



Trends of extreme US weather events in the changing climate

Saahil Shenoy^{a,1}, Dmitry Gorinevsky^{b,1}, Kevin E. Trenberth^c, and Steven Chu^{a,d,e,1}

Contributed by Steven Chu; received May 3, 2022; accepted September 23, 2022; reviewed by Richard Smith, Michael Wehner, and Xuebin Zhang

Trends in extreme 100-y events of temperature and rainfall amounts in the continental United States are estimated, to see effects of climate change. This is a nontrivial statistical problem because climate change effects have to be extracted from “noisy” weather data within a limited time range. We use nonparametric Bayesian methods to estimate the trends of extreme events that have occurred between 1979 and 2019, based on data for temperature and rainfall. We focus on 100-y events for each month in $1^\circ \times 1^\circ$ geographical areas looking at hourly temperature and 5-d cumulative rainfall. Distribution tail models are constructed using extreme value theory (EVT) and data on 33-y events. This work shows it is possible to aggregate data from spatial points in diverse climate zones for a given month and fit an EVT model with the same parameters. This surprising result means there are enough extreme event data to see the trends in the 41-y record for each calendar month. The yearly trends of the risk of a 100-y high-temperature event show an average 2.1-fold increase over the last 41 y of data across all months, with a 2.6-fold increase for the months of July through October. The risk of high rainfall extremes increases in December and January 1.4-fold, but declines by 22% for the spring and summer months.

extreme weather events | extreme value theory | climate change

Extreme weather events incur large costs to society and damage the environment (1, 2), and the changes in their frequency have sometimes been attributed to climate change (3, 4). As the public perception of changes may be biased by selective news reporting of such events, it is vital to improve rigorous, quantitative statistical methods, coupled to a strong physical basis, to detect changes in the frequency of the extreme events.

“Climate change” refers to changes in weather over periods of at least a decade and over large spatial regions. Climate change is expected because of human-induced changes in the composition of the atmosphere, most notably, increasing carbon dioxide from fossil fuel burning (5). The change in temperature is clearly found in statistical analysis of trends from global climate data (6–8). The trends of many other variables are discussed in the Intergovernmental Panel on Climate Change assessment (9). However, links to extreme weather changes are more tenuous although much more important in terms of adverse impacts. Also, while trends in global extreme events are important, people are most interested in what is happening where they live.

This paper applies extreme value theory (EVT) (see, e.g., ref. 10) to high-quality continental US data and quantifies changes in the frequency of 100-y extreme weather events in 41 y of data. We show that trends in local extreme weather events can be estimated with confidence. Hopefully, the results of this paper will motivate the collection of improved weather datasets in other regions of the world to allow early detection of statistically significant changes in extreme events.

Empirically based analyses of extremes have been used in studies of extreme precipitation events (11–15), including extreme temperature events (16), and the changes in their probability distributions (17, 18). Extreme events are statistically described by the tail of the distribution of the variable of interest. A change in the tail of the distribution might be modeled through a shift in the mean (mean shift model) (19, 20), or a shift in the mean and change in variance of the distribution (21). Scaling and mean-constrained shifting of discrete rainfall distributions have been considered as well (18, 20). Methods for attribution of particular extreme events have also been used (22, 23).

Normal probability distributions have been shown to adequately describe the bulk of the data. Climate changes affect both the variance and the mean of the normal distributions describing temperature (7, 8, 24, 25). However, the climate science community has also recognized that extreme temperature or precipitation events are often better described by long-tail distributions used in EVT models (23, 26, 27).

EVT models for rainfall extremes have been used to describe different locations and time periods (14) and for rainfall for multiperiods (28, 29), but linear trends of the parameters were assumed. Similarly, ref. 30 builds parametric trend models for frequency

Significance

Extreme weather events rarely occur but can have devastating impacts. This paper introduces an analysis method that determines whether one-in-a-hundred-years events are becoming more frequent. Based on a 41-y record in the continental United States, the change in frequency of extreme temperature events is determined with statistical confidence. We found it is possible to aggregate extreme event data over geographical locations with different climate zones so that meaningful inference is possible. The risk of high-temperature extremes is clearly increasing, but the results for rainfall vary with season.

Author affiliations: ^aPhysics Department, Stanford University, Stanford, CA 94305; ^bDepartment of Electrical Engineering, Stanford University, Stanford, CA 94305; ^cDistinguished Scholar, NCAR and Honorary Academic, Department of Physics, Auckland University, 1010 Auckland, NZ; ^dMolecular and Cellular Physiology Department, Stanford University, Stanford, CA 94305; and ^eEnvironmental Sciences and Engineering Department in the Doerr School of Sustainability, Stanford University, Stanford, CA 94305

Author contributions: S.S., D.G., and S.C. designed research; S.S., D.G., K.E.T., and S.C. performed research; S.S., and D.G. analyzed data; and S.S., D.G., K.E.T., and S.C. wrote the paper.

Reviewers: R.L.S., University of North Carolina, Chapel Hill; M.W., E. O. Lawrence Berkeley National Laboratory; and X.Z., Environment and Climate Change Canada.

The authors declare no competing interest.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: saahils@alumni.stanford.edu, gorin@stanford.edu, or schu@stanford.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2207536119/-DCSupplemental>.

Published November 14, 2022.

and severity of extreme events that change in time. In refs. 26 and 31 and related work, EVT models of 20-y extremes in climate simulation data were built, and linear trends of the model parameters were fitted. These works have recognized that climate extremes are nonstationary.

Analysis Approach Overview

This work analyzes a 41-y historical dataset for the continental United States and obtains statistically meaningful results for 100-y event trends. We start by defining a 100-y event as an extreme having 1/100 yearly probability in a $1^\circ \times 1^\circ$ geographical area in a given calendar month. For the 990-times-larger area of the United States, about 400 such events happened over the 41 y.

This paper provides a nonparametric analysis of time trends of distribution tails. It does not a priori assume any specific form of time dependence in the variables describing climate change. Our method applies Bayesian estimation to a multilevel EVT model and determines extreme event trends across years for each calendar month. The method of data analysis, introduced by S.S. and D.G. (32–34), is computationally scalable. It allows systematic nonparametric modeling of extreme weather trends from large-scale data, such as the terabyte-scale dataset used in this study.

The block diagram in Fig. 1 illustrates our overall data processing flow in analyzing risk trends. We use the dataset from North America Land Data Assimilation System (NLDAS) of the continental United States (Mosaic Land Surface Model L4 Hourly) (35, 36). See [SI Appendix, Appendix A](#) for a discussion of the dataset.

The ground station temperature and rainfall observations are mapped by NLDAS using satellite measurements and terrain models to create a dense set of fused (virtual) data

$$W = \{\{w(l, \tau)\}_{l=1}^L\}_{\tau=1}^T, \tag{1}$$

where locations l are on a spatial grid of $0.125^\circ \times 0.125^\circ$, and 1-h time samples τ are from 1979 through 2019. This corresponds to the data in Fig. 1. The units of temperature and rainfall observations are in degrees Celsius and kilograms per square meter, respectively.

For each data point in Eq. 1, we compute an “anomaly” (deviation from a baseline) as described below in *Anomaly Computation*.

We group the grid points into $1^\circ \times 1^\circ$ location boxes (denoted by subscript s) in a given calendar month m and year t . The 990 location boxes are sorted into Koeppen–Geiger climate zones for the continental United States (37) shown in Fig. 2. Within each time block of a given month and year, the maximum anomaly (high and low temperatures and rainfall) is chosen to form the block maxima data, $x_s(t, m)$. Only one maximum value is taken for the data block. This procedure does away with extreme event

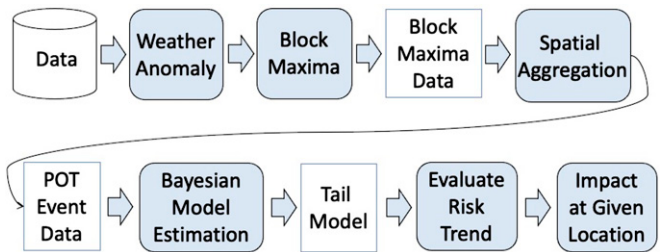


Fig. 1. Block diagram for Bayesian estimation of smoothed extreme event risk trends. *Top* shows extreme event data preprocessing and aggregation steps described in *Extreme Event Trend Modeling* and *Model Estimation*. *Bottom* shows estimations of the risk trends from the POT extreme event data as described in *Model Estimation* and *Risk Analysis Results*.

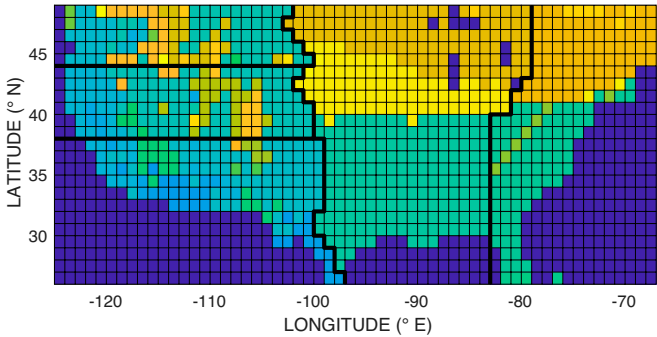


Fig. 2. Koeppen–Geiger zones for the continental United States used in data aggregation procedures. Each color represents a different Koeppen zone. In this work, the zones were further divided as shown by the thick black lines.

data clustering at a time scale less than a month and a spatial scale less than 1° . *Anomaly Computation* and [SI Appendix, Appendix A](#) provide the details.

The data $x_s(t, m)$ from different locations s but the same t and m are aggregated. The aggregated data are considered as independent samples of the tail distribution of extreme events. [SI Appendix, Appendix B](#) and *Tail Data Aggregation* provide more detail.

Next, we empirically determine the distribution of extreme events over some threshold value of temperature or rainfall. To estimate the statistical nature of the tail, we use the “peaks over threshold” (POT) procedure outlined below. These POT exceedances are used in the Bayesian MAP estimation procedure in *Model Estimation* to obtain the smoothed trend estimates of the tail model parameters ([SI Appendix, Appendix C](#)).

The tail of the distribution describes the statistical nature of the extreme events in this work. The risk of a 100-y event, which is defined in *Risk Analysis Results*, can be trended based on the estimated tail model trends. The risk of extreme events at a specific location can be predicted by specializing the estimated continent-wide risk model; see *Impact at Given Location*.

Baseline Extreme Event Model

The POT procedure creates a set of nonnegative exceedances defined as the differences between the spatially aggregated block maxima data and a threshold B ,

$$y = x - B, \tag{2}$$

conditional on $x \geq B$. A lower threshold B is used to determine the shape of the distribution tail of 100-y events by including 33-y events in POT data (Eq. 3) used for tail modeling. We assume the parameters of the distribution function determined by the lower threshold B will also describe the 100-y distribution.

The data subset formed out of Eq. 2 is called the POT exceedances

$$Y = \{y_i\}_{i=1}^n, \tag{3}$$

where $y_i = x_i - B$ for all $x_i \geq B$ and n is the events number.

The complete dataset in Eq. 3 can be considered to comprise N independent samples of the underlying distribution. Consider an “exceedance probability,” defined as

$$q = \mathbf{P}(x \geq B). \tag{4}$$

Each sample has probability q of being in Eq. 3. In a random sampling of the underlying distribution, we assume that number n

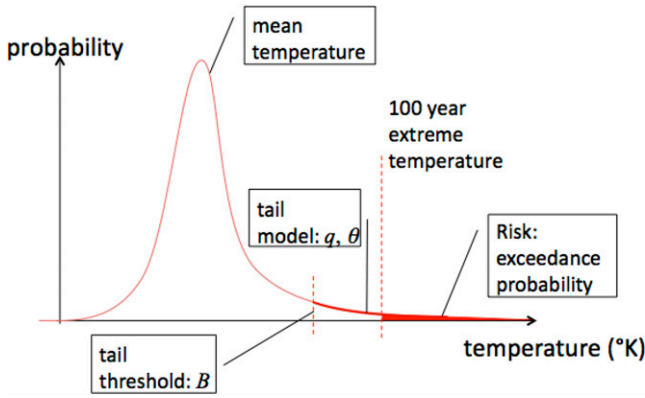


Fig. 3. Example of temperature distribution, the body and the tail. The tail starts at POT threshold B and is described by the height of the tail q and its decay rate θ . The tail distribution model allows computation of the risk of 100-y events.

of POT exceedances in Eq. 3 is described by a binomial distribution

$$p(n|q, N) = \text{Binomial}(N, q) \propto q^n (1 - q)^{N-n}. \quad [5]$$

As described in *SI Appendix, Appendix A and Appendix B*, the extreme value distribution that fits the analyzed data has an exponential tail. For sufficiently high B , the tail dataset Y (Eq. 3) can be considered as samples of a random variable y with the probability density function (PDF)

$$p(y|\theta) = \theta \cdot \exp(-\theta \cdot y), \quad [6]$$

where θ is the shape parameter.

Fig. 3 shows the parameters used to describe the tail in Fig. 1 of the distribution for extreme temperature events, which are the probability q in Eqs. 4 and 5 as well as a “shape” parameter θ defined in Eq. 6 estimated from the once-in-33-y event data. For the case of precipitation, the parameters are estimated using a data set of the logarithm of precipitation.

The parameters q and θ can be estimated using the maximum likelihood estimation (MLE) method (see, e.g., ref. 10). The POT data (Eq. 3) are used to form likelihoods following the probability models in Eqs. 5 and 6. The maximum of a positive function $f(x)$ and the maximum of its logarithm $\log[f(x)]$ occur at the same value. In our case, finding the maximum of $\log[f(x)]$ is computationally simpler. The logarithms of the likelihoods, up to an additive constant, are

$$L_q(q; n, N) = n \log q + (N - n) \log(1 - q), \quad [7]$$

$$L_\theta(\theta; Y) = \sum_{i=1}^n (\log \theta - y_i \theta). \quad [8]$$

The MLE parameter estimates \hat{q} and $\hat{\theta}$ are obtained by maximizing likelihoods (Eqs. 7 and 8) that are concave functions. The standard way to identify the maximum of a smooth concave function is to set the derivative of the function equal to zero. Hence, we set $\partial L_q(q; n, N)/\partial q = 0$ and $\partial L_\theta(\theta; Y)/\partial \theta = 0$, which yields

$$\hat{q} \equiv \arg \max_q L_q(q; n, N) = n/N, \quad [9]$$

$$\hat{\theta} \equiv \arg \max_\theta L_\theta(\theta; Y) = 1/\text{mean}(Y). \quad [10]$$

Fig. 3 illustrates the predictive power of the estimated tail model. If we assume the probability distribution shown in Fig. 3

has no time dependence, then the probability of “100-y extreme temperature” is 0.01. From Eqs. 4–6, we define the probability of exceeding a level C as $\mathbf{P}(y \geq C) = \hat{q}e^{-\hat{\theta}C}$. If \hat{q} and $\hat{\theta}$ are known, the 100-y return level C can be defined as $\mathbf{P}(y \geq C) = 1/100$, where

$$C = \hat{\theta}^{-1} \log(100 \cdot \hat{q}). \quad [11]$$

We define the “risk” of a 100-y event occurring in the shorter time period as

$$R = \hat{q}e^{-\hat{\theta}C}. \quad [12]$$

In what follows, the baseline extreme event model of this section is extended to describe seasonal variations and year-to-year trends.

Extreme Event Trend Modeling

Estimating the trend of extreme events over a short period of time requires aggregating weather data from multiple locations and time scales.

Climate and Weather Events. The sampled weather data (Eq. 1) are described by spatial location index l and time τ . The temporal weather scales range from the local time in hours $h = H(\tau)$, seasonal cycle described by calendar month $m = M(\tau)$, and years described by year index $t = Y(\tau)$. This work does not model weather events explicitly but models the impact of weather extremes by aggregating over each calendar month m and year t .

In this work, we aggregated the fine mesh spatial data (Eq. 1) in spatial boxes of longitude \times latitude size $1^\circ \times 1^\circ$ indexed as $s = S(l)$. With this definition, a weather event is characterized by a single extreme value for each box. We note that even close locations might have drastically different weather because of different altitude and terrain.

Model of Extreme Weather Events. Weather anomalies can be modeled as deviations from regular climate patterns and diurnal variations in data (Eq. 1). The diurnal variation can be significant. For example, a 20° diurnal variation amplitude can exceed a 10° difference between 10-y and 100-y events. The assumption we make is that weather events impact temperatures $w(l, \tau)$ at close locations l and times τ additively,

$$w(l, \tau) = u(l, M(\tau), H(\tau)) + v(l, \tau), \quad [13]$$

where $m = M(\tau)$ is the calendar month, $h = H(\tau)$ is the day hour for time sample τ , $u(l, m, h)$ is the baseline temperature for locality l in month m at hour h , and $v(l, \tau)$ is the weather anomaly.

Statistical extreme events depend on locality and are specific for year and calendar month. We define the block maxima extremes as

$$x_s(t, m) = \max_{M(\tau)=m, Y(\tau)=t, S(l)=s} v(l, \tau), \quad [14]$$

where $m = M(\tau)$ is calendar month for time τ , $t = Y(\tau)$ is the year, and $s = S(l)$ indexes the $1^\circ \times 1^\circ$ spatial box location.

The proposed probabilistic model describes extreme event data at all locations by a POT distribution

$$x(t, m) = y(t, m) + B(m, z), \quad [15]$$

where for POT events $x(t, m) \geq B(m, z)$, the deterministic threshold $B(m, z)$ depends on climate zone $z = Z(l)$ as described in the next section, and $y(t, m)$ is an exponentially distributed random variable. The extreme event trend is defined

by how the distribution parameters of the tail variable $y(t, m)$ depend on year t .

Data $x_s(t, m)$ in Eq. 14 are presumed to be samples of random variable $x(t, m)$ in Eq. 15. While thresholds $B(m, z)$ of temperature depend on the climate zone z , our model is that the POT data follow exponential distributions for all locations s with the same shape parameter. This empirical model fits data well and allows us to aggregate the data for all POT exceedances across all spatial locations, for a given calendar month m , in statistical estimation (see *Tail Data Aggregation*). In case of precipitation, we use a power law distribution model, as briefly mentioned in the paragraph below Eq. 6.

Model Estimation

Anomaly Computation. Our computation of weather anomalies differs from the sometimes used National Oceanic and Atmospheric Administration anomaly definition (38). In the estimate of the model parameters \hat{q} and $\hat{\theta}$ of extreme events, we begin by establishing a locality-dependent temperature baseline u in Eq. 13 by averaging data in Eq. 1 over all years,

$$u(l, m, h) = \text{mean}_{M(\tau)=m, H(\tau)=h} w(l, \tau). \quad [16]$$

For temperature, a single diurnal sequence of 24 hourly values $u(l, m, h)$ was calculated at each location l for each calendar month m . Examples of the diurnal mean temperature baselines $u(l, m, h)$ are shown in Fig. 4 for location l in Los Angeles for March ($m = 3$) and November ($m = 11$) and in Washington, DC ($m = 1$ and $m = 8$).

The temperature extreme events are analyzed by looking at temperature anomalies v , defined as the differences between the hourly temperature data w and the diurnal mean u at each grid location; see Eq. 13. We compute the temperature anomaly for each data point as $v(l, \tau) = w(l, \tau) - u(l, M(\tau), H(\tau))$. The plots in Fig. 4 show the hourly NLDAS temperature data $w(l, \tau)$

for two particular days, which deviate from the mean diurnal cycle baseline for that month.

The analysis of precipitation presents difficulties, since rainfall is not an everyday occurrence. At each grid location l , we use cumulative values for the last 5 d at daily samples τ to reflect the risk of flooding. In addition, we compute $w(l, \tau)$ in Eq. 13 as $\log(\text{rainfall} + \epsilon)$, where $\epsilon \approx 10^{-16} \text{ kg/m}^2$ to keep the result finite on days without rainfall. The baseline $u(l, m)$ in Eq. 13 is computed similarly to Eq. 16, except the cumulative precipitation does not depend on the hour. The precipitation anomaly in Eq. 13 is $v(l, \tau) = w(l, \tau) - u(l, M(\tau))$.

The temperature and precipitation anomaly data $v(l, \tau)$ inside each $1^\circ \times 1^\circ$ spatial box s for each month m of year t of data are used to compute extreme events data $x_s(t, m)$ according to Eq. 14. (See *SI Appendix, Appendix A and Appendix B* for details.)

Tail Data Aggregation. The estimation of distribution tail in Fig. 3 requires aggregating the POT data across the Koeppen zone. Threshold B in Eq. 2 was initially selected as the 33-y event level based on the block maxima anomaly data in the zone.

We have found that a surprisingly simple exponential tail model Eq. 6, where the tail rates θ are the same for all locations, fits the extreme event data well. This finding allows aggregation of all data above the Koeppen zone-dependent threshold B and obtaining consistent estimates for θ in Eq. 10.

The validity of using the same exponential tail model for a given month and year across different Koeppen zones is illustrated by QQ plots in Fig. 5. (QQ plots and their use are explained below *SI Appendix, Eq. S13*.) These QQ plots show the aggregate data fits for estimated tail models across all years and zones for a given calendar month. In the vicinity of a 100-y extreme event return level, the tail distribution fits the data fairly well. There is not much variation of tail data fits across the zones. For more on the same tail rate fitted across the locations, see *Cross-Validation* and *SI Appendix, Appendix D*. For more on modeling the tail as exponential, see *SI Appendix, Appendix A and Appendix B*.

The distribution tail theoretically extends to infinity, while there are physical limits to the maximum temperature and rainfall. (This is true for either long-tail or normal distributions.) What ultimately matters is model prediction of quantile (return level) for extreme events of interest.

The quantile predictions in the QQ plots of Fig. 5 seem to show that the estimated model has predictive power beyond 300-y events. This model estimated from 41 y of data is limited to a reasonable range of the climate variables and is not suitable for predicting the risk of exceedingly rare events.

Spatial Dependency of Extremes. Many weather patterns have a large scale in both space and time. This creates clusters of interdependent extreme data points for the same weather event. The block maxima (Eq. 14) for data collected at 64 grid points in each $1^\circ \times 1^\circ$ box location over a month does away with extreme event data clustering at a time scale less than a month and a spatial scale less than 1° since only one maximum value is taken for the data block.

However, extreme weather events at different spatial $1^\circ \times 1^\circ$ locations and different months were assumed to be independent. If an extreme event is described by an average size location clustering $k > 1$, the extreme events are overcounted by a factor of k . The spatial dependency is discussed in more detail in *SI Appendix, Appendix D*. The calibration factors provided in *SI Appendix, Data Clustering* are $k = 3.35$ and 3.79 for high- and low-temperature events, and $k = 1.55$ for rainfall events. Based on the observed cluster sizes, the tails will be $\sim 10\%$ longer than

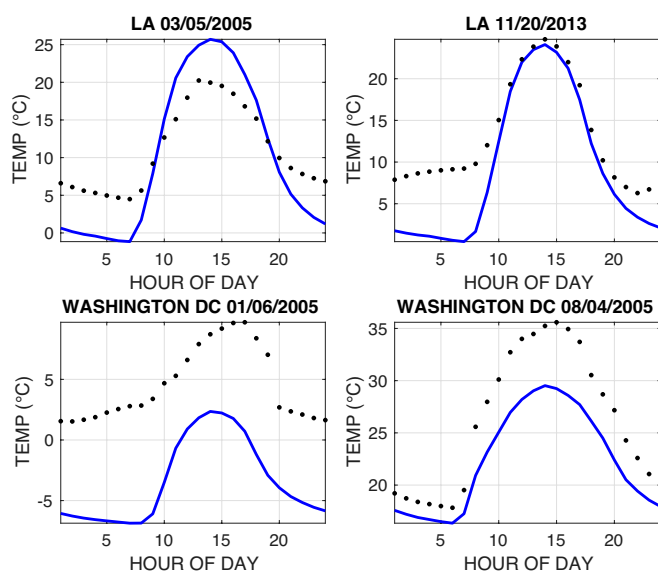


Fig. 4. Example plots of temperature and the diurnal mean for Los Angeles, CA, and Washington, DC, locations showing 2 d in March/November and January/August, respectively. The dotted curve is the temperature recorded at the listed date, and the solid line curve is the mean diurnal cycle. For the Los Angeles examples, the maximum deviations from the mean occur in the morning. For the Washington, DC, examples, the maximum deviations occur in either the morning or afternoon.

estimated in *Trend Estimation*, and the trends of the 100-y events are expected to be less affected.

Trend Estimation. The extreme weather trends are estimated by fitting Eq. 15 to block maxima data $x_s(t, m)$ (Eq. 14). We set $B(m, z)$ in Eq. 15 as the 33-y return level of the aggregated extreme event data $x_s(t, m)$ within each Koeppen zone $z = Z(s)$ for all years t of the data. The POT dataset $Y_{tm} = \{x_s(t, m) - B(m, z)\}$ is aggregated across all Koeppen zones for each calendar month m and year t . With this procedure, the maximum likelihood formulation for POT data is extended for multiple months m and years t .

The distribution parameters θ_{tm} and q_{tm} are estimated from the POT data. One method would be to use MLE, which assumes that the parameters for a given month and year are independent of each other. Instead, we introduce a prior distribution (the assumed distribution of the parameters in the absence of data) where the values of θ_{tm} and q_{tm} are assumed not to strongly depart from $\theta_{t'm'}$ and $q_{t'm'}$ for adjacent years of the same month and adjacent months of the same year. Across years and months of data, the distribution of the parameters is modified to what is referred to as the posterior probability distribution.

A posterior distribution of the parameters is created by forming the normalized product of the likelihood and the prior distribution. Estimating the parameters by maximizing the posterior distribution is known as the Bayesian maximum a posteriori procedure (referred to as “MAP”).

The MAP maximization of the parameters of the posterior distribution is most easily done (similar to MLE) by considering the logarithm of the posterior distribution function. The log-posterior is then the sum of log-likelihoods, which have forms similar to Eq. 7, and Eq. 8 and log-priors. The details of MAP problem formulation and solution are presented in [SI Appendix, Appendix C](#). The MAP estimates are computed by solving two problems of maximizing the log-posterior distributions for $q_{t,m}$ and $\theta_{t,m}$,

$$\begin{aligned} \theta_{tm} = \arg \max & \left[\sum_{t=1}^T \sum_{m=1}^{12} L_{\theta}(\theta_{tm}, Y_{tm}) \right. \\ & + \zeta_1 \sum_{t=2}^T \sum_{m=1}^{12} U_{\theta}(\theta_{tm}, \theta_{t-1,m}) \\ & + \zeta_2 \sum_{t=1}^T \sum_{m=2}^{12} U_{\theta}(\theta_{tm}, \theta_{t,m-1}) \\ & \left. + \zeta_2 \sum_{t=2}^T U_{\theta}(\theta_{t,1}, \theta_{t-1,N}) \right], \end{aligned} \quad [17]$$

$$\begin{aligned} q_{tm} = \arg \max & \left[\sum_{t=1}^T \sum_{m=1}^{12} L_q(q_{tm}, n_{tm}, N_{tm}) \right. \\ & + \eta_1 \sum_{t=2}^T \sum_{m=1}^{12} U_q(q_{tm}, q_{t-1,m}) \\ & + \eta_2 \sum_{t=1}^T \sum_{m=2}^{12} U_q(q_{tm}, q_{t,m-1}) \\ & \left. + \eta_2 \sum_{t=2}^T U_q(q_{t,1}, q_{t-1,N}) \right], \end{aligned} \quad [18]$$

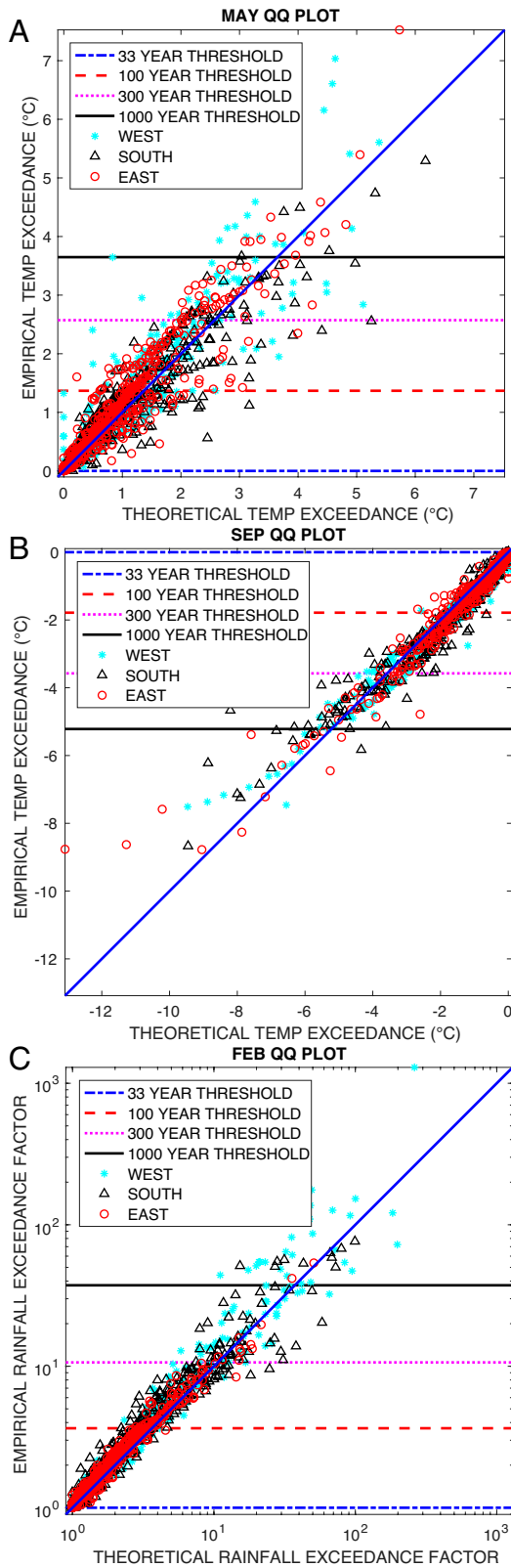


Fig. 5. QQ plots for all zones: (A) high temperature, (B) low temperature, and (C) precipitation. The horizontal axes are the quantiles of exponential distributions for the temperatures and Pareto distribution for the precipitation. The vertical axes are the empirical quantiles for each year across all zones: POT exceedances of threshold B for the temperatures and scaled empirical data x/e^B for the precipitation. We further break down quantile data that are from the west, south, and east regions as used in Fig. 14; see *Cross-Validation*. These QQ plots show that the tail model fit is a fairly good fit across the Koeppen zones.

2019	$\theta_{T,1}$	$\theta_{T,2}$	\cdots	$\theta_{T,11}$	$\theta_{T,12}$
\vdots	\updownarrow	\updownarrow	\cdots	\updownarrow	\updownarrow
1980	$\theta_{2,1}$	$\theta_{2,2}$	\cdots	$\theta_{2,11}$	$\theta_{2,12}$
1979	$\theta_{1,1}$	$\theta_{1,2}$	\cdots	$\theta_{1,11}$	$\theta_{1,12}$
	Jan	Feb	\cdots	Nov	Dec

Fig. 6. The structure of the year-to-year prior. This prior relates parameters for a given month to the same month of the previous year. The illustration is for θ_{tm} ; the same prior structure is used for q_{tm} .

where $t = 1, \dots, T$; $m = 1, \dots, 12$. Expressions for log-likelihoods $L_\theta(\theta, Y)$ and $L_q(q, n, N)$ are given by Eqs. 7 and 8; n_{tm} is the exceedance number for year t and month m , and N_{tm} is the total number of data points. In Eqs. 17 and 18, the parameters ζ_1 , ζ_2 , η_1 , and η_2 are fixed. *SI Appendix, Appendix C* describes concave log-priors $U_\theta(r, s)$, and $U_q(r, s)$ and prior parameters ζ_1 , ζ_2 , η_1 , and η_2 that control the smoothness of the solution.

The prior connection between year-over-year parameters is defined by the parameters ζ_1 and η_1 in Eqs. 17 and 18 and illustrated in Fig. 6. If, for example, there is no year-to-year change in θ_{tm} across all years, $\zeta_1 = \infty$. Similarly, if q_{tm} remains constant, $\eta_1 = \infty$. The exact form of the prior distributions for the parameters $\theta_{t,m}$ and $q_{t,m}$ does not significantly alter the result. The main requirement is that the prior PDFs $U_\theta(x, y)$ and $U_q(x, y)$ have a single maximum at $x = y$.

In addition to the previously described priors, we use a second set of priors for smoothing across calendar months defined by the terms $U_\theta(\theta_{tm}, \theta_{t,m-1})$ in Eq. 17 and $U_q(q_{tm}, q_{t,m-1})$ in Eq. 18 as illustrated in Fig. 7. In the case of the month of January, the indexes in the equations change to reflect that the previous month is December of the previous year.

2019	$\theta_{T,1} \leftrightarrow \theta_{T,2} \leftrightarrow \cdots \leftrightarrow \theta_{T,11} \leftrightarrow \theta_{T,12}$
\vdots	$\cdots \cdots \cdots \cdots$
1980	$\theta_{2,1} \leftrightarrow \theta_{2,2} \leftrightarrow \cdots \leftrightarrow \theta_{2,11} \leftrightarrow \theta_{2,12}$
1979	$\theta_{1,1} \leftrightarrow \theta_{1,2} \leftrightarrow \cdots \leftrightarrow \theta_{1,11} \leftrightarrow \theta_{1,12}$
	Jan Feb \cdots Nov Dec

Fig. 7. The structure of the month-to-month prior. This prior relates parameters for a given year and month to the parameters for the previous month. The illustration is for θ_{tm} ; the same prior structure is used for q_{tm} .

Discussion of Bayesian Estimation Method. Unlike signal processing methods commonly used for filtering (e.g., moving averages), the optimization-based smoothing (Eqs. 17 and 18) yields optimal estimates in the ends of the interval $[1, T]$. Parameters in Eq. 18 characterize the trend for extreme events frequency, while Eq. 17 describes the trend of magnitude for extreme events at a given frequency (e.g., 100-y events). The trend estimation is nonparametric: It does not assume any specific shapes of the trend curves. Convex optimization approaches of this type are commonly used in machine learning. The method takes POT data $\{Y_{tm}\}$ as an input and computes the tail parameter trends $\{\theta_{t,m}, q_{t,m}\}$ as the output. It uses four scalar parameters ζ_1 , ζ_2 , η_1 , and η_2 ; see *SI Appendix, Tuning of the Smoothing Parameters*.

Risk Analysis Results

The estimated model can be used for risk analysis. The changing number and severity of extreme events expected to occur once in 100 y at any location in the continental United States are estimated. The 100-y event return level $C(m)$ is determined for a given month m . Aggregating extreme event POT data $y_s(t, m) = x_s(t, m) - B(m, Z(s))$ over all location boxes s and years t gives the extreme event set Y_m defined in Eq. 3. The 100-y extreme events are supposed to make 0.01 of the total events times the number of boxes s .

The risk $R_{t,m}$ of an extreme event in a given calendar year and month is defined by the probability of an event exceeding $C(m)$, as discussed in Eq. 11. By using risk formula Eq. 12,

$$R_{t,m} = \hat{q}_{tm} \cdot e^{-\hat{\theta}_{tm} C(m)}, \quad [19]$$

where $\hat{\theta}_{tm}$ and \hat{q}_{tm} are parameter estimates (Eqs. 17 and 18). By the definition of a 100-y event, relative risks in temperature and rainfall over the entire time period $100 \cdot R_{t,m}$ are one. *SI Appendix, Appendix F* shows the results for the relative risks $100 \cdot R_{t,m}$ in tabular form. *SI Appendix, Appendix D* discusses confidence in the results, including CIs for risk $R_{t,m}$.

Extreme High Temperature Results. The US geographic area results (Fig. 8) for the 12 calendar months reveal that the risk increases by a factor of 2.1, on average, over the 41 y. By our statistical methodology, the 2011 North American heat wave,

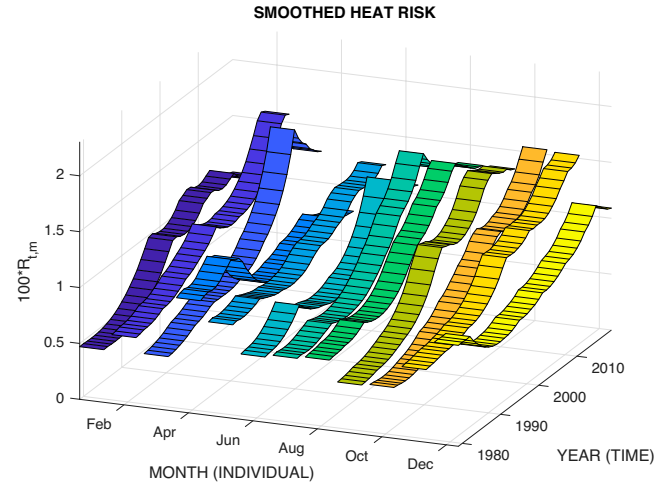


Fig. 8. Plots of the risk (in percent) of 100-y high-temperature events in the US geographic area. Each ribbon represents a calendar month. The smoothed trends are shown for years from 1979 to 2019. Apart from April, all risks trend upward.

SMOOTHED COLD RISK

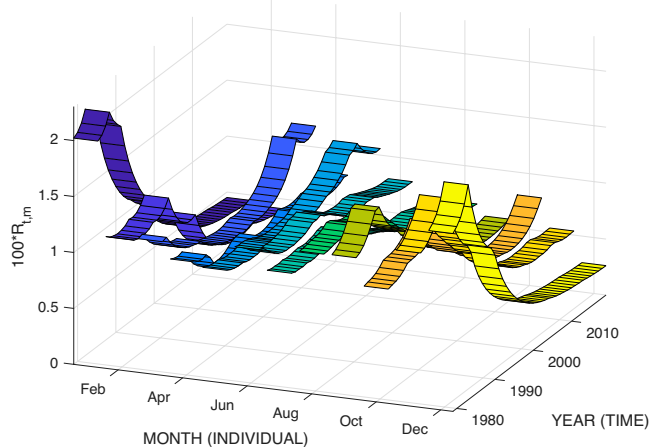


Fig. 9. Plots of the relative risk of 100-y low temperature events between 1979 and 2019. Almost all months have downward trends, but May trends upward.

which brought record high temperatures to the Midwest and East of the United States, is considered to be a 300-y event for the month of July. Similarly, the 2012 North American heat wave in June that resulted in 82 heat-related deaths, and the June 2015 Northwest US heat wave, are also considered to be 300-y events.

Extreme Low Temperature Results. Fig. 9 shows the extreme low-temperature trends for the United States. The risk of an extreme low-temperature event increases by 5% for the month of May and is almost flat for March, over the 41 y. The risk decreases by 51% on average for the remaining months, as expected for a warming climate.

Extreme Precipitation Results. Precipitation in the United States occurs mainly in the form of rainfall and snow. Fig. 10 shows the trends of the risk for extreme rainfall in the $1^\circ \times 1^\circ$ boxes for 5-d cumulative rainfall. The risk increased by a factor of 1.4 for the winter months of December and January, over the 41 y. As we did not include snowfall, the total precipitation is even more pronounced. The risk trend is nearly flat for spring, summer, and fall. The May risk is the highest. By our statistical methodology,

SMOOTHED RAINFALL RISK

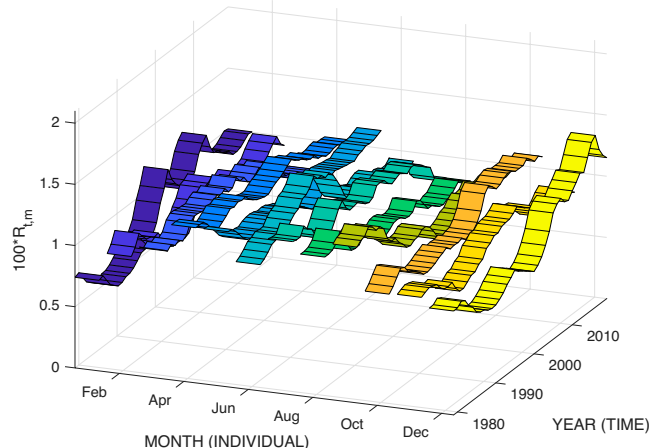


Fig. 10. The relative risk of 100-y rainfall events in the United States from 1979 to 2019. The winter months of January and December have an upward trend, while, for other seasons, the trends are relatively flat.

flooding of southeast Texas in May 2015 and of Corpus Christi in May 2016 exceeded the 300-y level.

Impact at a Given Location

As discussed in *Model Estimation*, estimating parameters of the extreme events distributions at different climate zones allows us to predict the approximate relative risk for extreme event tails at all locations. This allows determination of the extreme event impact at a given location and time.

The model parameters $\theta_{m,t}$ and $q_{m,t}$ of the tail distribution obtained for month m of year t define the risk in Eq. 19. We define the predicted 100-y temperature return level $D_{100}(t, m)$ for POT variable $y_{t,m} = x_{t,m} - B$ in Eq. 5, 6 by equating the risk $q_{m,t} \cdot e^{-D_{100}(t,m) \cdot \theta_{m,t}}$ to 100-y event probability $p_{100} = 1/100$. Solving for $D_{100}(t, m)$ yields

$$D_{100}(t, m) = \theta_{m,t}^{-1} \log(100 \cdot q_{m,t}). \quad [20]$$

This formalism for the 100-y event can be easily generalized to $D_K(t, m)$, where $K = 300$ or 500.

For temperature anomaly block maxima $x_{t,m}$, the 100-y return level becomes $B(z, m) + D_{100}(t, m)$, where $B(z, m)$ depends on the Koeppen zone z (Eq. 15). The anomaly $x_{t,m}$ is related to temperature $w = T(l, \tau)$ at location l at time τ through the baseline $u(l, m, h)$, where $m = M(\tau)$ is the month and $h = H(\tau)$ is the day hour. (See Eqs. 13 and 14.) This allows computing the 100-y temperature return level $T_{100}^H(t, m)$ for high-temperature $w(l, \tau)$ in a given month m and year t . The maximum value $x_{t,m}$ for anomaly $v(l, \tau) - u(l, m, h)$ in location box $s = S(l)$ and month $m = M(\tau)$ exceeds the 100-y return level $B(z, m) + D_{100}(t, m)$ with probability $p_{100} = 1/100$.

This procedure gives us the 100-y return level at given location l and time τ ,

$$T_{100}^H(l, \tau) = u(l, m, h) + B(z, m) + \theta_{m,t}^{-1} \log(100 \cdot q_{m,t}), \quad [21]$$

where $m = M(\tau)$, $h = H(\tau)$, $t = Y(\tau)$, and $z = Z(l)$. Fig. 11 shows an example for a heat wave in Tacoma, WA.

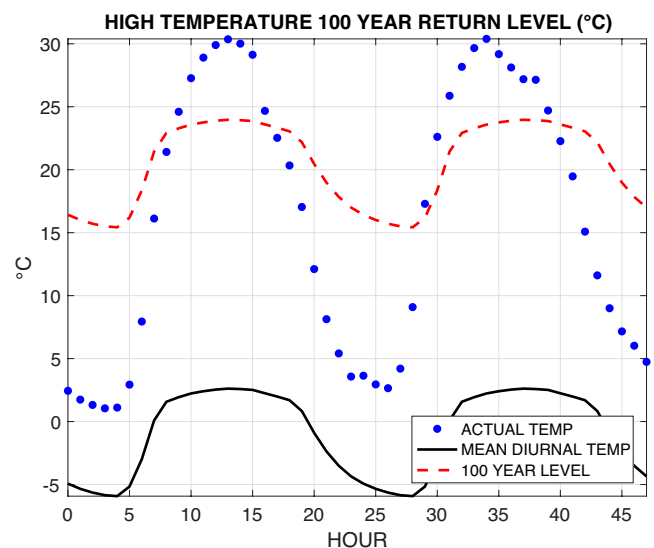


Fig. 11. Comparison of hourly temperature data $w(l, \tau)$, mean diurnal cycle $u(l, m, H(\tau))$, and 100-y high-temperature event return level $T_{100}^H(l, \tau)$ in Eq. 21. The data are plotted from June 26, 2015 at 12 AM to June 27, 2015 at 11 PM for location grid point l near the city of Tacoma, WA. The plot argument is the number of hours after the start time. For two plotted days, the temperature exceeds the 100-y return level. This corresponds to the 2015 northwestern heat wave.

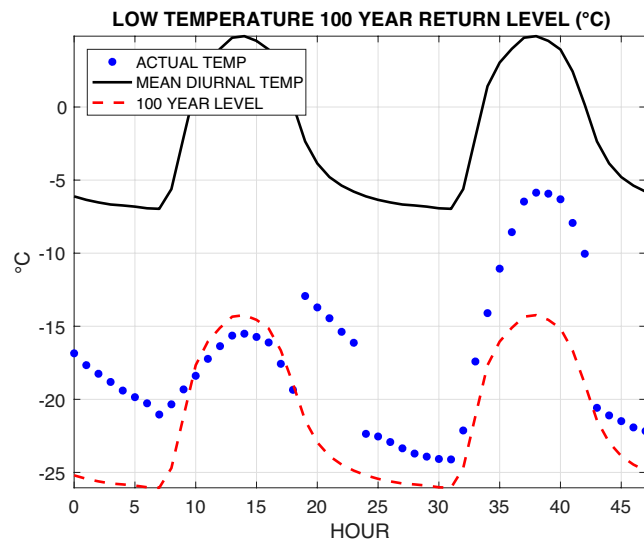


Fig. 12. Comparison of hourly temperature data $w(l, \tau)$, mean diurnal cycle $u(l, m, H(\tau))$, and 100-y cold event return level $T_{100}^C(l, \tau)$ in Eq. 22. The data are plotted from February 19, 2015 at 12 AM to February 20, 2015 at 11 PM for location grid point l near Cherokee National Forest in Elizabethtown, TN. The plot argument is the number of hours after the start time. The temperature dips below the 100-y cold threshold for 8 h on the first plotted day. This corresponds to the 2014–2015 North American winter event.

Similarly, for 100-y extreme cold temperature level,

$$T_{100}^C(l, \tau) = u(l, m, h) - B(z, m) - \theta_{m,t}^{-1} \log(100 \cdot q_{m,t}), \quad [22]$$

where $z = Z(l)$, $m = M(\tau)$, $t = Y(\tau)$, and $h = H(\tau)$. The mean diurnal cycle $u(l, m, h)$, threshold $B(z, m)$, and tail parameters $\hat{\theta}_{tm}$, \hat{q}_{tm} are estimated as described in *Model Estimation*. Fig. 12 shows an example for a cold wave in Cherokee National Forest in Elizabethtown, TN.

A 100-y return level for extremes of the log-precipitation has a similar form, except we model a 5-d average precipitation per day (see *Anomaly Computation*).

$$\log P_{100}(l, \tau) = u(l, m, h) + B(z, m) + \theta_{m,t}^{-1} \log(100 \cdot q_{m,t}). \quad [23]$$

By exponentiating Eq. 23, we get

$$P_{100}(l, \tau) = e^{u(l, m)} \cdot e^{B(z, m)} \cdot (100 \cdot q_{tm})^{\theta_{tm}^{-1}}, \quad [24]$$

where the model parameters are estimated as described in *Model Estimation*. An example for a 2016 flood in Corpus Christi, TX, is shown in Fig. 13. This flood event is classified as just above the 100-y return level by our statistical analysis for the selected $1^\circ \times 1^\circ$ location. By comparison, the rainfall from Hurricane Harvey in 2017 in the Houston area was significantly above the 100-y level.

The described prediction of the risk at given locations interpolates the exponential tail model across calendar months, years, and locations. We found that, occasionally, the computed 100-y return level $B + D_{100}(t, m)$ is smaller than the 33-y event threshold $B = B(z, m)$ in Eq. 15. This is caused by sparsity of the data for some periods. To address the problem, one can reduce B (e.g., to a 20-y event threshold) or increase longitudinal smoothing in Eq. 18 (parameter η_1).

Discussion

Cross-Validation. The presented extreme event model is tested by cross-validation. We break the continental US geographic area

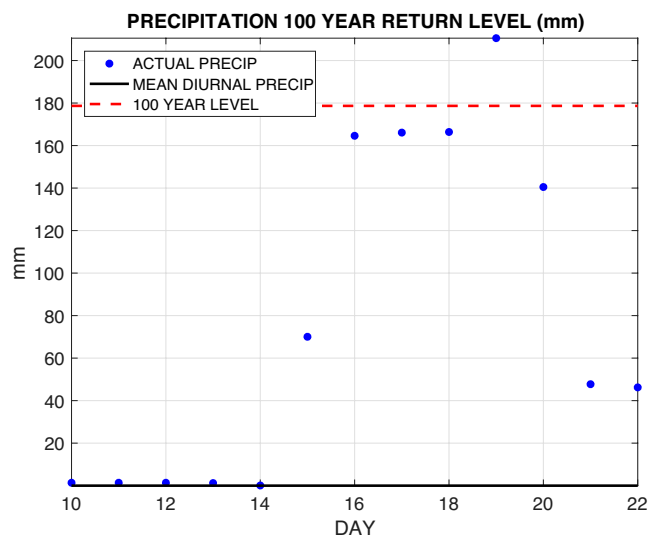


Fig. 13. Comparison of 5-d daily average precipitation data $e^{u(l, m, H(\tau))}$, monthly mean precipitation $e^{u(l, m, H(\tau))}$, and 100-y event return level $P_{100}(l, \tau)$ in exponentiated Eq. 23. The data are plotted from May 10 to May 22, 2016 for location grid point l near the city of Corpus Christi, TX. This corresponds to the mid-May flooding event.

into west, south, and east partitions shown in Fig. 14. The POT data for two of the partitions are combined to estimate the tail model for a given month. This model is then tested with the POT data for the remaining partition. There are three cross-validation tests for each of the 12 mo.

Fig. 15 shows QQ plots as diagnostics for the cross-validation. The high-temperature, low-temperature, and high-rainfall models are fitted to 41 y of the test data for a selected calendar month. For example, Fig. 15A plots empirical quantile data for all years of the selected month in the west partition against the model quantiles. The model distribution (Eqs. 5 and 6) is determined from the data in the south and east partitions. Similar plots were made for predicting south and east partitions data. Fig. 15 shows a reasonable model fit.

Welch's t test was used to test the hypotheses that each two cross-validation sets have equal θ_{tm} for each year t and month m , the total of $12 \times 41 \times 3 = 1,476$ tests. *SI Appendix, Appendix D* provides more detail. The hypothesis holds at the 95% level for the high and low temperature models in the three test zones and all months m and years t , except in less than 2% of the tests. For the

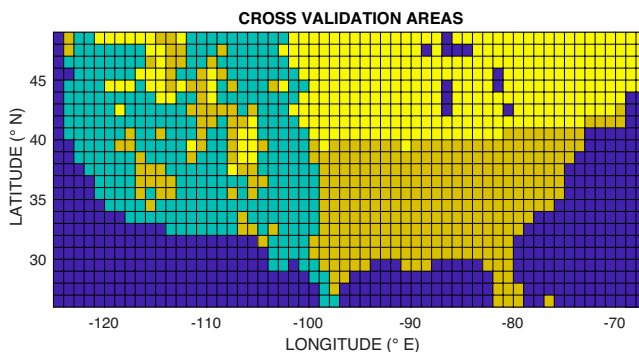


Fig. 14. Cross-validation partitioning of the continental US area. Each colored area represents a partition used for testing the tail model. The tail model is trained on the two remaining partitions. Partitions have 330 location boxes, on average. There are four colors in total, and all the dark blue location boxes correspond to bodies of water that are not included in the analysis.

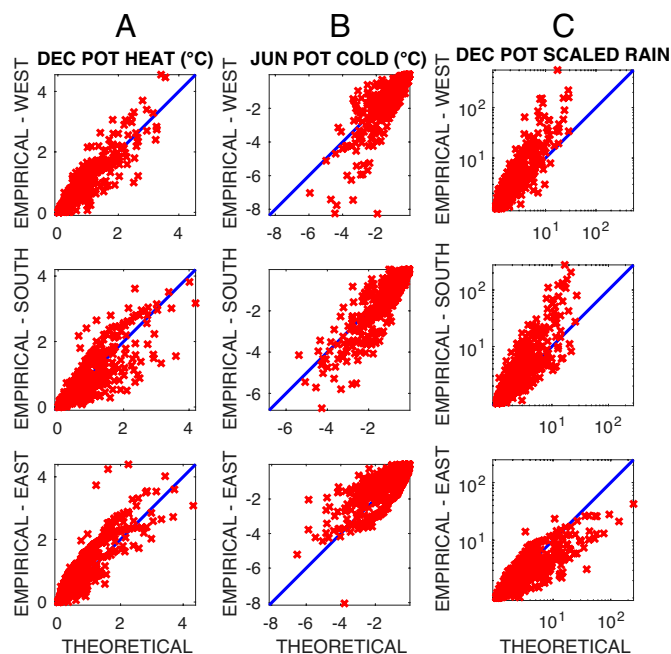


Fig. 15. QQ plots to illustrate how well the tail model trained on data for two partitions fits the remaining test partition: (A) high-temperature exceedances (in degrees Celsius) for the month of December, (B) low-temperature exceedances for the month of June, and (C) extreme rainfall for the month of December. The y and x axes show the empirical quantiles vs. the tail model (theoretical) quantiles for the 33-y event threshold exceedances; exponential tail for the temperatures and Pareto tail for the rainfall.

precipitation, it holds at the 99% level in 64% of all tests, which makes the tail model universality somewhat less certain.

Mean Shift Tail Model. A simple model attributes the change in the number of extreme events to a shift in the mean of the distribution, with the distribution shape assumed the same, as illustrated in Fig. 16. Such mean shift models have been considered in earlier work, for example, refs. 18–21. The question then is, “How do the results of this paper relate to such modeling?” To answer, we use the “mean shift” model discussed in detail in *SI Appendix, Appendix E*. It is more accurate than the prior work, because it more precisely describes the shape of the tail. This model can also be used to correct for the degree of spatial correlation of rare events. Our Bayesian model is preferred, since it describes year-over-year changes in the tail shape. Interestingly, the Bayesian model, while more rigorous, is found to be generally consistent with our mean shift model for temperature extremes. (See *SI Appendix, Appendix E* and ref. 39.)

SI Appendix, Appendix E shows a discrepancy between the Bayesian and the mean shift model risk trends for extreme rainfall events. The risk trends estimated by the Bayesian model have considerably less year-to-year variation than the mean shift model,

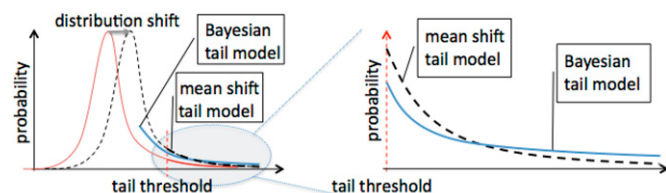


Fig. 16. Conceptual comparison of the mean shift model and the Bayesian extreme event model. *Left* illustrates the extreme event increase caused by the distribution shift. *Right* zooms in on the distribution tails and illustrates the difference between the tails for the shifted distribution and the Bayesian model.

as shown in *SI Appendix, Fig. S17*. Despite the fact that the average rainfall is decreasing, this analysis shows that extreme rainfall events are occurring with similar probabilities from year to year. If total precipitation were included, the trend is likely to be toward increased risk of flooding.

Climate Science Perspective on Results. Much has been written about extreme weather events. In comparing those results to the analysis in this paper, one should note that there are differences in how “extreme events” are defined.

The trends of extreme events in this paper suggest increasing damage risks (especially droughts and wildfires) related to high-temperature extremes. During the period from 1979 to 2019, the global mean land and surface temperature increased by $\approx 0.8^\circ\text{C}$. The average monthly conditions in this record can be extended for well over a century and help place the post-1979 period in some context.

Higher temperatures increase the water-holding capacity of the atmosphere, and the increased buoyancy and moisture are projected to lead to an increased risk of heavier rainfall rates and stronger storms provided that moisture is available (3, 15, 18, 20, 40), such as the many rainfall extremes in the eastern United States associated with hurricanes.

Our statistical analysis shows that the main increases in precipitation extremes are in the colder and drier months of December and January. A warm anomaly allows a large increase of moist air which promotes higher-intensity precipitation and may yield bigger rainfalls, or more snow (which was not monitored in this dataset). Winter storms are generally bigger, draw moisture from larger regions such as the Gulf and adjacent oceans, and convert the moisture into precipitation. Hence moisture limitations are not so much in play. On the other hand, summer storms are much smaller and dependent on more local sources of moisture, that is, more convective complexes. They are more likely to be starved of a moisture source, and their full water-holding capacity is not realized. The overall relative risk in extreme precipitation amount shows a modest increase, but is likely a function of space and time scales.

The other seasonal factor in play is that, in winter, the precipitation is from extratropical cyclones in which warm moist advection occurs. The southerlies are warm and moist, and there is a positive correlation between moisture and temperature. In contrast, in the summer, there tends to be a negative correlation because of a strong positive feedback that occurs on land. A drier region means less evaporative cooling, which, in turn, increases the temperature and drives more atmospheric demand for moisture, resulting in still more drying and drought conditions.

Conclusions

This paper introduces a nonparametric Bayesian approach to characterize statistically significant trends of extreme, 100-y temperature and rainfall events from weather data over a 41-y time period. This statistical analysis is independent of any climate models that attribute changes in extreme events to anthropogenic or other factors (3, 4).

The most important enabling factor of the analysis presented in this work was our discovery that the parameters of a distribution tail model used to describe the risk of 100-y events are approximately the same across all climate zones.

The analysis uses hourly (or, at a minimum, daily) data collected over several decades and covering a large spatial area of the continental United States with high spatial resolution based on ground station and satellite measurements, but the same analysis

needs to be applied to different US datasets. We were unable to apply the same analysis to Europe; although central Europe has complete data, the historical ground station data in other parts of Europe were insufficient.

Improved methods of detecting trends in extremes in rainfall and drought conditions in the summer months will have an impact on food security and the risk of wildfires. Similarly, extreme rainfall in populated, flood-prone areas raises more immediate concerns than extreme rainfall in remote arid regions. Increased sea levels or seasonal river heights will make certain regions more prone to major flooding events. More-sophisticated risk models are required to understand how these extreme flood events are affected by climate change. Understanding groundwater levels in aquifers is also crucial for water resource management for extended periods of extreme drought.

Future Work

This paper presents an analysis method that shows the robustness of tail structures. This work opens up future possibilities such as the regional analysis of the extreme events we considered, or other types of extreme events.

In the current analysis, rainfall anomalies are computed in a similar manner to temperature anomalies, but the occurrence of rainfall is a sparse event. A 5-d moving cumulative sum for rainfall was used in order to reduce sparsity in the rainfall data,

but extreme rainfall over a few hours is the most life threatening. To properly address rainfall extremes, hourly data of the intensity of precipitation need to be measured.

Given the importance of monitoring changes in the frequency and intensity of extreme events, all countries could benefit from data-driven modeling of increases in heat waves, floods, and threats to local agriculture and water supplies. The deployment of solar/anemometer-powered, autonomous, and tamper-resistant weather stations with cellular/satellite communications could be used to create a sufficiently dense set of ground stations, especially in sparsely populated locations. Given the social and economic costs of extreme events, there is a critical need for high-quality data and improved statistical analysis methods to extract crucial trends in disastrous weather events around the world.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*. NLDAS Mosaic Land Surface Model L4 V002 data are available at https://disc.gsfc.nasa.gov/datasets/NLDAS_MOS0125_H_002/summary (35).

ACKNOWLEDGMENTS. We express our gratitude, for help and valuable discussions, to Prof. Bradley Efron and Mr. Stefan Wager from the Stanford Statistics Department and to Dr. Eberhard Faust from MunichRe. We also thank Professors Gabriel Veechi, Michael Wehner, Xuebin Zhang, and Richard Smith for their invaluable feedback in improving the paper. National Center for Atmospheric Research is sponsored by the NSF.

1. J. Sander, J. F. Eichner, E. Faust, M. Steuer, Rising variability in thunderstorm-related US losses as a reflection of changes in large-scale thunderstorm forcing. *Weather Clim. Soc.* **5**, 317–331 (2013).
2. A. B. Smith, R. W. Katz, US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. *Nat. Hazards* **67**, 387–410 (2013).
3. K. E. Trenberth, J. T. Fasullo, T. G. Shepherd, Attribution of climate extreme events. *Nat. Clim. Chang.* **5**, 725–730 (2015).
4. Engineering National Academies of Sciences and Medicine, *Attribution of Extreme Weather Events in the Context of Climate Change* (The National Academies Press, 2016).
5. Intergovernmental Panel on Climate Change, *Climate Change 2013—The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, New York, 2013).
6. A. J. Simmons *et al.*, Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Q. J. R. Meteorol. Soc.* **140**, 329–353 (2014).
7. M. G. Donat, L. V. Alexander, The shifting probability distribution of global daytime and night-time temperatures. *Geophys. Res. Lett.* **39**, L14707 (2012).
8. J. Hansen, M. Sato, R. Ruedy, Perception of climate change. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2415–E2423 (2012).
9. H. O. Pörtner *et al.*, *Climate Change 2022: Impacts, Adaptation and Vulnerability* (Intergovernmental Panel on Climate Change, 2022).
10. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer-Verlag, London, 2001).
11. O. Zolina *et al.*, Precipitation variability and extremes in Central Europe: New view from STAMMEX results. *Bull. Am. Meteorol. Soc.* **95**, 995–1002 (2014).
12. O. Zolina, C. Simmer, K. Belyaev, S. K. Gulev, P. Koltermann, Changes in the duration of European wet and dry spells during the last 60 years. *J. Clim.* **26**, 2022–2047 (2013).
13. J. Lehmann, D. Coumou, K. Frieler, Increased record-breaking precipitation events under global warming. *Clim. Change* **132**, 501–515 (2015).
14. S.-K. Min, X. Zhang, F. W. Zwiers, G. C. Hegerl, Human contribution to more-intense precipitation extremes. *Nature* **470**, 378–381 (2011).
15. K. Mahoney *et al.*, Climatology of extreme daily precipitation in Colorado and its diverse spatial and seasonal variability. *J. Hydrometeorol.* **16**, 781–792 (2015).
16. M. G. Donut *et al.*, Global land-based datasets for monitoring climatic extremes. *Bull. Am. Meteorol. Soc.* **94**, 997–1006 (2013).
17. J. Sillmann, V. V. Kharin, X. Zhang, F. W. Zwiers, D. Bronaugh, Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. D Atmospheres* **118**, 1716–1733 (2013).
18. A. G. Pendergrass, D. L. Hartmann, Changes in the distribution of rain frequency and intensity in response to global warming. *J. Clim.* **27**, 8372–8383 (2014).
19. S. C. Herring, M. P. Hoerling, J. P. Kossin, T. C. Peterson, P. A. Stott, Explaining extreme events of 2014 from a climate perspective. *Bull. Am. Meteorol. Soc.* **96**, S1–S172 (2015).
20. A. G. Pendergrass, D. L. Hartmann, Two modes of change of the distribution of rain. *J. Clim.* **27**, 8357–8371 (2014).
21. L. O. Mearns, R. W. Katz, S. H. Schneider, Extreme high-temperature events: Changes in their probabilities with changes in mean temperature. *J. Clim. Appl. Meteorol.* **23**, 1601–1613 (1984).
22. C. J. Paciorek, D. A. Stone, M. F. Wehner, Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Weather Clim. Extrem.* **20**, 69–80 (2018).
23. M. D. Risser, M. F. Wehner, Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. *Geophys. Res. Lett.* **44**, 12–457 (2017).
24. G. A. Meehl *et al.*, An introduction to trends in extreme weather and climate events: Observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. *Bull. Am. Meteorol. Soc.* **81**, 413–416 (2000).
25. S. Rahmstorf, D. Coumou, Increase of extreme events in a warming world. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17905–17909 (2011).
26. V. V. Kharin, F. W. Zwiers, X. Zhang, M. Wehner, Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim. Change* **119**, 345–357 (2013).
27. G. J. Van Oldenborgh *et al.*, Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environ. Res. Lett.* **12**, 124009 (2017).
28. S. Begueria *et al.*, Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: A case study in northeast Spain from 1930 to 2006. *Int. J. Climatol.* **31**, 2102–2114 (2011).
29. M. Roth, T. A. Buishand, G. Jongbloed, A. M. G. Klein Tank, J. H. Zanten, A regional peaks-over-threshold model in a nonstationary climate. *Water Resour. Res.* **48**, W11533 (2012).
30. B. Renard, X. Sun, M. Lang, “Bayesian methods for non-stationary extreme value analysis” in *Extremes in a Changing Climate*, A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, S. Soroshian, Eds. (Springer, 2013), pp. 39–95.
31. F. W. Zwiers, X. Zhang, Y. Feng, Anthropogenic influence on long return period daily temperature extremes at regional scales. *J. Clim.* **24**, 881–892 (2011).
32. S. Shenoy, D. Gorinevsky, “Risk adjusted forecasting of electric power load” in *2014 American Control Conference* (Institute of Electrical and Electronics Engineers, 2014), pp. 914–919.
33. S. Shenoy, D. Gorinevsky, “Predictive analytics for extreme events in big data” in *2015 IEEE First International Conference on Big Data Computing Service and Applications (BigDataService)* (Institute of Electrical and Electronics Engineers, 2015) pp. 184–193.
34. S. Shenoy, D. Gorinevsky, Estimating long tail models for risk trends. *IEEE Signal Process. Lett.* **22**, 968–972 (2015).
35. Y. Xia *et al.*, NLDAS Mosaic Land Surface Model L4 Hourly 0.125 × 0.125 degree V002. https://disc.gsfc.nasa.gov/datasets/NLDAS_MOS0125_H_002/summary. Accessed 22 August 2020.
36. Y. Xia *et al.*, Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res. D Atmospheres* **117**, D03109 (2012).
37. M. Kottek, J. Grieser, C. Beck, B. Rudolf, F. Rubel, World map of the Köppen-Geiger climate classification updated. *Meteorol. Z. (Berl.)* **15**, 259–263 (2006).
38. R. S. Vose *et al.*, NOAA’s merged land-ocean surface temperature analysis. *Bull. Am. Meteorol. Soc.*, **93**, 1677–1685, 2012.
39. M. G. Donat *et al.*, Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. D Atmospheres* **118**, 2098–2118 (2013).
40. K. E. Trenberth, A. Dai, R. M. Rasmussen, and D. B. Parsons, The changing character of precipitation. *Bull. Am. Meteorol. Soc.* **84**, 1205–1217 (2003).