMS data, whereas the PoT model is fitted to PDS data. Further recall that only the PoT and the PGEV models can be used to study the occurrence of cases C1–C3 in the introduction. Naturally, a goal of this section is to compare the results of the different models to see, for example, if the use of the higher-quality AMS data and the PGEV model changes the conclusions that are drawn from the lower-quality PDS data and the PoT model.

Section 5a estimates trends in the declustered *NE-PDS* data and discusses the fit of the PGEV model. The PGEV model is then applied to estimate trends in extreme rainfall events in the U.S. Northeast (section 5b) and in three other large regions in

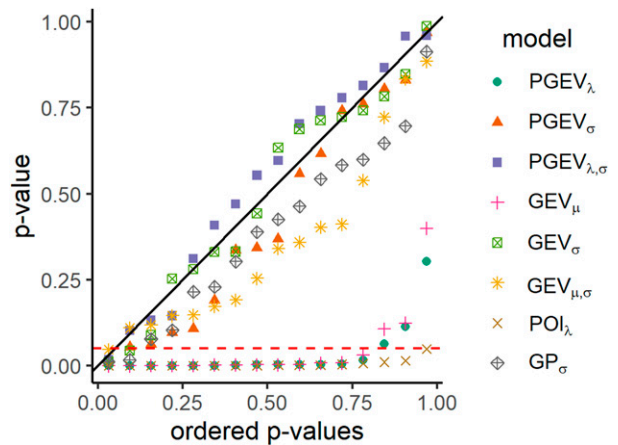


FIG. 3. Ordered p values for likelihood ratio tests for the declustered *NE-PDS* dataset and for the AMS for the same stations. The models represented are trend in Poisson intensity in the PGEV model for AMS (PGEV_{λ}), trend in scale in the PGEV model for AMS (PGEV_{σ}), trend in both Poisson intensity and scale in the PGEV model for AMS ($\text{PGEV}_{\lambda,\sigma}$), trend in location of AMS (GEV_{μ}), trend in scale of AMS (GEV_{σ}), trend in both location and scale in AMS ($\text{GEV}_{\mu,\sigma}$), trend in intensity in PoT (POI_{λ}), and trend in scale in PoT (GP_{σ}). Red dotted line is the 0.05 value.

the contiguous United States (section 5d). Section 5c explores the effect of trends on the frequency and intensity of future rainfall events.

a. The *NE-PDS* data

The *NE-PDS* stations were selected because they showed an increase in extreme rainfall events. The existence of trends in this special dataset, as expected, is confirmed by Fig. 3, which shows p values from eight different likelihood ratio tests. The tests of trends in both λ and σ were made against the best submodel, which was that with trend in λ but no trend in σ . No trend corresponds to p values falling close to the 45° line, while small p values indicate a trend. There is little indication of trends in σ in the GEV, PGEV, and PoT models. However, for both the PGEV and the PoT model, the plots point to clear trends in the rate of the Poisson model for the number of exceedances. Additionally, there is a trend in the location parameter of the GEV model, which, by Eq. (4), is as expected if there is a trend in the Poisson rate. Figure 4 shows that the λ_1 estimates obtained from the PoT trend model and the PGEV-trend model are similar. Hence, for this dataset individual extreme daily rainfall events are becoming more frequent but not more intense (i.e., C1 holds, but not C2). Further, this analysis confirms that the PGEV and the PoT models give similar results. But, of course, this dataset cannot be used to confirm or otherwise the existence of increasing trends in all of the Northeast.

b. The U.S. Northeast: Trends in Poisson rate, no trend in scale

By Fig. 5, the p values for tests of trends in both scale σ and Poisson rate λ in the PGEV model, against trend only in λ , give no

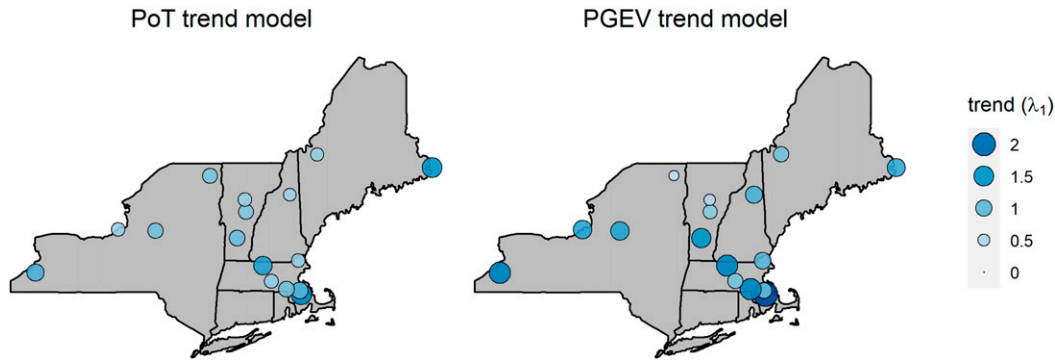


FIG. 4. (left) Estimated λ_1 from the PoT trend model for the *NE-PDS* data and (right) the $PGEV_\lambda$ trend model for the AMS data from the same stations, with less smoothed temperature for the Northern Hemisphere as covariate.

reason to believe that including a trend in σ improves model fit. Instead, the p values for the tests of trend in intensity against no trend indicate that there are trends in the Poisson rate for many stations. In particular, 34% of the p values were smaller than 0.05. As for the *NE-PDS* data, the curve for the p values of the PGEV model test of trend only in scale against no trend is as expected from Eq. (4). To simplify Fig. 5, we have not included the p values from the GEV tests that are part of Fig. 3. However, as in Fig. 3 the p -value plot for the tests of trends in location in the GEV trend model was similar to the plot for the tests of trend in intensity in the PGEV model, and the other test gave no indication of significance. The validity of these conclusions is supported by simulation in appendix D.

The estimated trends for the PGEV model with trend in intensity are shown in Fig. 6. Visually, the figure seems to indicate a tendency for trends in intensity to be larger in the southeastern part of the area. However, this may just be a result of the denser net of stations in this area. The corresponding plot of estimated location parameters in the GEV model (not shown) showed that similar trends are caught station-wise by both models, but less clearly by the GEV model.

The histogram in Fig. 6 shows that for 91% of the stations the estimated trends λ_1 in the frequency of extreme rainfall events are positive. The mean trend is 0.65, the median trend is 0.61, and the standard deviation of the trends is 0.49. The time periods that were included in the station data were related to the estimates of the size of the trend in λ_1 , and, in particular, stations with later ending times often had higher λ_1 values.

c. The U.S. Northeast: Implications of climate change on the frequency and intensity of large rainfall events

Let $\Delta_t = t_2 - t_1$ be the temperature change, where t_1 and t_2 are the average yearly temperatures at two distinct times. At temperature t the expected number of exceedances is given by $\lambda(t) = \exp(\lambda_0 + \lambda_1 t)$, and thus the relative difference in the number of yearly extreme rainfall events is

$$\Delta_\lambda = \frac{\lambda(t_2) - \lambda(t_1)}{\lambda(t_1)} = \exp(\lambda_1 \Delta_t) - 1.$$

The left panel of Fig. 7 shows a histogram of Δ_λ for the stations in the *NE-PDS* data for different values of Δ_t . For $\Delta_t = 0.5^\circ\text{C}$, the median increase is around 35%, and for $\Delta_t = 1^\circ\text{C}$, the median increase is around 83% and the mean increase is around 116%.

As discussed in the introduction, even though the distribution of individual extreme rainstorms does not change, the increase in frequency will lead to an increase of intensity of the yearly or, say, 20-yearly maximum rainstorm. A common way of reporting sizes of extreme storms is through the intensity of the 20-yr storm, which in a stationary climate is the storm that has probability 5% of occurring in a year. In changing climate, static references such as 20-yr storms

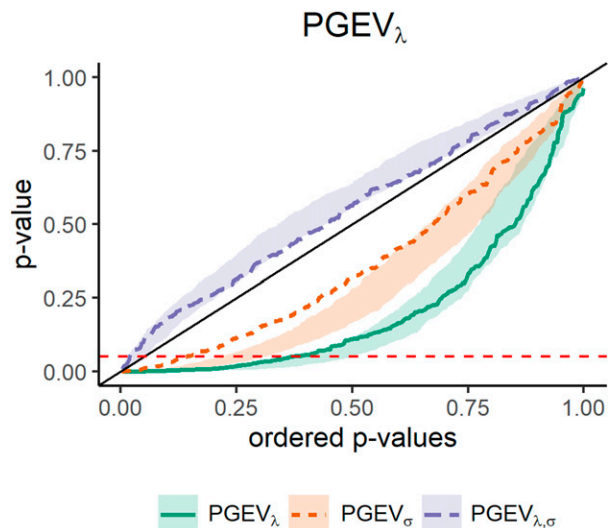


FIG. 5. Ordered p values for likelihood ratio tests for the AMS in the 333 stations in the *NE* dataset. The models represented are trend in Poisson intensity in the PGEV model for AMS ($PGEV_\lambda$), trend in scale in the PGEV model for AMS ($PGEV_\sigma$), and trend in both Poisson intensity and scale in the PGEV model for AMS ($PGEV_{\lambda,\sigma}$). Red dashed line is the 0.05 value. Colored regions are confidence bands obtained by simulation (see appendix D).

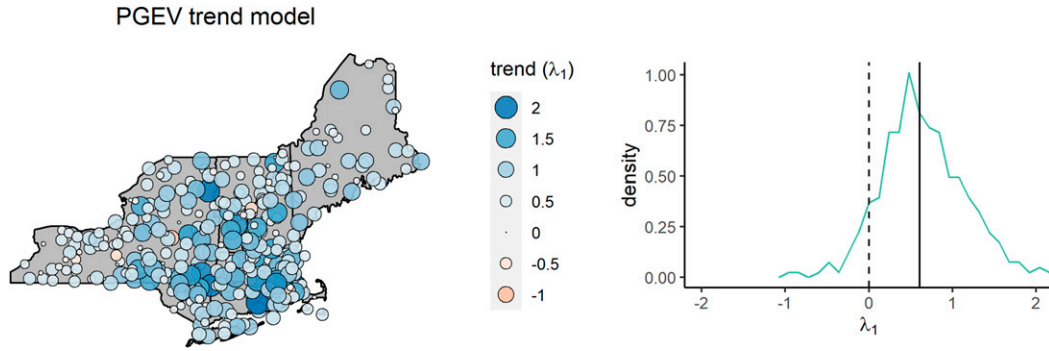


FIG. 6. Trends in frequency of extreme rainfalls are described by estimated parameters λ_1 for the PGEV $_{\lambda}$ model for the NE AMS data. Sizes of trends are shown (left) by both color and size of circles on map and (right) by the histogram of the λ_1 estimates with median as a solid vertical line.

lose meaning. Therefore, we instead of the 20-yr rainstorm talk about the 5% rainstorm, and more generally about the $P\%$ rainstorm. This is the storm intensity that has the probability $P\%$ of being exceeded in a year, obtained from the PGEV model as

$$S(p, t) = u + \frac{[\lambda(t)^\gamma - 1]\sigma}{\gamma} + \frac{\lambda(t)^\gamma \sigma}{\gamma} \{[-\log(1 - p)]^{-\gamma} - 1\}.$$

A similar result holds for $\gamma = 0$. The probability of a storm of intensity $S(p, t_1)$ at temperature t_2 is

$$P[X(t_2) > S(p, t_1)] = 1 - \exp \left[- \left(1 + \gamma \frac{S(p, t_1) - \left\{ u + \frac{[\lambda(t_2)^\gamma - 1]\sigma}{\gamma} \right\}}{\lambda(t_2)^\gamma \sigma} \right)^{-1/\gamma} \right]$$

$$= 1 - \exp \left[\frac{\lambda(t_2)}{\lambda(t_1)} \log(1 - p) \right] = 1 - (1 - p)^{\exp(\lambda_1 \Delta)}.$$

This means that if the relative frequency of exceeding a threshold increases by 83%, as seen for a 1°C increase, then a 5% storm will occur with approximately 9% probability, under the new temperature.

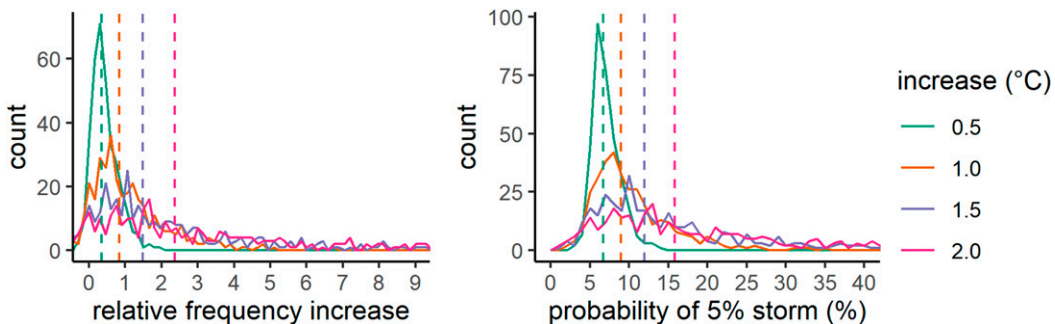


FIG. 7. Histograms of (left) the expected relative increase in the number of extreme rainfall events and (right) the future probability of a 5% rainstorm for the 333 stations in the U.S. Northeast as obtained by assuming different increases of yearly mean temperature. Medians are included in both plots (vertical dashed lines). For readability, the x axis has been truncated in both plots, excluding 14 values for the 1.5°C increase and 55 values for the 2°C increase.

d. The contiguous United States: Similar but weaker trends

The four regions in the USA dataset are shown in Fig. 1. The p values obtained from the PGEV model for the other three regions in the contiguous United States were similar to those for the U.S. Northeast, and indicated no or little trend in the scale parameter in the PGEV model but trends in the intensity parameter λ_1 (Fig. 8). Positive estimated λ_1 values dominate but are smaller than for the Northeast (Fig. 9). The scale parameters were highest in Texas and the Southeast (Fig. 10), as expected. The shape parameters are similar throughout the contiguous United States, except for Texas (Fig. 10) where they are somewhat larger, and hence the risks of unusually extreme rainfall events are higher.

6. Discussion and conclusions

In this paper we use a new model, the PGEV model, to show that in the U.S. Northeast extreme daily rainfall events are becoming more frequent but that there is little evidence of increasing trends in the distribution of intensities of individual extreme daily rainfall events. The estimated median trend in frequency for stations in the U.S. Northeast corresponds to exceedances of high thresholds becoming 83% more frequent

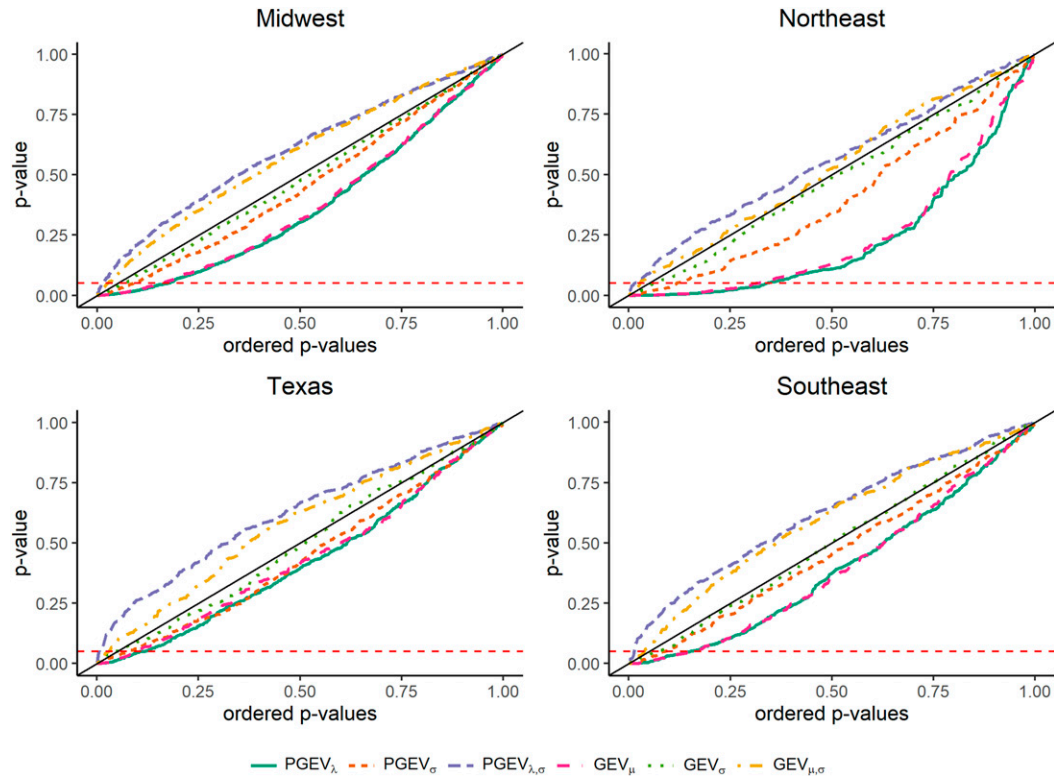


FIG. 8. Ordered p values for the different models applied to AMS from the stations in the contiguous United States, by region. The models represented are trend in Poisson intensity in the PGEV model for AMS ($PGEV_{\lambda}$), trend in scale in the PGEV model for AMS ($PGEV_{\sigma}$), trend in both Poisson intensity and scale in the PGEV model for AMS ($PGEV_{\lambda,\sigma}$), trend in location of AMS (GEV_{μ}), trend in scale of AMS (GEV_{σ}), and trend in both location and scale in AMS ($GEV_{\mu,\sigma}$). The red dotted line shows the 0.05 value.

per degree of increase in average yearly temperature, and for many stations the frequency increase exceeds 150% per 1°C of temperature increase. Further, for example, if the temperature increases by 1°C , then the median increase of the probability of the 5% rainstorm is to 9%, so the 5% probability rainstorms become almost twice as likely. As a consequence, very extreme yearly maxima of daily rainfall event intensities also become more probable. Trends, as estimated for three other large areas in the contiguous United States (Midwest, Southeast, Texas) are similar, but weaker. The results used high-quality annual maximum series (AMS) data, and were validated using simulations and PDS data from 16 stations. These data had undergone manual quality control and are of higher quality than standard PDS data.

The PGEV model is based on using the connection between the block maxima and PoT method in extreme value statistics to derive a new parameterization that is similar to the Langbein formula. An important property of our model is that it can be used to predict the implications of different GCM scenarios on risks of extreme rainfall amounts. A practical way for communities and cities to use our results for infrastructure planning could be to find one or several nearby stations with similar topography and estimate PGEV parameters, in particular trends, at those, and then use the result for informing dimensioning of infrastructure.

A further important consideration for planning is “storm dependence”: If an extreme rainstorm only hits a small area (and thus perhaps only a single station) consequences could be less severe than if it hits a large area (and then perhaps several stations). The extreme value statistical methodology for modeling storm dependence (and dependence between station measurements) is still at an early stage (see, e.g., Kiriliouk et al. 2019). Storm dependence will not affect biases or standard deviations of trend estimates for individual station or biases of averages of such estimates, but it will influence standard deviations of means of station trends.

In appendix E, we provide a spatial Matérn model for the storm dependence and use it to estimate a common trend for each region. The result of this analysis is that there is little evidence for differences in trends between stations in the same region, but that there are differences in the trends for the different regions. For more details and estimated values, see appendix E.

Many of our results, in particular that trends are strongest in the U.S. Northeast, agree with earlier literature. The discussion in Easterling et al. (2017) about changes in extratropical cyclones and how they cause a large portion of extreme rainfall events in the Northeast may give one explanation of why our models fit well above the 99% quantile, but not below it. However, there are also differences. For example, Kunkel et al. (2020) finds that trends for rare events

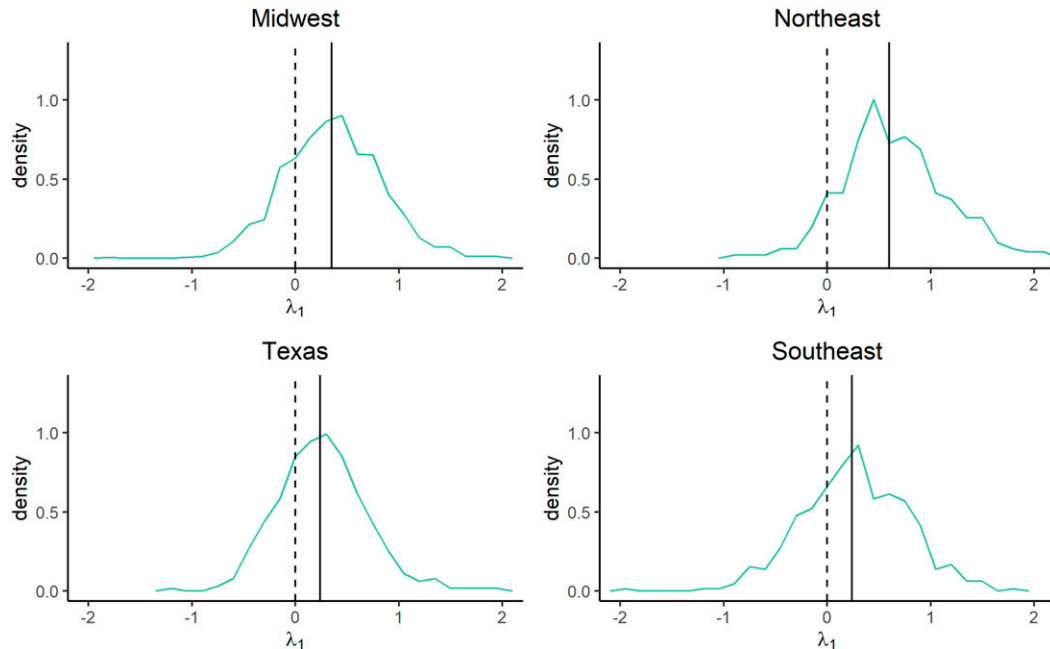


FIG. 9. Histograms of the estimated trend parameters λ_1 for the $PGEV_\lambda$ model for the *USA* dataset describing the trends in frequency of extreme rainfall, shown by region.. The solid vertical line in each panel indicates the median.

are higher than for less rare events, in contradiction to our results. This paper uses PDS data for the period 1949–2016 whereas the *NOAA Atlas 14* AMS data for the Northeast that are analyzed here cover the time period 1814–2014, so different ending times presumably do not explain this difference in conclusions. Instead possible explanations include that Kunkel et al. (2020) compute grid cell averages with different cell sizes for more common and more rare events before using standard linear regression to estimate trends; that having temperature as a covariate makes it possible for us to use much earlier starting times; and that there are missing observations in the GHCN dataset.

Empirical climate precipitation research usually relies on translation of station data to grid cells by various forms of interpolation and averaging, and the climate models also

provide gridded output. The extent to which gridded data agree with station data depends on grid size and the details of how the averaging implicit in gridding has been done (see, e.g., Huang et al. 2017; Risser et al. 2019). Compared to a local analysis, gridding has the advantage of reducing variances and therefore adding robustness to the conclusions. However, there often are important local differences inside of the grid cells (see, e.g., Martel et al. 2020), and this may make gridded results less directly useful for the engineers who build our infrastructures.

Our results apply to the maximum of the daily rain amounts in separate storm events and mean that extreme rain storms are becoming more frequent for many stations. However, a rainstorm may contain several days with high amounts of rain. If this is important, it should be studied

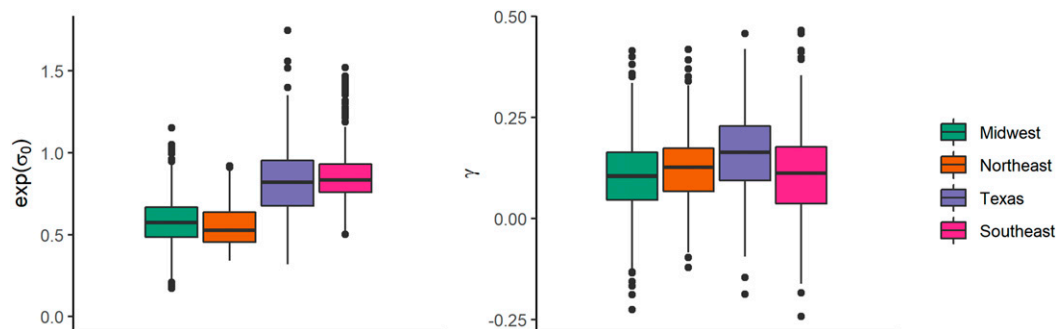


FIG. 10. Boxplots of estimated values of the (left) scale parameter $\exp\{\sigma_0\}$ and (right) shape parameter γ for stations in the contiguous United States.

separately, for individual stations. Some topics for future research are including spatial and topographic information into the models, and taking storm dependency between stations into account in the modeling.

Acknowledgments. This work is funded by the Knut and Alice Wallenberg Foundation (KAW Grant 20012.0067) and the Swedish Research Council (Grant 2016-04187). Special thanks to Sandra Pavlovic, Ben Shaby, and Gregory Bopp for very useful discussions, and to Philippe Naveau for his suggestions. We also thank three anonymous reviewers for comments that led to important improvements of the paper.

Data availability statement. The data used in this article are publicly available and are obtained from NOAA (NOAA NCEI (2019) and NOAA’s National Weather Service (2019).

APPENDIX A

Connection between Estimated Trends in BM and PoT and the Langbein Formula

Let X denote the rainfall intensity at a randomly chosen day, M the yearly maxima, and u the threshold. If threshold excesses $X - u|X > u$ have a GP(σ_u, γ) distribution, and the number N of exceedances over threshold for a given year has a Poisson distribution with rate parameter λ_u , then

$$P(N = n) = \frac{\lambda_u^n}{n!} e^{-\lambda_u},$$

and by Eq. (3),

$$\begin{aligned} P(M < x) &\approx \sum_{n=0}^{\infty} \left[1 - \left(1 + \gamma \frac{x-u}{\sigma_u} \right)^{-1/\gamma} \right]^n \frac{\lambda_u^n}{n!} e^{-\lambda_u} \\ &= e^{-\lambda_u} \exp \left\{ \lambda_u \left[1 - \left(1 + \gamma \frac{x-u}{\sigma_u} \right)^{-1/\gamma} \right] \right\} \\ &= \exp \left(- \left\{ 1 + \gamma \frac{x - \left[u + \frac{\sigma_u (\lambda_u^\gamma - 1)}{\gamma} \right]}{\sigma_u \lambda_u^\gamma} \right\}^{-1/\gamma} \right). \end{aligned}$$

Comparing the above to the GEV distribution, we see that $M \sim \text{GEV}(\tilde{\mu}, \tilde{\sigma}, \tilde{\gamma})$, where

$$\tilde{\mu} = u + \frac{(\lambda_u^\gamma - 1)\sigma_u}{\gamma}, \quad \tilde{\sigma} = \sigma_u \lambda_u^\gamma, \quad \text{and} \quad \tilde{\gamma} = \gamma \tag{A1}$$

when $\gamma \neq 0$. When $\gamma = 0$, the parameters are connected as

$$\tilde{\mu} = u + \sigma \log(\lambda_u), \quad \tilde{\sigma} = \sigma, \quad \text{and} \quad \tilde{\gamma} = \gamma = 0. \tag{A2}$$

Langbein (1949) derived the relation

$$T_a = 1/[1 - \exp(-1/T_p)] \tag{A3}$$

between the recurrence intervals T_a and T_p of the same event for the annual maximum and peaks over threshold

(PoT) series under the assumption of stationarity (see also Takeuchi 1984). Assuming the GEV distribution for the annual maximum series and the GP distribution for the PoT series, we have that the probability of the annual maximum M exceeding a value x is

$$P(M \geq x) = 1 - \exp \left\{ - \left[1 + \tilde{\gamma} \left(\frac{x - \tilde{\mu}}{\tilde{\sigma}} \right) \right]^{-1/\tilde{\gamma}} \right\},$$

and the distribution of the peaks X over threshold u is

$$P(X \geq x|X > u) = \left[1 + \gamma \left(\frac{x-u}{\sigma} \right) \right]^{-1/\gamma}.$$

This gives the corresponding recurrence intervals

$$T_a = 1 / \left(1 - \exp \left\{ - \left[1 + \tilde{\gamma} \left(\frac{x - \tilde{\mu}}{\tilde{\sigma}} \right) \right]^{-1/\tilde{\gamma}} \right\} \right), \tag{A4}$$

and

$$T_p = \left[1 + \gamma \left(\frac{x-u}{\sigma} \right) \right]^{1/\gamma}. \tag{A5}$$

Inserting Eq. (A4) into Langbein’s formula (A3) gives the recurrence interval

$$T_p = \left[1 + \tilde{\gamma} \left(\frac{x - \tilde{\mu}}{\tilde{\sigma}} \right) \right]^{1/\tilde{\gamma}}. \tag{A6}$$

Comparing Eqs. (A5) and (A6), shows that Langbein’s formula holds if $\tilde{\mu} = u$, $\tilde{\sigma} = \sigma$, and $\tilde{\gamma} = \gamma$. From Eq. (A1), we can see that this means that λ_u^γ should be equal to one. If $\gamma \neq 0$, this holds when $\lambda_u = 1$ and approximately when $\gamma \ln \lambda_u \approx 0$. In the Gumbel case, when $\gamma = 0$, the connection between the models becomes $\mu = u + \sigma_u \ln(\lambda_u)$ and $\sigma = \sigma_u$, which also is close to the Langbein formula when $\lambda_u \approx 1$.

APPENDIX B

Trends Do Not Depend on the Threshold

When using the PoT method to estimate the parameters of the GP for the distribution of excesses and the Poisson process describing the number of exceedances over a threshold, the threshold u needs to be chosen. Ideally, it is chosen high enough that the GP distribution gives a good fit to the excesses, but as low as possible to get as much data to work with as possible. Both the scale parameter σ_u and the rate parameter λ_u depend on the choice of the threshold. If the threshold is increased by v , the rate parameter decreases to

$$\lambda_{u+v} = P(X > u + v|X > u)\lambda_u. \tag{B1}$$

Therefore, $\lambda_u = \exp(\lambda_0 + \lambda_1 t)$ gives $\lambda_{u+v} = \exp[\log(k) + \lambda_0 + \lambda_1 t]$, where $k = P(X > u + v|X > u)$. Thus, in the PoT model, changing the threshold does not change the trend parameter λ_1 . However, of course the estimated λ_1 is different if one uses different thresholds for the estimation since changing the

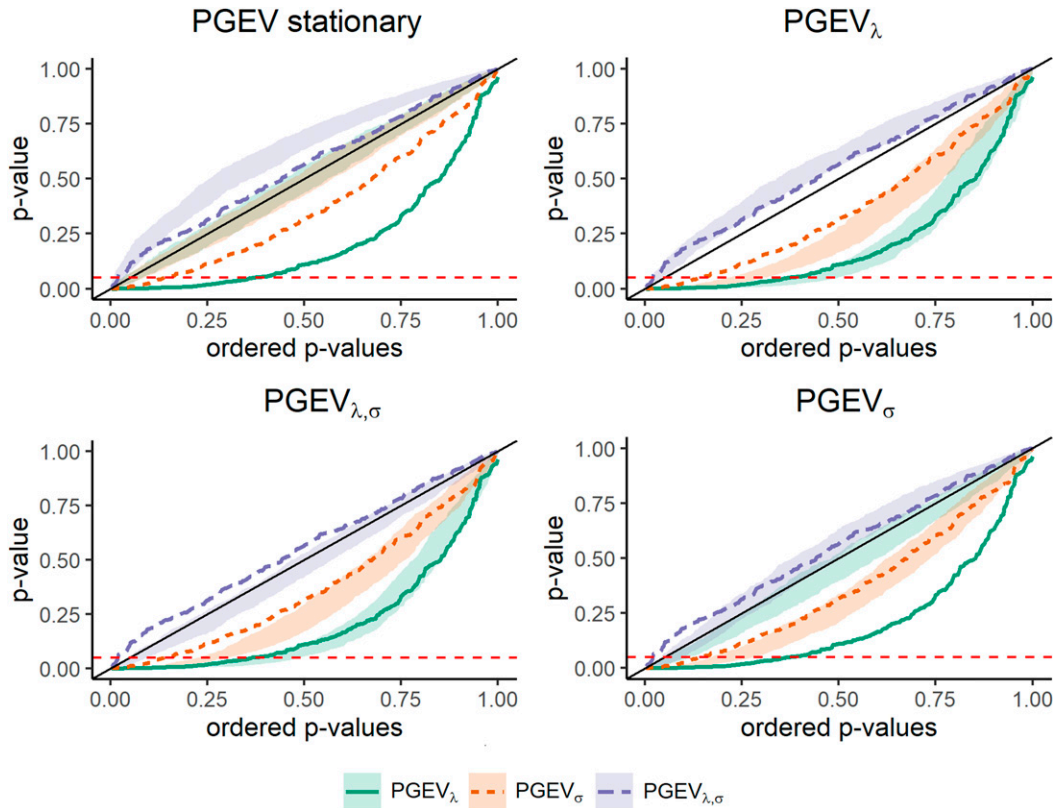


FIG. D1. Ordered p values from model fits on the *NE* dataset (lines) along with confidence bands from simulation (shaded regions) using four different PGEV models. Red dotted line is the 0.05 value.

threshold changes what parts of the data are used for estimation.

APPENDIX C

Declustering

The standard technique to handle the dependency of threshold exceeding is to decluster—that is, to gather the observed excesses into clusters and use the cluster maxima as the basic observations in the PoT model and in the relation (4) (see Coles 2001). To find the clusters we first, guided by data as described below, determine a cluster length δ and let the first cluster start at the time, t_1 , of the first exceedance of u . The observed rain amounts $x_{t_1}, x_{t_1+1}, \dots, x_{t_1+\delta-1}$ on these days are then modified as follows:

$$x'_t = \begin{cases} x_t, & \text{if } x_t = \max\{x_j; j \in (t_1, \dots, t_1 + \delta - 1)\} \\ \min(x_t, u), & \text{otherwise,} \end{cases} \tag{C1}$$

so that only the maximum value in the cluster remains an exceedance. In case the maximum occurs multiple times within a cluster, the maximum is assigned to the first such occurrence. The second cluster starts at the time, t_2 , of the first exceedance of u after time t_1 , and the rain amounts in the

cluster are modified in the same way, and so on. We thus obtain a time series of cluster maxima, and use it as the observations instead of the original data when estimating the PoT model.

APPENDIX D

Verification via Simulation

As a background to the p plots (Figs. 3 and 8), daily rainfall for 300 stations was simulated using parameters similar to the estimated PGEV parameters in the U.S. Northeast. Thereafter, model parameters were estimated from the simulated data and the ordered p values for each model were computed. This was repeated for 100 populations and from those, 95% confidence bounds were computed and compared to the results obtained from the *NE* dataset.

For each station, parameters $\theta = (\lambda_0, \lambda_1, \sigma_0, \sigma_1, \gamma)$ were drawn from an $N(\hat{\theta}, \hat{\Sigma})$ distribution, independently for the different stations. Here, $\hat{\theta}$ and $\hat{\Sigma}$ were first obtained by fitting a multivariate normal distribution to the estimated PGEV model parameters from the stations in the *NE* dataset. Then for each station 80 yearly maxima were simulated from a GEV distribution with the parameterization of Eq. (7), with the covariate t set to the lowest temperature for years 1934 to 2014.

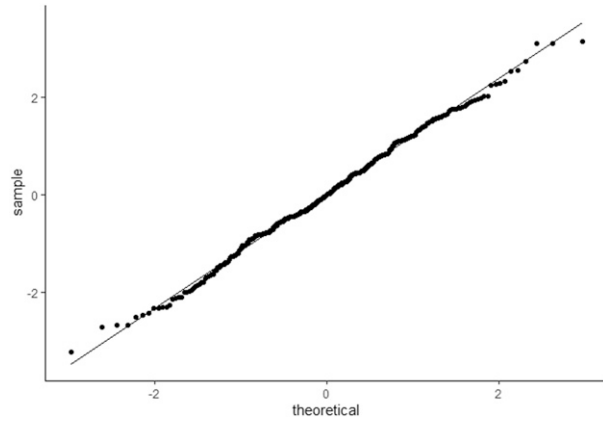


FIG. E1. A Q–Q plot of standardized station-wise estimates of the trend parameter λ_1 in the Northeast against standard normal distribution.

The threshold u was set to 2.88. Figure D1, where the resulting confidence bounds are compared to the p plots from the NE dataset, indicates that the PGEV $_{\lambda}$ model describes the data reasonably well. In this it should be noted that there are differences between the simulated data and the NE data, such as that the threshold was chosen to be constant for all stations in simulations and that there is spatial dependency that is not modeled in the simulated data.

APPENDIX E

Spatial Analysis

For each station we obtained an estimated trend $\lambda_{1,i}$ together with an associated estimated standard deviation σ_i of the estimate via the maximum likelihood estimation method in section 4. The variation between estimated trends $\lambda_{1,i}$ at different stations depends on both estimation uncertainty and on differences between real trends. If there are no differences in real trends, then standardizing the station-wise trend estimates by subtracting the mean trend and dividing by the estimated standard deviations should give standard normal values. Instead, if there are substantial variations in the true trends these values would be expected to deviate from a

normal distribution. Figure E1 shows that for the Northeast a standard normal distribution gives a good fit and hence that there is little indication of variation between the true trends in the different stations in this region. Similar plots for the other three regions gave the same result. However, there are clear differences in the trends between the different regions.

We hence for each region fitted the simple spatial model

$$Y_i = \mu + X(s_i) + e_i, \tag{E1}$$

to the estimated station-wise trends. In this model μ is a common mean trend and X is a mean-zero Gaussian random field (GRF) with Matérn covariance function

$$C(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa h)^{\nu} K_{\nu}(\kappa h).$$

Here Γ is the gamma function; K_{ν} is the modified Bessel function of the second kind; ν , κ , and σ^2 are positive shape, scale, and variance parameters, respectively; and h is geodesic distance measured in tens of kilometers. The terms $e_i \sim N(0, r^2\sigma_i^2)$ are independent nugget effects where the station-wise nugget standard deviations, $r\sigma_i$, are assumed to be proportional to the estimated standard deviations σ_i of the trend estimates, and r is the proportionality parameter.

The parameters of the model, $\theta = (\mu, \nu, \kappa, \sigma, r)$, were estimated by maximizing the log-likelihood of model (E1). The practical dependency range (i.e., the distance where the correlation is approximately 0.1) is given by the estimate of $\rho = 10\sqrt{8\nu\kappa}^{-1}$ (Lindgren et al. 2011). The results for the four regions are shown in Table E1. Extreme rainfall events in the U.S. Northeast become more frequent with rising temperatures, but their intensity distribution remains stable. The effective range of the GRF differs slightly between regions, but is between 144 and 218 km (Table E1).

The average of the estimated station-wise trend parameters are contained in the confidence interval for the common trend μ for all regions, again suggesting that there is little variation between the true trends inside of the different regions. Further, within each of the four regions we computed estimates of the contour avoiding set, $E_{u=0,\alpha}(X)$. This is defined as the union of the joint contour $u = 0$ excursion sets with probability $1 - \alpha$; see definition 4 in Bolin and Lindgren (2015). The probability is

TABLE E1. Parameter estimates for each of the four regions by using individual nugget effects with common scale r along with mean trend λ_1 , confidence intervals for μ and σ_1 , and approximated practical range ρ .

	Northeast	Southeast	Midwest	Texas
$\overline{\lambda_1}$	0.6430	0.2571	0.3559	0.2567
$\hat{\mu}$	0.5467	0.2599	0.3357	0.2135
$\hat{\nu}$	8.099	7.238	0.6385	6.033
$\hat{\kappa}$	0.369	0.4620	0.085 61	0.4822
$\hat{\sigma}_1$	0.1354	0.1977	0.2432	0.2075
\hat{r}	1.106	1.077	0.9740	0.9460
CI $\hat{\mu}$	[0.4281, 0.6652]	[0.1710, 0.3488]	[0.2412, 0.4301]	[0.1101, 0.3169]
CI $\hat{\sigma}_1$	[0.078 87, 0.232 48]	[0.1399, 0.2794]	[0.1791, 0.3303]	[0.1461, 0.2946]
$\hat{\rho}$ (km)	218.2	164.7	264.0	144.1

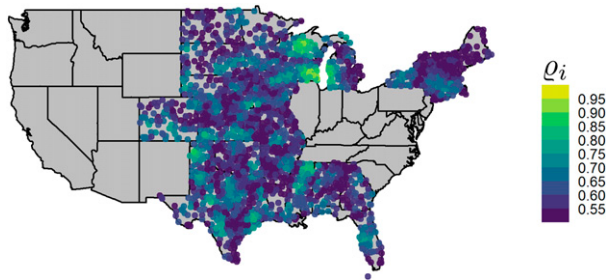


FIG. E2. Marginal contour avoiding probabilities, q_i under the conditional distribution for $X|Y$, calculated separately for each of the four regions: Midwest, Northeast, Texas, and Southeast. The contour-avoiding sets, $E_{u=0,\alpha}(X)$, remain empty in all regions under the significance level $\alpha = 0.05$.

based on the conditional distribution of $X|Y$, which is a $N(\tilde{\mu}, \tilde{\Sigma})$ distribution with

$$\tilde{\mu} = \Sigma(\Sigma + \mathbf{R})^{-1}(Y - \mu), \quad \tilde{\Sigma} = \Sigma - \Sigma(\Sigma + \mathbf{R})^{-1}\Sigma, \quad (\text{E2})$$

and Σ and \mathbf{R} are the covariance matrices of the GRF and the nugget effects, respectively. For further information on excursion sets see, for example, Bolin and Lindgren (2015). Using the R package excursions (Bolin and Lindgren 2018), we computed the marginal contour avoiding probabilities q_i for the different station locations \mathbf{s}_i , defined as $q_i = P[X(\mathbf{s}_i) > 0]$ if $E[X(\mathbf{s}_i)] > 0$, and $q_i = P[X(\mathbf{s}_i) < 0]$ otherwise, and the contour avoiding sets using the error probability $\alpha = 0.05$ (Fig. E2). These contour avoiding sets did not contain any station, lending additional support to the approximation that there is a common trend within each region.

REFERENCES

- Acero, F. J., J. A. García, and M. C. Gallego, 2011: Peaks-over-threshold study of trends in extreme rainfall over the Iberian Peninsula. *J. Climate*, **24**, 1089–1105, <https://doi.org/10.1175/2010JCLI3627.1>.
- ASCE, 2016: Failure to act: Closing the infrastructure investment gap for America's economic future. American Society of Civil Engineers, 29 pp., <https://www.infrastructureusa.org/failure-to-act-closing-the-infrastructure-investment-gap-for-americas-economic-future/>.
- Barlow, M., and Coauthors, 2019: North American extreme precipitation events and related large-scale meteorological patterns: A review of statistical methods. *Climate Dyn.*, **53**, 6835–6875, <https://doi.org/10.1007/s00382-019-04958-z>.
- Bolin, D., and F. Lindgren, 2015: Excursion and contour uncertainty regions for latent Gaussian models. *J. Roy. Stat. Soc.*, **77B**, 85–106, <https://doi.org/10.1111/rssb.12055>.
- , and —, 2018: Calculating probabilistic excursion sets and related quantities using excursions. *J. Stat. Software*, **86** (5), 1–20, <https://doi.org/10.18637/jss.v086.i05>.
- Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 224 pp.
- De Paola, F., M. Giugni, F. Pugliese, A. Annis, and F. Nardi, 2018: GEV parameter estimation and stationary vs. non-stationary analysis of extreme rainfall in African test cities. *Hydrology*, **5**, 28, <https://doi.org/10.3390/hydrology5020028>.
- Easterling, D. R., and Coauthors, 2017: Precipitation change in the United States. *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, D. J. Wuebbles et al., Eds., U.S. Global Change Research Program, 207–230, <https://doi.org/10.7930/JOH993CC>.
- Ganguli, P., and P. Coulibaly, 2017: Does nonstationarity in rainfall require nonstationary intensity–duration–frequency curves? *Hydrol. Earth Syst. Sci.*, **21**, 6461–6483, <https://doi.org/10.5194/hess-21-6461-2017>.
- Gilleland, E., and R. W. Katz, 2016: extRemes 2.0: An extreme value analysis package in R. *J. Stat. Software*, **72** (8), 1–39, <https://doi.org/10.18637/jss.v072.i08>.
- Huang, H., J. M. Winter, E. C. Osterberg, R. M. Horton, and B. Beckage, 2017: Total and extreme precipitation changes over the northeastern United States. *J. Hydrometeorol.*, **18**, 1783–1798, <https://doi.org/10.1175/JHM-D-16-0195.1>.
- Katz, R. W., M. B. Parlange, and P. Naveau, 2002: Statistics of extremes in hydrology. *Adv. Water Resour.*, **25**, 1287–1304, [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8).
- 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>.
- Kiriliouk, A., H. Rootzén, J. Segers, and J. L. Wadsworth, 2019: Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics*, **61**, 123–135, <https://doi.org/10.1080/00401706.2018.1462738>.
- Kundzewicz, Z. W., and Z. Kaczmarek, 2000: Coping with hydrological extremes coping with hydrological extremes. *Water Int.*, **25**, 66–75, <https://doi.org/10.1080/02508060008686798>.
- Kunkel, K. E., T. R. Karl, M. F. Squires, X. Yin, S. T. Stegall, and D. R. Easterling, 2020: Precipitation extremes: Trends and relationships with average precipitation and precipitable water in the contiguous United States. *J. Appl. Meteor. Climatol.*, **59**, 125–142, <https://doi.org/10.1175/JAMC-D-19-0185.1>.
- Kyselý, J., J. Picek, and R. Beranová, 2010: Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global Planet. Change*, **72**, 55–68, <https://doi.org/10.1016/j.gloplacha.2010.03.006>.
- Langbein, W. B., 1949: Annual floods and the partial-duration flood series. *Trans. Amer. Geophys. Union*, **30**, 879–881, <https://doi.org/10.1029/TR030i006p00879>.
- Li, Y., W. Cai, and E. P. Campbell, 2005: Statistical modeling of extreme rainfall in southwest Western Australia. *J. Climate*, **18**, 852–863, <https://doi.org/10.1175/JCLI-3296.1>.
- Lindgren, F., H. Rue, and J. Lindström, 2011: An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. Roy. Stat. Soc.*, **73B**, 423–498, <https://doi.org/10.1111/j.1467-9868.2011.00777.x>.
- Martel, J.-L., A. Mailhot, and F. Brissette, 2020: Global and regional projected changes in 100-yr subdaily, daily, and multiday precipitation extremes estimated from three large ensembles of climate simulations. *J. Climate*, **33**, 1089–1103, <https://doi.org/10.1175/JCLI-D-18-0764.1>.
- 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>.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, W. Zbigniew, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, **5863**, 573–574, <https://doi.org/10.1126/science.1151915>.

- Myhre, G., and Coauthors, 2019: Frequency of extreme precipitation increases extensively with event rareness under global warming. *Sci. Rep.*, **9**, 16063, <https://doi.org/10.1038/s41598-019-52277-4>.
- NOAA NCEI, 2019: Climate at a glance: Regional time series. National Centers for Environmental Information, accessed 25 February 2021, <https://www.ncdc.noaa.gov/cag/>.
- , 2020: U.S. billion-dollar weather and climate disasters. National Centers for Environmental Information, accessed 9 November 2020, <https://www.ncdc.noaa.gov/billions/>.
- NOAA's National Weather Service, 2019: NOAA Atlas 14 time series data. NOAA, accessed 1 December 2020, https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_series.html.
- Pendergrass, A. G., 2018: What precipitation is extreme? *Science*, **360**, 1072–1073, <https://doi.org/10.1126/science.aat1871>.
- Perica, S., S. Pavlovic, M. St. Laurent, C. Trypaluk, D. Unruh, D. Martin, and O. Wilhite, 2015 (revised 2019): *Precipitation-Frequency Atlas of the United States*. Vol. 10, *NOAA Atlas 14*, NOAA, National Weather Service, 265 pp., https://www.weather.gov/media/owp/oh/hdsc/docs/Atlas14_Volume10.pdf.
- R Core Team, 2017: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, <https://www.R-project.org/>.
- Risser, M. D., C. J. Paciorek, T. A. O'Brien, M. F. Wehner, and W. D. Collins, 2019: Detected changes in precipitation extremes at their native scales derived from in situ measurements. *J. Climate*, **32**, 8087–8109, <https://doi.org/10.1175/JCLI-D-19-0077.1>.
- Rootzén, H., and R. Katz, 2013: Design life level: Quantifying risk in changing climate. *Water Resour. Res.*, **49**, 5964–5972, <https://doi.org/10.1002/wrcr.20425>.
- Solomon, S., and Coauthors, 2007: Technical summary. *Climate Change 2007: The Physical Science Basis*, Cambridge University Press, S. Solomon et al., Eds., Cambridge University Press, 19–91.
- Takeuchi, K., 1984: Annual maximum series and partial-duration series—Evaluation of Langbein's formula and Chow's discussion. *J. Hydrol.*, **68**, 275–284, [https://doi.org/10.1016/0022-1694\(84\)90215-4](https://doi.org/10.1016/0022-1694(84)90215-4).
- Towler, E., B. Rajagopalan, E. Gilleland, R. S. Summers, D. Yates, and R. W. Katz, 2010: Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resour. Res.*, **46**, W11504, <https://doi.org/10.1029/2009WR008876>.