Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL098855

Diverse Physical Processes Drive Upper-Tail Flood Quantiles in the US Mountain West

Guo Yu[1,2], **Daniel B. Wright**[1], **and Frances V. Davenport**[3,4]

[1]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA, [2]Division of Hydrologic Sciences, Desert Research Institute, Las Vegas, NV, USA, [3]Department of Earth System Science, Stanford University, Stanford, CA, USA, [4]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

Key Points:

- The prevalence of mixed flood populations and their implications for Flood Frequency Analysis
- Using estimated shape parameters to reflect differences in "tail behavior" among different flood types
- The importance of more physically informed approaches for predicting flood statistics in the region and beyond due to climate warming

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

G. Yu,
guo.yu@dri.edu

Citation:

Yu, G., Wright, D. B., & Davenport, F. V. (2022). Diverse physical processes drive upper-tail flood quantiles in the US mountain west. *Geophysical Research Letters*, *49*, e2022GL098855. https://doi.org/10.1029/2022GL098855

Received 24 MAR 2022
Accepted 3 MAY 2022

Author Contributions:

Conceptualization: Guo Yu, Daniel B. Wright
Data curation: Frances V. Davenport
Methodology: Guo Yu, Daniel B. Wright, Frances V. Davenport
Supervision: Daniel B. Wright
Writing – original draft: Guo Yu
Writing – review & editing: Daniel B. Wright, Frances V. Davenport

Abstract Floods in the mountainous western U.S. typically arise from rainfall, snowmelt, or rain-on-snow (ROS). This implies that streamflow records comprise "mixtures" of flood regimes with differing physical characteristics. This runs counter to the assumption, made in flood hazard practice, that flood observations are independent and identically distributed. In this study, we examine 308 watersheds and 281 (91%) of which exhibited substantial roles of at least two flood regimes, with rainfall and ROS accounting for the very largest floods. We demonstrate that flood mixtures can have dramatic impacts on upper-tail flood quantiles—rarer than the 0.01 annual exceedance probability. While the geographic distribution of mixed flood regimes is explained by climate and elevation, such regimes are subject to future change due to climate warming. The complexities brought about by diverse flood mixtures require process-based approaches for understanding and modeling future flood frequency distributions.

Plain Language Summary Floods in the mountainous western United States can be caused by various physical drivers such as rainfall, snowmelt, or their combination. This implies that statistics such as the so-called "100-year flood"—important for floodplain mapping and infrastructure design—oftentimes belie complex "mixtures" of different physical flood mechanisms. The existence of such mixtures confounds conventional analysis methods and assumptions. Results from our study of 308 watersheds suggest that neglect of mixture effects can lead to large uncertainties in the estimated magnitudes of extreme flood statistics in the region. Current mixtures will also change due to climate warming; floods in historically snowmelt-dominated watersheds, for example, may see an increasing importance of rainfall processes in the future. Our findings thus highlight the importance of more physically informed approaches for predicting flood statistics in the region and beyond.

1. Introduction

Riverine flooding can be triggered by a variety of distinct hydrometeorological drivers. In the eastern U.S., most floods are due to extreme tropical cyclone rainfall or springtime extratropical systems (Smith et al., 2011; Villarini & Smith, 2010). Looking further west, extratropical systems combined with summertime mesoscale convective systems make up the flood climatology in the midwestern U.S. (Villarini et al., 2011), while floods in the mountainous western United States can be caused by extreme rainfall (often from atmospheric rivers [ARs]; e.g., Barth et al., 2017), snowmelt, or their combination (Berghuijs et al., 2016; Davenport et al., 2020). Such "mixtures" of flood types associated with different physical drivers are also well-documented in Europe (Berghuijs et al., 2019; Blöschl et al., 2017, 2019) and likely exist in most terrestrial regions around the globe. Watersheds with these mixtures generally do not experience each type with equal frequency or severity (Smith et al., 2018).

Flood mixtures have two important implications. The first is related to estimation of extreme streamflow quantiles such as the 100-year average recurrence interval (ARI; corresponding to a 1% annual exceedance probability [AEP]). These and other quantiles from extreme streamflow distributions—derived via methods broadly referred to as flood frequency analysis (FFA)—are central to infrastructure design, dam safety analysis, and floodplain mapping (e.g., NRC, 1988, 1994; USBR, 2006, 2011). For mathematical convenience, FFA practices typically treat a sample of streamflow observations at a given site as independent and identically distributed (iid; e.g., Sivapalan & Samuel, 2009) regardless of whether the sample stems from one or multiple causes. The difficulty of mixed sample FFA was recognized decades ago in FFA guidelines ("Bulletin 17B"; ICWD, 1982), and a variety of "mixture distribution FFA" techniques have since appeared (e.g., Gotvald et al., 2012; Murphy, 2001; Parrett

et al., 2011; Waylen & Woo, 1982). To the best of our knowledge, however, these "remedies" have not achieved widespread use (England Jr. et al., 2018; Kjeldsen et al., 2008). It thus appears that more emphasis on mixtures is needed in flood research and practice.

The second implication of flood mixtures is related to climatic nonstationarity. The existence of mixtures suggests that attempts to identify historical trends or to estimate flood quantiles under changing conditions are unlikely to succeed if they cannot account for changes in flood "subsamples" associated with different physical drivers. Failure to do so may partially explain the lack of observational evidence for a climate-related increase in flooding (Sharma et al., 2018). While nonstationarity is not an explicit focus of this study, results presented below suggest that the prevalence of mixtures in the study region will have important implications for FFA in a changing climate.

While "methodological literature" on mixture distribution FFA abounds (e.g., Gotvald et al., 2012; Murphy, 2001; Waylen & Woo, 1982), few examples take a regional look at the prevalence of mixed flood populations and their FFA implications. This study aims to show both the regional prevalence of mixtures and what this means for upper tail flood quantiles. These aims share some similarity with Smith et al. (2011) and Barth et al. (2017). Compared with Smith et al. (2011), we bring a different geographic focus and explicitly investigate the effects of mix types on flood quantiles. While our study region largely overlaps with that of Barth et al. (2017)—who examined the role of ARs in flood quantiles—we consider a wider range of flood drivers. Although the objective of this study is not to provide practical methods for mixture distribution FFA, it adds to emerging research into process-based understanding and prediction of flood frequency. We also demonstrate the value of new datasets, including from large-scale land surface models (LSMs), to support such research.

2. Background—Floods in the US Mountain West

Winter stratiform precipitation—often associated with ARs (e.g., Gershunov et al., 2017; Ralph et al., 2006)—brings moisture to the U.S. west coast, where it often experiences orographic enhancement and snowpack accumulation (James & Houze, 2005). When rain falls on existing snowpack, it accelerates melt and can produce rain-on-snow (ROS) flooding (e.g., Marks et al., 1998). In the late spring and early summer, snowmelt and ROS are the main flood drivers for watersheds at high elevations (e.g., Northern Cascades and Sierra Nevada) and high latitudes (e.g., Montana; Berghuijs et al., 2016; McCabe et al., 2007). In the late summer, rainfall-driven floods occur across the semiarid southwest, often associated with the North American monsoon (e.g., Higgins et al., 1997; Li et al., 2003; Vivoni et al., 2006). Large-scale atmospheric circulation anomalies, oftentimes outside of the snowmelt season, account for some of the largest floods in the region (Hirschboeck, 1987; Maddox et al., 1979, 1980). For example, 13 out of the 21 very large (on a unit watershed area basis) floods in the conterminous US (CONUS) analyzed in Hirschboeck (1987) occurred in the west and were attributed to such phenomena. These include the well-known west coast "Christmas Flood" in 1964 (e.g., Fredriksen, 1965) and the 1976 Big Thompson flood in Colorado (e.g., Costa, 1978).

Though precipitation extremes have increased only modestly in the western US since 1950s (e.g., Karl et al., 2009; Wright et al., 2019), this trend is projected to continue with climate warming (e.g., Diffenbaugh et al., 2005; Dominguez et al., 2012; O'Gorman, 2015). Past warming has shifted the snowmelt season earlier, accompanied by declining melt rate (Fritze et al., 2011; Musselman et al., 2017). ROS events have become more frequent at high elevation due to higher freezing levels, but less frequent at low elevation as snowpack declines (Kampf & Lefsky, 2016; McCabe et al., 2007). Thus, the relative balance of different flood drivers has and will continue to shift, posing challenges to FFA and potentially causing important changes in flood distributions.

3. Data and Methods

3.1. Streamflow and Other Observational/Model Data

This study uses an updated version of the data set from Davenport et al. (2020), who identified rainfall-, snowmelt-, and ROS-driven streamflow observations in the western US. The longest streamflow records in the data set range from 1980 to 2020. Davenport et al. (2020) identified US Geological Survey daily streamflow observations that exceed the long-term median value and calculated preceding 8-day basin-averaged precipitation (NLDAS-2; Mitchell, 2004) and simulated snow water equivalent from the North American Land Data Assimilation System Variable Infiltration Capacity (NLDAS-VIC) land surface model (LSM) (Xia et al., 2012). To

ensure independence, Davenport et al. (2020) required at least 7 days between streamflow events, which then were classified as: (a) rainfall-driven (total rainfall > 10 mm and snowmelt < 5 mm), (b) snowmelt-driven (total rainfall < 5 mm and snowmelt > 10 mm) and (c) ROS when both rainfall and snowmelt exceed 5 mm. While the final Davenport et al. (2020) data set only includes 410 watersheds that exhibit multiple flood types, this includes virtually all higher-elevation watersheds (>500 m above sea level [masl]). Watersheds with a only a single type—not analyzed in Davenport et al. (2020) nor here—were rainfall-dominated and 65 of 69 are on the Pacific coast or at low elevation in Arizona. Thus, the overwhelming majority of watersheds outside of those two subregions exhibited multiple flood types according to the that study's criteria.

For each watershed, we selected the largest m streamflow events irrespective of flood type, as well as the largest n streamflow events of each type, where m (n) is the largest value between 25 (20) and the number of unique water years. The m events are henceforth referred to as single-sample floods while the n events are referred to by type. If the number of a particular type is less than 20, it is neglected because its distribution cannot be modeled well. In general, the magnitudes of flood types rejected due to limited sample sizes were much smaller than retained types at the same site (Figure S1 in Supporting Information S1). Nonetheless, rejection of flood types on the basis of sample size is a noteworthy limitation of our work and is discussed further in Section 5. Twenty eight sites with $m < 25$ were discarded. To further ensure that flood events are independent, we removed 74 sites that exhibited significant Pearson correlation (p-value < 0.05) between successive peaks. This left 308 watersheds with average sample sizes of 37 for the single-sample group, and 37, 35, and 36 for rainfall-, snowmelt- and ROS-driven groups, respectively. We also applied nonparametric bootstrapping to evaluate sample size effects (e.g., from 25 to 41) and obtained very similar shape parameters despite uncertainties in the estimated 100-year floods (Figures S2–S5 in Supporting Information S1). For the purposes of this study, that is, to understand the extent of mixture effects (e.g., shape parameters), these sample sizes appear to be sufficient. We do not necessarily recommend these results for usage in operational FFA.

It should be noted that most prior studies that consider mixture flood distributions (Barth et al., 2017; Gotvald et al., 2012; Smith et al., 2011; Villarini et al., 2011) separated annual instantaneous peak flows into flood types. This can seriously limit the number of observations for each subtype. Our approach based on continuous daily streamflow timeseries remedies this somewhat. To bridge this gap between daily and instantaneous peak flows, ratios between the two (where available) were calculated. These ratios decrease with watershed size and are (larger) smaller values for (rainfall) snowmelt floods (Figure S6 in Supporting Information S1). This latter fact implies that different flood types diverge more—and thus will exhibit stronger mixing effects—at the instantaneous scale than at the daily scale.

3.2. Flood Envelope Curves

Envelope curves depict the upper bound of regional streamflow observations (Costa, 1987; Enzel et al., 1993). We examined how the largest floods of each flood type vary with watershed scale by estimating envelope curves from the largest observations via (Fuller, 1914)

$$Q_i = \alpha_i A^{\theta_i} \tag{1}$$

where Q_i denotes the largest daily streamflow with respect to drainage area A and flood type i; α_i and θ_i are scaling intercept and power law coefficients, respectively.

3.3. Extreme Value and Mixture Modeling and Boundedness

We fitted Generalized Pareto distributions (GPD; e.g., Coles, 2001) to observations of different flood types, as well as the single-sample floods using L-Moments (Hosking & Wallis, 1997). The GPD's cumulative distribution function (CDF) is:

$$F_i\left(x_i; \mu_i, \sigma_i, \xi_i\right) = 1 - \left(1 + \frac{\xi_i(x_i - \mu_i)}{\sigma_i}\right)^{-\frac{1}{\xi_i}} \text{ for } \xi_i \neq 0 \tag{2}$$

where x_i denotes streamflow of flood type i, and μ_i, σ_i, and ξ_i are the threshold, scale, and shape parameters, respectively. μ_i and σ_i indicate the central tendency and variability, respectively, while ξ_i is indicative of skewness and the "thickness" of the GPD tail. Notably, in GPD and related distributions (e.g., the generalized extreme value

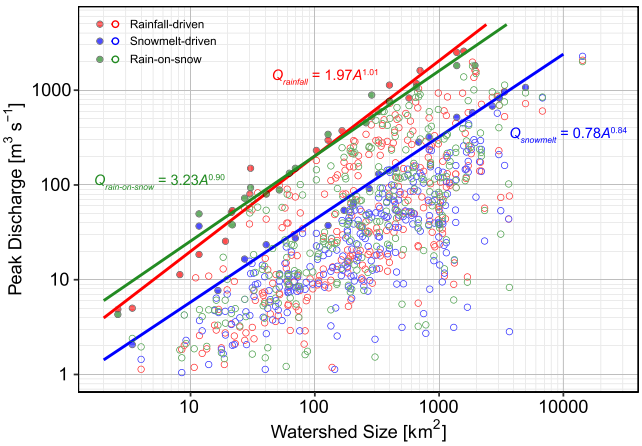

Figure 1. The largest rainfall-, snowmelt-, and ROS-driven flood peaks for each watershed and their corresponding envelope curves. Filled circles denote peaks used for envelope curve fitting.

distribution), $\xi_i > 0$ indicates that the upper tail is heavy and unbounded, meaning that there is nonzero probability density as $x_i \to \infty$. $\xi_i < 0$, in contrast, implies an upper bound to the distribution, while $\xi_i = 0$ indicates a thin tail. Estimated shape parameters have been widely used in the study of extreme events such as rainfall, floods, and water vapor transport (e.g., Villarini & Smith, 2010; Su & Smith, 2021). In this study, record length thresholds described in Section 3.1 provided μ_i; this approach has been used in prior research (e.g., Papalexiou & Montanari, 2019) and practice (e.g., Bonnin et al., 2004; Perica et al., 2018).

Having ensured independence of flood events, we also derived "mixture distributions" by taking the product of CDFs of different flood types for watersheds exhibiting multiple flood types (Nadarajah, 2008; Waylen & Woo, 1982):

$$F_{\text{mixture}}(x) = F_{\text{rainfall}}(x) * F_{\text{snowmelt}}(x) * F_{\text{ROS}}(x) \quad (3)$$

If only rainfall and snowmelt peaks are considered for a watershed, for example, the cumulative distribution function (CDF) of its mixture distribution is the product of F_{snowmelt} and F_{rainfall}. We also use "upper tail flood type" to describe the type that generates the highest 100- to 500-year ARIs. To partly account for the distribution uncertainty, we have also repeated analyses using the log-Pearson type III (LP3) distribution (Equations S1–S3 in Supporting Information S1; e.g., Asquith et al., 2017).

4. Results

4.1. Envelope Curves of Different Flood Types

The largest rainfall- and ROS-driven flood peaks are similar in magnitude and are substantially larger than snowmelt floods (Figure 1). At larger scales (>2,000 km^2), however, there are few rainfall or ROS flood events that exceed the snowmelt envelope curve. This is likely due to rainfall partial coverage limits the magnitude of rainfall-driven and ROS-driven floods in large watersheds.

The pronounced differences between flood types evident in Figure 1 indicate that extreme floods are almost always tied to rainfall. This result is in line with the fact that snowmelt rates are constrained by the available energy and tend to be smaller than extreme rainfall rates (e.g., Jarrett, 1989; Jarrett & Costa, 1988; Kampf & Lefsky, 2016). For example, the largest NLDAS-2 single grid cell daily rainfall depth within the study region was 415 mm, far exceeding the largest NLDAS-VIC daily snowmelt of 285 mm. Nonetheless, these envelope curves conceal important aspects of flood behavior linked to geography and elevation.

4.2. Mixture Distributions and Their Prevalence

To illustrate the concept of mixture distributions and how we calculated them, we show examples from Gallatin River near Gallatin Gateway, MT and Chiwawa River near Plain, WA (Figure 2). Though both watersheds exhibit three flood types (Figures 2a and 2b), the associated GPDs are quite distinct (Figures 2c and 2d). In both watersheds, mixture distributions resemble the single-sample GPDs for ARIs smaller than 10–20 years, but much higher estimates for the upper tail (ARI ≥ 100 years; Figures 2c and 2d). Similarly, estimated ARIs show substantial differences depending on whether mixture distributions or single-sample GPDs are used. For instance, the average recurrence interval (ARI) of a 300 m^3 s^{-1} event in Gallatin River is estimated to be ~70 years using the mixture distribution, but over 500 years using the single sample GPD.

Snowmelts dominate the tail of the mixture distribution for the higher elevation Gallatin River (2,400 masl; Figure 2c), while rainfall-driven peaks dominate the tail for Chiwawa River (1,400 masl; Figure 2d). Crucially, single-sample GPDs for these sites fail to capture the upper tail flood distribution behavior.

Among the 308 watersheds in this study, 27 (9%) exhibit a single flood type (i.e., rainfall-driven), while 82 (27%) and 199 (64%) exhibit mixtures of two and three flood types, respectively. GPD shape parameters for the rainfall

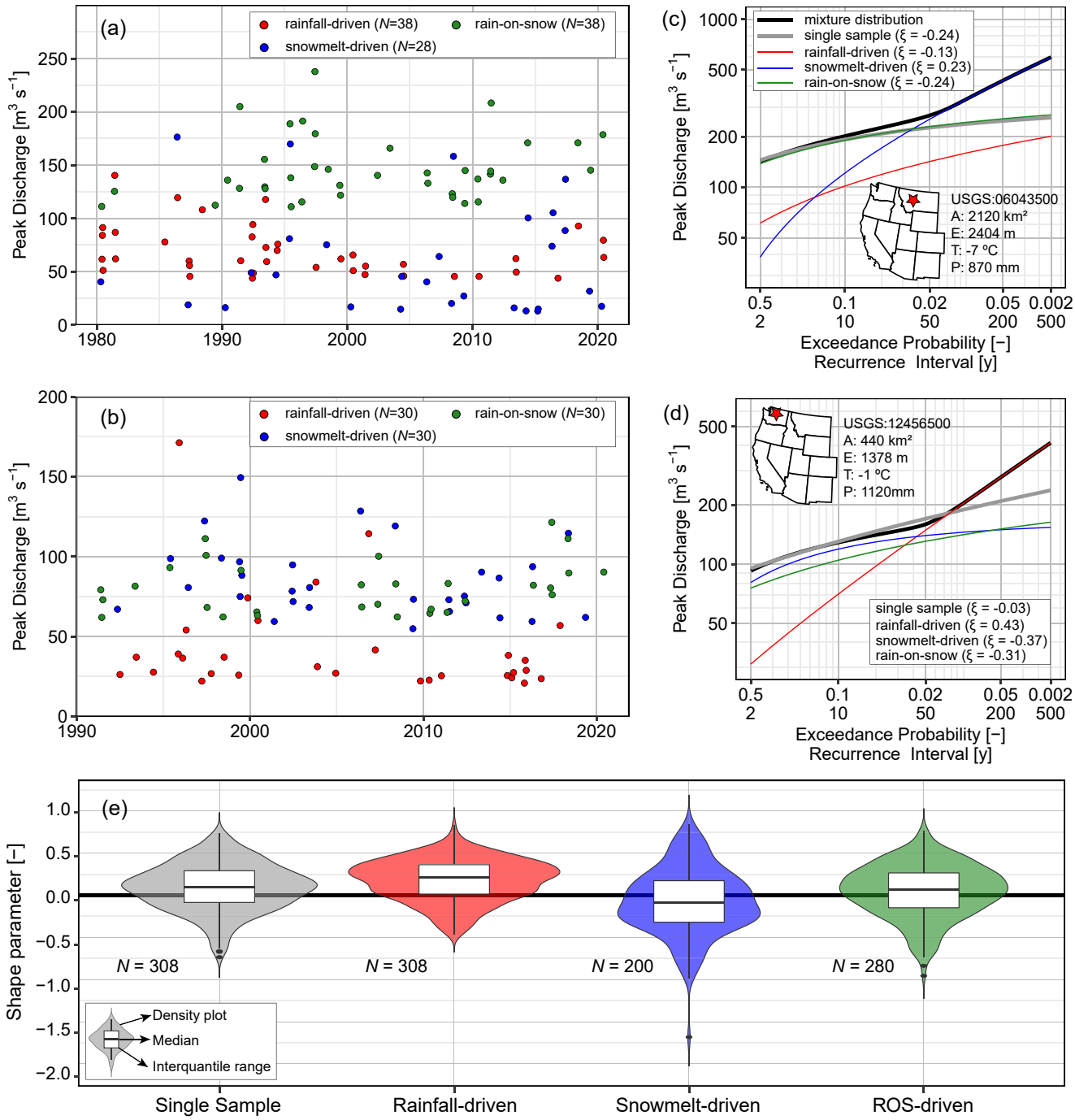

Figure 2. The selected peaks of three different flood groups for (a) Gallatin River near Gallatin Gateway, MT (USGS:06043500) and (b) Chiwawa River near Plain, WA (USGS:12456500). Genearlized Pareto Distributions (GPDs) for different groups, along with mixture distributions for (c) Gallatin and (d) Chiwawa. Inset maps in (c and d) roughly show watershed locations. (e) Violin plots of all watersheds' single sample and distinct type GPD shape parameters.

peaks are generally positive while snowmelt values are negative (Figure 2e). Log-Pearson type III (LP3) shape parameters show similar patterns (Figure S7 in Supporting Information S1). These findings are consistent with the envelope curves in Figure 1: the largest rainfall-driven floods are much more severe than the snowmelt-driven events except at the largest watershed scales.

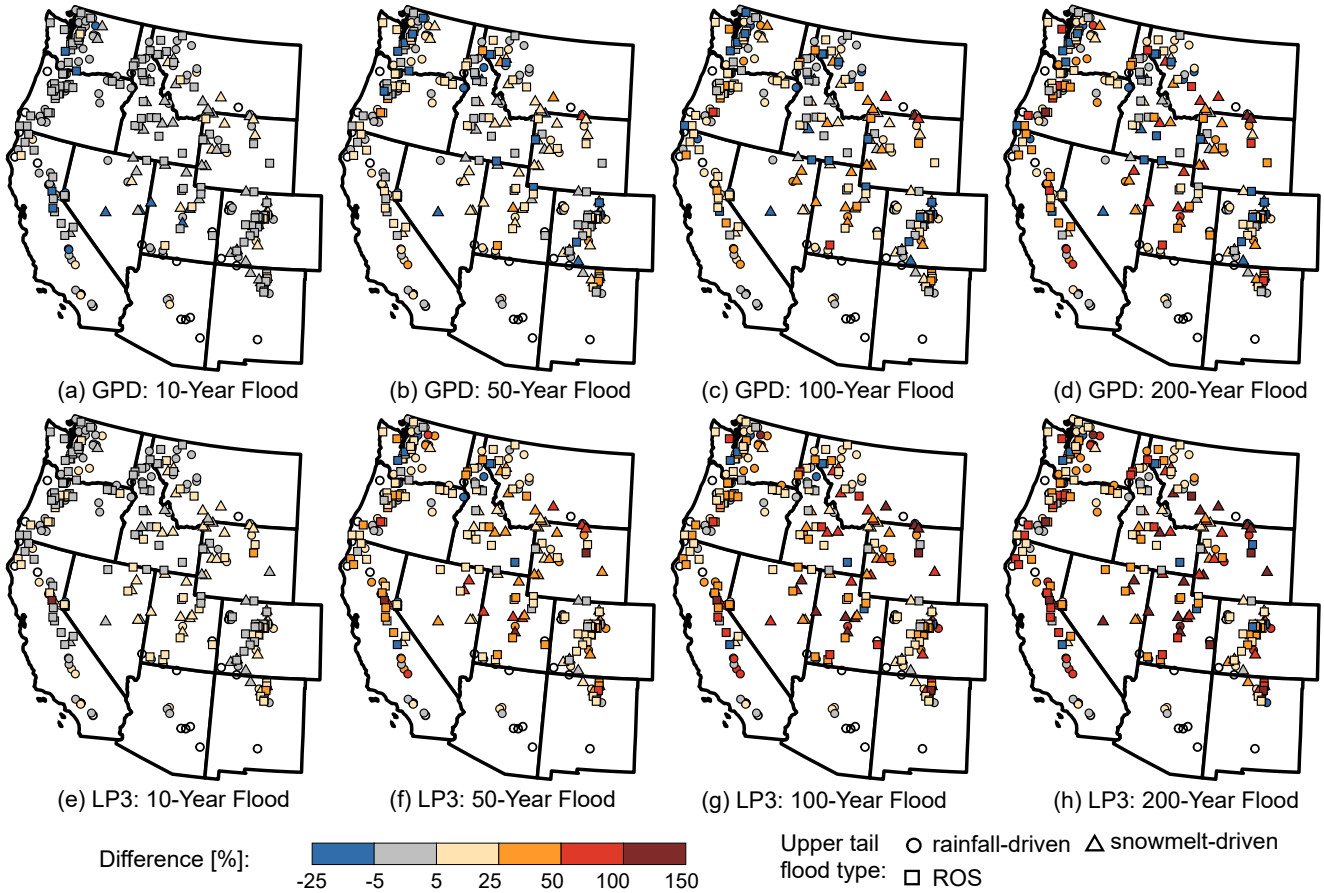

Figure 3. Percent differences for (a and e) 10-year, (b and f) 50-year, (c and g) 100-year and (d and h) 200-year floods between single-sample and mixture distribution methods. (a–d) and (e–h) are based on Generalized Pareto distributions (GPD) and log-Pearson type III (LP3) distributions, respectively. Unfilled circles represent the watersheds that only had sufficient data to model a single flood type.

4.3. Importance of Mixture Distributions in Flood Quantiles

We derived both single-sample GPD/LP3 and mixture distributions for all but 19 sites—denoted with unfilled circles in Figure 3—which had sufficient records only to derive rainfall-driven distributions. Consistent with examples in the previous subsection, quantile estimates differ markedly at high ARIs. For 10-year floods, for example, differences between the two distributions are negligible (generally $<\pm5\%$; Figure 3a) while for 200-year floods, the mixture distribution-based estimates are 50%–150% higher than the single-sample GPD-based values for most watersheds (Figure 3d). LP3-based results show even larger percent differences, indicating stronger "mixing effects" (Figures 3e–3h).

The upper tail flood type shows geographic patterns as rainfall- and snowmelt floods dominate in watersheds along the Pacific Coast and the Rocky Mountains, respectively (Figure 3). ROS upper tail flood types can be found throughout the study region. Irrespective of upper tail flood type, the mixture distributions yielded $>+5\%$ higher estimates than single sample method for 100- and 200-year floods for a majority (57% using GPD and 74% using LP3) of watersheds. In these watersheds, single-sample shape parameters are smaller than the values from at least one individual flood type (Figure S8a in Supporting Information S1).

Forty-six (seventeen) watersheds show negative percent differences ($<-5\%$) even for 200-year floods between using two methods, meaning the single-sample GPD (LP3) yields higher estimates (dark blue symbols in Figure 3). In these watersheds, shape parameters for single-sample distributions are generally larger than the corresponding values from any individual flood type (Figure S8b in Supporting Information S1). This physically-unreasonable result is due to the statistical artifact known as "skew separation" which results in increased skewness when mixing statistically different samples (Dawdy & Gupta, 1995; see Section 5 for a brief explanation).

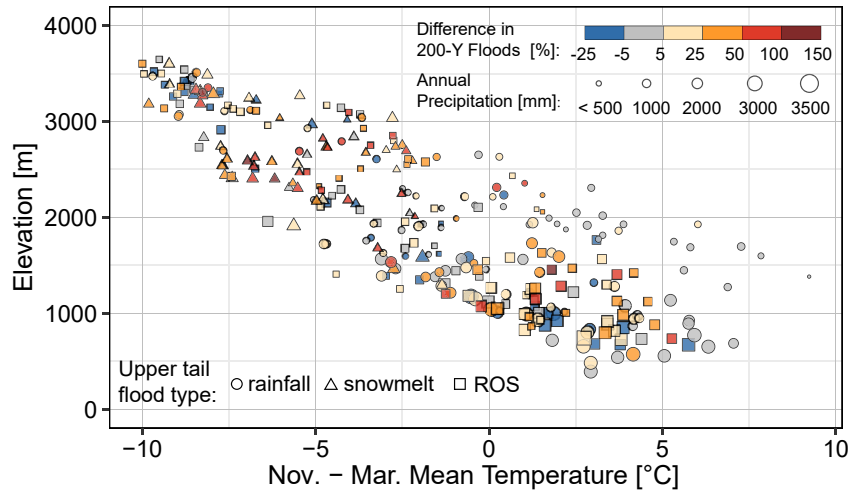

Figure 4. The relationship between the percent differences in 200-year floods and their explanatory variables, including winter mean temperature, annual precipitation, and watershed elevations.

Eighty five (sixty two) watersheds showed comparable estimates (<5% difference) between the single-sample GPD/LP3 and mixture distributions (gray symbols in Figures 3d and 3h). This can occur in two ways: (a) GPD (LP3) of the three (or two) individual flood types can be similar, or (b) floods of one type are always larger than the those of the other types. An example of the former is the Thompson River near Thompson Falls, MT (USGS: 12389500); an example of the later is the Sauk River near Darrington, WA (USGS: 12186000), where rainfall floods are always higher than the snowmelt and ROS events (Figure S9 in Supporting Information S1).

4.4. Linkage of Watershed Characteristics to Flood Quantile Differences

We investigated differences in 200-year floods between single-sample and mixture distributions from Figure 3d with respect to elevation, basin-averaged annual precipitation, and winter mean temperature. Annual precipitation and winter mean temperature decrease with elevation across the western US (Figure 4). Large differences in 200-year floods are prevalent across climate and elevation regimes, except for relatively warm watersheds (winter mean temperature >5°C) where snow rarely accumulates and where rainfall dominates both single-sample and mixture distributions. For higher and colder watersheds, snowmelt events yield higher rare quantiles (e.g., Figure 2c); for lower and warmer watersheds, rainfall and ROS events dominate the tail (e.g., Figures 2d and 4). Rainfall-driven floods can also dominate the tail distribution even for watersheds at high elevation (>3,000 m), consistent with previous research showing that rainfall-driven floods can occur at such elevations (especially in the intermittent snow zone; e.g., Kampf & Lefsky, 2016; Mahoney et al., 2015).

5. Discussion and Conclusions

This study examines the behavior of mixture flood populations and its impacts on upper tail distributions in the western US. Here, we discuss limitations of our work as well as implications for FFA practice. We reiterate that this study does not propose a method that can replace current operational FFA techniques, for example, Bulletin 17C (England et al., 2018), but rather to highlight the potential need for such a replacement.

Envelope curves show that the largest flood events in the region are almost entirely associated with rainfall (including ROS) associated with anomalous atmospheric circulations (Gochis et al., 2015; Hirschboeck, 1987; Maddox et al., 1980); the largest snowmelt events are roughly four times smaller. 91% of the study watersheds exhibited relatively large samples (n > 20 for the 1980–2020 period) of least two flood types. Distribution tail behavior further highlights key differences among flood types. Positive GPD/LP3 shape parameters, which indicate the potential for rare but very large floods, are much more common for rainfall and ROS samples than for snowmelt samples. The latter type has, on average, negative shape parameters, indicating an upper bound in flood magnitude from snowmelt tied to limited daytime net radiation.

A total of 223 (246) sites showed undesirable over- or under-estimation in GPD (LP3)-based flood quantiles resulting from neglect of mixture effects. Our results indicate that the influence of mixtures on flood frequency is most prevalent in the upper tail, that is, above the 100-year average recurrence interval (ARI). Lower return periods are much less affected. In more than half of sites (57% using GPD and 74% using LP3), mixture distributions resulted in 200-year floods that are at least 5% larger—and much larger, in many cases—than the results of single sample distributions (Figures 3d and 3h). In a warming future, watersheds at lower elevations will experience less snowpack and thus reduced frequency of ROS events (e.g., Huang et al., 2018; Musselman et al., 2018), which will potentially shift ROS-dominated tails to rainfall-dominated. Meanwhile, high-elevation watersheds will see more precipitation falling as rain rather than snow (e.g., Freudiger et al., 2014; Fritze et al., 2011), which may change snowmelt-dominated tails to ROS- or rainfall-dominated. These projected changes imply further divergence between snowmelt- and rainfall-driven flood distributions.

A further 46 (17) watersheds show 200-year floods greater than 5% lower using mixture distributions than using the single sample GPD (LP3). These are attributable to "skew separation" (Matalas et al., 1975)—the artificial increase in skewness that results when two or more samples from different populations are mixed. An example of this can be seen in Figure S10 in Supporting Information S1. Dawdy and Gupta (1995) showed that skew separation can result from heterogeneity in flood generating mechanisms.

Taken together, the prevalence of "mixture distributions" and disparities in streamflow magnitudes associated with different types of floods seriously undercut the independent and identically distributed (iid) assumption that underpins much of FFA practice. Nonetheless, we take time here to highlight six limitations of our study: (a) we used a relatively small sample sizes (ranging from 20 to 41). While such samples are admittedly small to accurately capture tail behavior, sensitivity analysis showed relatively limited and predictable quantile estimation behavior as a function of sample size (Figures S2–S5 in Supporting Information S1). Nonetheless, sample size issues should probably preclude usage of our results for decision-making purposes. (b) Rainfall and ROS floods in this study could have been further divided into subtypes according to their rainfall generating mechanisms (e.g., ARs vs. non-ARs; Barth et al., 2017) though at the expense of even smaller sample sizes. (c) A flood type was only included in mixture analysis if sufficient samples were available. This could in theory lead to the exclusion of anomalously large events for certain sites, if that event's type is infrequent. Due to the inclusion of the rainfall type at all sites and the bounded behavior of snowmelt floods, ROS is the only type subject to this limitation in our study region; 27 sites (9%) excluded the ROS type. (d) We neglected the role of "low floods" in arid watersheds and can color FFA results (e.g., Cohn et al., 2013)—though our usage of peaks-over-threshold observations rather than annual maxima may mitigate this issue to some extent. (e) Our usage of daily streamflow records differs from most FFA applications, which use instantaneous records. As argued in Section 3.1 (supported by Figure S6 in Supporting Information S1), mixture effects are likely stronger at the instantaneous timescale, meaning our findings likely understate the importance of mixing for such FFA applications. (f) We did not assess the impacts of temporal trends in flood types on mixture results. The existence of trends would complicate analyses further, but cannot be ruled out given observed water cycle changes in the region.

This study, along with Smith et al. (2011) and Barth et al. (2017), highlights the existence of mixed flood samples stemming from a wide variety of hydrometeorological drivers (e.g., snowmelt, ROS, and rainfall from ARs, tropical cyclones, and other storm systems). Our results imply a widespread violation of the conventional FFA independent and identically distributed (iid) assumption in our study region; further work is needed to see whether such independent and identically distributed (iid violations are common elsewhere. In addition, projected climate warming in the western US raises questions as to whether past flood observations and FFA results and methods will remain valid in the future. We believe that more explicitly process-based approaches have much to offer (see Sivapalan & Samuel, 2009; Wright et al., 2020 for expanded arguments). Here, we use large-scale land surface models (LSMs) simulations to help separate flood samples into distinct groups to study mixture distributions. Land surface models (LSMs) and land-atmosphere reanalyzes could be used to extend this or other analyses into a longer-term and geographically broader investigation of mixture flood distributions. Even more physically rooted FFA approaches such as Yu et al. (2019, 2020, 2021), which resolve probable combinations of different hydrometeorological drivers within physically-based numerical model simulations, can also provide insights and test hypotheses about the connections between mixed flood regimes, flood frequency, and how these are changing in a warming climate.

Data Availability Statement

The flood data used in this study are available at: https://doi.org/10.5281/zenodo.5645525.

Acknowledgments
GY's and DBW's contributions were supported by the U.S. National Science Foundation Hydrologic Sciences Program (award number EAR-1749638). GY was also partly supported by the Maki Postdoctoral fellowship by the Desert Research Institute. FD's contributions were supported by Stanford University.

References

Asquith, W. H., Kiang, J. E., & Cohn, T. A. (2017). Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities (USGS Numbered Series No. 2017–5038). *Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities* (Vols. 2017–5038). U.S. Geological Survey. https://doi.org/10.3133/sir20175038

Barth, N. A., Villarini, G., Nayak, M. A., & White, K. (2017). Mixed populations and annual flood frequency estimates in the Western United States: The role of atmospheric rivers. *Water Resources Research*, *53*(1), 257–269. https://doi.org/10.1002/2016WR019064

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The relative importance of different flood-generating mechanisms across Europe. *Water Resources Research*, 2019WR024841. https://doi.org/10.1029/2019WR024841

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States: Flood mechanisms across the U.S. *Geophysical Research Letters*, *43*(9), 4382–4390. https://doi.org/10.1002/2016GL068070

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., et al. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351), 588–590. https://doi.org/10.1126/science.aan2506

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., et al. (2019). Changing climate both increases and decreases European river floods. *Nature*, *573*(7772), 108–111. https://doi.org/10.1038/s41586-019-1495-6

Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M., & Riley, D. (2004). *Precipitation-frequency atlas of the United States. Volume 2 version 3.0*. Retrieved from https://repository.library.noaa.gov/view/noaa/22610

Cohn, T. A., England, J. F., Berenbrock, C. E., Mason, R. R., Stedinger, J. R., & Lamontagne, J. R. (2013). A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series. *Water Resources Research*, *49*(8), 5047–5058. https://doi.org/10.1002/wrcr.20392

Coles, S. G. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Costa, J. E. (1978). Colorado Big Thompson flood: Geologic evidence of a rare hydrologic event. *Geology*, *6*(10), 617–620. https://doi.org/10.1130/0091-7613(1978)6<617:CBTFGE>2.0.CO;2

Costa, J. E. (1987). Hydraulics and basin morphometry of the largest flash floods in the conterminous United States. *Journal of Hydrology*, *93*(3), 313–338. https://doi.org/10.1016/0022-1694(87)90102-8

Davenport, F. V., Herrera-Estrada, J. E., Burke, M., & Diffenbaugh, N. S. (2020). Flood size increases nonlinearly across the Western United States in response to lower snow-precipitation ratios. *Water Resources Research*, *56*(1). https://doi.org/10.1029/2019WR025571

Dawdy, D. R., & Gupta, V. K. (1995). Multiscaling and skew separation in regional floods. *Water Resources Research*, *31*(11), 2761–2767. https://doi.org/10.1029/95WR02078

Diffenbaugh, N. S., Pal, J. S., Trapp, R. J., & Giorgi, F. (2005). Fine-scale processes regulate the response of extreme events to global climate change. *Proceedings of the National Academy of Sciences*, *102*(44), 15774–15778. https://doi.org/10.1073/pnas.0506042102

Dominguez, F., Rivera, E., Lettenmaier, D. P., & Castro, C. L. (2012). Changes in winter precipitation extremes for the Western United States under a warmer climate as simulated by regional climate models: Precipitation extremes Western US. *Geophysical Research Letters*, *39*(5). https://doi.org/10.1029/2011GL050762

England, J. F., Jr., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas, W. O., Jr., Veilleux, A. G., et al. (2018). *Guidelines for determining flood flow frequency—Bulletin 17C* (Report No. 4-B5) (p. 168). https://doi.org/10.3133/tm4B5

Enzel, Y., Ely, L. L., House, P. K., Baker, V. R., & Webb, R. H. (1993). Paleoflood evidence for a natural upper bound to flood magnitudes in the Colorado River Basin. *Water Resources Research*, *29*(7), 2287–2297. https://doi.org/10.1029/93WR00411

Fredriksen, R. L. (1965). *Christmas storm damage on the HJ andrews experimental forest*. Pacific Northwest Forest and Range Experiment Station, US Department of Agriculture.

Freudiger, D., Kohn, I., Stahl, K., & Weiler, M. (2014). Large-scale analysis of changing frequencies of rain-on-snow events with flood-generation potential. *Hydrology and Earth System Sciences*, *18*(7), 2695–2709. https://doi.org/10.5194/hess-18-2695-2014

Fritze, H., Stewart, I. T., & Pebesma, E. (2011). Shifts in Western North American snowmelt runoff regimes for the recent warm decades. *Journal of Hydrometeorology*, *12*(5), 989–1006. https://doi.org/10.1175/2011JHM1360.1

Fuller, W. E. (1914). Flood flows. *Transactions of the American Society of Civil Engineers*, *77*(1), 564–617.

Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A., & Rutz, J. J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting Western North America. *Geophysical Research Letters*, *44*(15), 7900–7908. https://doi.org/10.1002/2017GL074175

Gochis, D., Schumacher, R., Friedrich, K., Doesken, N., Kelsch, M., Sun, J., et al. (2015). The great Colorado flood of September 2013. *Bulletin of the American Meteorological Society*, *96*(9), 1461–1487. https://doi.org/10.1175/BAMS-D-13-00241.1

Gotvald, A. J., Barth, N. A., Veilleux, A. G., & Parrett, C. (2012). *Methods for determining magnitude and frequency of floods in California, based on data through water year 2006*. US Geological Survey.

Higgins, R. W., Yao, Y., & Wang, X. L. (1997). Influence of the North American monsoon system on the U.S. Summer precipitation regime. *Journal of Climate*, *10*, 23.

Hirschboeck, K. K. (1987). Catastrophic flooding and atmospheric circulation anomalies. In *Catastrophic flooding*. Routledge.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press. https://doi.org/10.1017/CBO9780511529443

Huang, X., Hall, A. D., & Berg, N. (2018). Anthropogenic warming impacts on today's Sierra Nevada snowpack and flood risk. *Geophysical Research Letters*, *45*(12), 6215–6222. https://doi.org/10.1029/2018GL077432

Interagency Committee on Water & Data. (1982). *Guidelines for determining flood flow frequency*. (Technical Report). Bulletin 17B.report.

James, C. N., & Houze, R. A. (2005). Modification of precipitation by coastal orography in storms crossing Northern California. *Monthly Weather Review*, *133*(11), 3110–3131. https://doi.org/10.1175/MWR3019.1

Jarrett, R. (1989). Hydrology and paleohydrology used to improve the understanding of flood hydrometeorology in Colorado. *Design of Hydraulic Structres*.

Jarrett, R., & Costa, J. (1988). *Evaluation of the flood hydrology in the Colorado front range using precipitation, streamflow, and paleoflood data for the Big Thompson River basin*. https://doi.org/10.3133/wri874117

Kampf, S. K., & Lefsky, M. A. (2016). Transition of dominant peak flow source from snowmelt to rainfall along the Colorado Front Range: Historical patterns, trends, and lessons from the 2013 Colorado Front Range floods. *Water Resources Research*, *52*(1), 407–422. https://doi.org/10.1002/2015WR017784

Karl, T. R., Melillo, J. M., Peterson, T. C., & Hassol, S. J. (2009). *Global climate change impacts in the United States: A state of knowledge report*. Cambridge University Press.

Kjeldsen, T. R., Jones, D. A., & Bayliss, A. C. (2008). *Improving the FEH statistical procedures for flood frequency estimation*. Environment Agency.

Li, J., Maddox, R. A., Gao, X., Sorooshian, S., & Hsu, K. (2003). A numerical investigation of storm structure and evolution during the July 1999 Las Vegas flash flood. *Monthly Weather Review*, *131*(9), 2038–2059. https://doi.org/10.1175/1520-0493(2003)131<2038:anioss>2.0.co;2

Maddox, R. A., Canova, F., & Hoxit, L. R. (1980). Meteorological characteristics of flash flood events over the Western United States. *Monthly Weather Review*, *108*(11), 1866–1877. https://doi.org/10.1175/1520-0493(1980)108<1866:MCOFFE>2.0.CO;2

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

Mahoney, K., Ralph, F. M., Wolter, K., Doesken, N., Dettinger, M., Gottas, D., et al. (2015). Climatology of extreme daily precipitation in Colorado and its diverse spatial and seasonal variability. *Journal of Hydrometeorology*, *16*(2), 781–792. https://doi.org/10.1175/JHM-D-14-0112.1

Marks, D., Kimball, J., Tingey, D., & Link, T. (1998). The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: A case study of the 1996 Pacific northwest flood. *Hydrological Processes*, *12*(10–11), 1569–1587. https://doi.org/10.1002/(sici)1099-1085(199808/09)12:10/11<1569::aid-hyp682>3.0.co;2-l

Matalas, N. C., Slack, J. R., & Wallis, J. R. (1975). Regional skew in search of a parent. *Water Resources Research*, *11*(6), 815–826. https://doi.org/10.1029/WR011i006p00815

McCabe, G. J., Clark, M. P., & Hay, L. E. (2007). Rain-on-snow events in the Western United States. *Bulletin of the American Meteorological Society*, *88*(3), 319–328. https://doi.org/10.1175/BAMS-88-3-319

Mitchell, K. E. (2004). The multi-institution North American land data assimilation system (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, *109*(D7), D07S90. https://doi.org/10.1029/2003JD003823

Murphy, P. J. (2001). Evaluation of mixed-population flood-frequency analysis. *Journal of Hydrologic Engineering*, *6*(1), 62–70. https://doi.org/10.1061/(ASCE)1084-0699

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change*, *7*(3), 214–219. https://doi.org/10.1038/nclimate3225

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., et al. (2018). Projected increases and shifts in rain-on-snow flood risk over Western North America. *Nature Climate Change*, *8*(9), 808–812. https://doi.org/10.1038/s41558-018-0236-4

Nadarajah, S. (2008). On the product of generalized Pareto random variables. *Applied Economics Letters*, *15*(4), 253–259. https://doi.org/10.1080/13504850500425378

National Research Council (NRC). (1988). *Estimating probabilities of extreme floods: Methods and recommended research*. The National Academies Press. https://doi.org/10.17226/18935

National Research Council (NRC). (1994). *Estimating bounds on extreme precipitation events: A brief assessment*. The National Academies Press. https://doi.org/10.17226/9195

O'Gorman, P. A. (2015). Precipitation extremes under climate change. *Current Climate Change Reports*, *1*(2), 49–59. https://doi.org/10.1007/s40641-015-0009-3

Papalexiou, S. M., & Montanari, A. (2019). Global and regional increase of precipitation extremes under global warming. *Water Resources Research*, *55*(6), 4901–4914. https://doi.org/10.1029/2018WR024067

Parrett, C., Veilleux, A., Stedinger, J. R., Barth, N. A., Knifong, D. L., & Ferris, J. C. (2011). *Regional skew for California, and flood frequency for selected sites in the Sacramento-San Joaquin River Basin, based on data through water year 2006*. U. S. Geological Survey.

Perica, S., Pavlovic, S., St. Laurent, M., Trypaluk, C., Unruh, D., & Wilhite, O. (2018). *Precipitation-frequency atlas of the United States volume 11. Silver spring*. National Weather Service, National Oceanic and Atmospheric Administration. Retrieved from https://www.weather.gov/media/owp/oh/hdsc/docs/Atlas14_Volume11.pdf

Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R., & White, A. B. (2006). Flooding on California's Russian river: Role of atmospheric rivers. *Geophysical Research Letters*, *33*(13), L13801. https://doi.org/10.1029/2006GL026689

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If Precipitation extremes are increasing, why aren't floods? *Water Resources Research*, *54*(11), 8545–8551. https://doi.org/10.1029/2018WR023749

Sivapalan, M., & Samuel, J. M. (2009). Transcending limitations of stationarity and the return period: Process-based approach to flood estimation and risk assessment. *Hydrological Processes*, *23*(11), 1671–1675. https://doi.org/10.1002/hyp.7292

Smith, J. A., Cox, A. A., Baeck, M. L., Yang, L., & Bates, P. (2018). Strange floods: The upper tail of flood peaks in the United States. *Water Resources Research*, *54*(9), 6510–6542. https://doi.org/10.1029/2018WR022539

Smith, J. A., Villarini, G., & Baeck, M. L. (2011). Mixture distributions and the hydroclimatology of extreme rainfall and flooding in the Eastern United States. *Journal of Hydrometeorology*, *12*(2), 294–309. https://doi.org/10.1175/2010JHM1242.1

Su, Y., & Smith, J. A. (2021). An atmospheric water balance perspective on extreme rainfall potential for the contiguous US. *Water Resources Research*, *57*(4). https://doi.org/10.1029/2020WR028387

U.S. Bureau of Reclamation (USBR). (2006). *Guidelines for evaluating hydrologic hazards*. (Technical Report) (p. 83).report.

U.S. Bureau of Reclamation (USBR). (2011). *Dam safety public protection guidelines (A risk framework to support dam safety decision-making)*.

Villarini, G., & Smith, J. A. (2010). Flood peak distributions for the eastern United States: Eastern United States floods. *Water Resources Research*, *46*(6). https://doi.org/10.1029/2009WR008395

Villarini, G., Smith, J. A., Baeck, M. L., & Krajewski, W. F. (2011). Examining flood frequency distributions in the midwest U.S. *JAWRA Journal of the American Water Resources Association*, *47*(3), 447–463. https://doi.org/10.1111/j.1752-1688.2011.00540.x

Vivoni, E. R., Bowman, R. S., Wyckoff, R. L., Jakubowski, R. T., & Richards, K. E. (2006). Analysis of a monsoon flood event in an ephemeral tributary and its downstream hydrologic effects: Monsoon flood event and its downstream effects. *Water Resources Research*, *42*(3). https://doi.org/10.1029/2005WR004036

Waylen, P., & Woo, M. (1982). Prediction of annual floods generated by mixed processes. *Water Resources Research*, *18*(4), 1283–1286. https://doi.org/10.1029/WR018i004p01283

Wright, D. B., Bosma, C. D., & Lopez-Cantu, T. (2019). U.S. hydrologic design standards insufficient due to large increases in frequency of rainfall extremes. *Geophysical Research Letters*, *46*(14), 8144–8153. https://doi.org/10.1029/2019GL083235

Wright, D. B., Yu, G., & England, J. F. (2020). Six decades of rainfall and flood frequency analysis using stochastic storm transposition: Review, progress, and prospects. *Journal of Hydrology*, *585*, 124816. https://doi.org/10.1016/j.jhydrol.2020.124816

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (2012). 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. *Journal of Geophysical Research*, *117*(D3). https://doi.org/10.1029/2011JD016048

Yu, G., Wright, D. B., & Holman, K. D. (2021). Connecting hydrometeorological processes to low-probability floods in the mountainous Colorado front range. *Water Resources Research*, *57*(4), e2021WR029768. https://doi.org/10.1029/2021WR029768

Yu, G., Wright, D. B., & Li, Z. (2020). The upper tail of precipitation in convection-permitting regional climate models and their utility in nonstationary rainfall and flood frequency analysis. *Earth's Future*, *8*(10), e2020EF001613. https://doi.org/10.1029/2020EF001613

Yu, G., Wright, D. B., Zhu, Z., Smith, C., & Holman, K. D. (2019). Process-based flood frequency analysis in an agricultural watershed exhibiting nonstationary flood seasonality. *Hydrology and Earth System Sciences*, *23*(5), 2225–2243. https://doi.org/10.5194/hess-23-2225-2019