Smith, J., Cox, A., Baeck, M. L., Yang, L., & Bates, P. (2018). Strange Floods: The Upper Tail of Flood Peaks in the United States. *Water Resources Research*. https://doi.org/10.1029/2018WR022539

Publisher's PDF, also known as Version of record

Link to published version (if available):
10.1029/2018WR022539

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via AGU at https://doi.org/10.1029/2018WR022539 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE

10.1029/2018WR022539

Key Points:

- The upper tail of flood peaks in the United States is examined through analyses of the upper tail ratio for annual flood peak observations
- Two important aspects of strange floods in the United States are location near mountainous terrain and intense thunderstorms as causative agents
- A fundamental distinction between record floods in the United States and the broader population of annual flood peaks is their seasonal distribution

Correspondence to:
J. A. Smith,
jsmith@princeton.edu

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

Received 4 JAN 2018
Accepted 2 AUG 2018
Accepted article online 13 AUG 2018
Published online 15 SEP 2018

©2018. American Geophysical Union. All Rights Reserved.

Strange Floods: The Upper Tail of Flood Peaks in the United States

James A. Smith[1], Alexander A. Cox[1], Mary Lynn Baeck[1], Long Yang[1], and Paul Bates[2]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, [2]School of Geographical Sciences, University of Bristol, Clifton Bristol, UK

Abstract We examine the upper tail of flood peak distributions through analyses of annual peak observations from more than 8,000 U.S. Geological Survey (USGS) stream gaging stations and through hydrometeorological analyses of the storms that produce the most extreme floods. We focus on the distribution of the *upper tail ratio*, which is defined as the peak discharge for the flood of record at a stream gaging station divided by the sample 10-year flood magnitude. The 14 June 1903 Heppner storm, which produced an upper tail ratio of 200, was the product of a hailstorm that formed along the Blue Mountains in eastern Oregon, a region dominated by snowmelt flooding. A striking contrast between record flood peaks and the larger distribution of annual flood peaks in the United States is in the seasonality of flood occurrence, with record floods reflecting a much stronger contribution from warm season thunderstorm systems. Mountainous terrain and intense convective rainfall are important elements of the geography and hydrometeorology of extreme upper tail ratio flood peaks. The distribution of upper tail ratio values for USGS stream gaging stations does not depend on basin area, a result which is consistent with scaling results based on extreme value theory. Downscaling simulations with the Weather Research and Forecasting model are used to examine the storm environment of the 1903 Heppner storm, along with two other record flood peaks near the Blue Mountains of eastern Oregon from the USGS *miscellaneous* flood record, the July 1956 Meyers Canyon flood and the July 1965 Lane Canyon flood.

1. Introduction

The strangest flood in the systematic stream gaging record of the U.S. Geological Survey (USGS) is arguably the 14 June 1903 Heppner flood in eastern Oregon (Figure 1; Byrd, 2009; Whistler, 1903). The June 1903 flood peak in Balm Fork (Figure 1) is more than two orders of magnitude larger than the 10-year flood peak at the USGS stream gaging station for Balm Fork. The downstream flood peak in Willow Creek at Heppner (Figure 1) is 80 times larger than the 10-year flood for the Willow Creek stream gaging station. The 14 June 1903 Heppner flood, which resulted in more than 250 fatalities, was the product of a hailstorm that formed along the north slope of the Blue Mountains of Oregon (Whistler, 1903). There are no flood peaks in the systematic stream gaging record of the Blue Mountains region that are comparable to the June 1903 Heppner flood, but there are two *miscellaneous* flood events that have even more extreme estimated flood peaks than the Heppner flood. The July 1956 Meyers Canyon flood (Cooper, 2006; Costa, 1987a; Hendricks, 1964; Levish & Ostenaa, 1996) and the July 1965 Lane Canyon flood (Cooper, 2006; Costa, 1987a; Rostvedt, 1970) control the U.S. envelope curve (Costa, 1987a) and the world envelope curve (Costa, 1987b; Herschy, 2003) for basin scales between 10 and 40 km^2. In a report on flood frequency for eastern Oregon, Cooper (2006) notes that annual peaks for USGS stream gaging stations in eastern Oregon occurred most frequently in spring and resulted from snowmelt. The largest unit discharge flood peaks are "all due to thunderstorms, [but] thunderstorms are essentially unrepresented in the systematic record" (Cooper, 2006).

We will use the upper tail ratio, which is defined as the peak discharge for the flood of record at a gaging station divided by the sample 10-year flood magnitude, to examine upper tail properties of floods in the United States. The 10-year flood is a commonly used threshold to distinguish properties of the upper tail of flood distributions (see, e.g., O'Connor & Costa, 2004a; Villarini et al., 2011). In this paper, we will examine strange floods through analyses of USGS flood records from more than 8,000 stream gaging stations and through more detailed analyses of the most extreme floods and the storms that produce them.

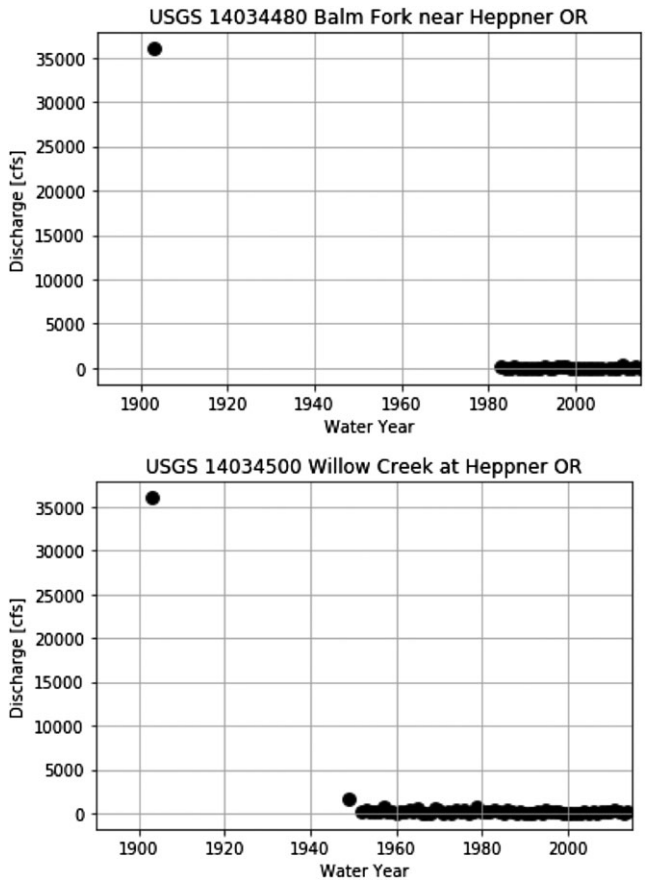

Figure 1. Annual maximum flood peaks for Balm Fork (top) and Willow Creek (bottom). Balm Fork has a drainage area of 68 km^2 and Willow Creek has a drainage area of 251 km^2 (see also Table 1). USGS = U.S. Geological Survey.

One of the key notions behind strange floods, as defined by large upper tail ratios at stream gaging stations, is that annual flood peak observations for a drainage basin may reflect mixtures of different populations of flood agents (see, e.g., Hirschboeck, 1987, 1988; Rossi et al., 1984; Smith et al., 2011a), like the snowmelt and thunderstorm floods in eastern Oregon (for similar examples of flood mixtures in Switzerland, see Fundel & Zappa, 2011). We use the eastern Oregon floods to illustrate the nature of strange floods for which the flood-generating mechanism (orographic thunderstorms) is rare and contrasts markedly with the common flood-generating mechanism (snowmelt). We will show that strange floods can also arise from unprecedented extremes in rainfall for the dominant flood-generating mechanism in a region, as is the case for a collection of record floods in the High Plains and along the Front Range of the Rocky Mountains. Mixtures of flood-generating mechanisms can be examined through analyses of the seasonality of flood peaks (see Villarini, 2016, for a recent discussion of seasonality of U.S. flood records). We will show that there is a striking contrast between the seasonality of record flood peaks at USGS stations and the larger record of annual flood peaks from USGS stream gaging stations. The contrast grows even sharper as we restrict attention to record floods with large upper tail ratio values.

The notion of strange floods is also tied to mountainous terrain. Orographic thunderstorms are agents of strange floods in eastern Oregon and in many other settings around the United States (Costa, 1987a) and the world (see, e.g., Ducrocq et al., 2008; Herschy, 2003). On the extreme end of terrain, Jarrett concluded that 2,300 m is the upper limit for the occurrence of extreme floods in the Front Range of the Rocky Mountains (Jarrett & Costa, 1988; Jarrett & Tomlinson, 2000; see also ; Mahoney et al., 2012, for hydrometeorological discussion). Spatial heterogeneity of flood peaks and scaling of flood peaks with drainage area are central topics for characterization of upper tail properties of floods. The literature on flood regionalization and flood peak scaling (see, e.g., Ayalew et al., 2014; Blöschl & Sivapalan, 1997; Hosking & Wallis, 1997; Nguyen et al., 2013; Robinson & Sivapalan, 1997) provides important background for the examination of strange floods.

Hailstorms have been discounted as heavy rainfall producers based on arguments revolving around low precipitation efficiency and rapid storm motion. Cotton et al. (2010) note that "storms producing the largest hailstones occur in strongly sheared environments; thus, in general, we should not expect that the storm systems producing the largest hailstones are also heavy rain producing storms" (see also Hamada et al., 2015; Zipser et al., 2006). Doswell et al. (1996) and Smith et al. (2001) take a different perspective, pointing to the most intense thunderstorms as underappreciated agents of extreme flooding (see also Bunkers & Doswell, 2016; Hitchens & Brooks, 2013; Llasat et al., 2016; Nielsen et al., 2015; Nielsen & Schumacher, 2018; Rogash & Racy, 2002).

The 1903, 1956, and 1965 floods in eastern Oregon seem strange partly because we have relatively little insight to the storms that produced them. We utilize downscaling simulations from the Weather Research and Forecasting (WRF) model to reconstruct the environment of the Heppner, Meyers Canyon, and Lane Canyon storms and to examine aspects of the structure and evolution of the storms. We have used downscaling simulations to examine the hydrometeorology of the 1927 flooding in the Lower Mississippi River (Smith & Baeck, 2015) and the 18–19 July 1942 Smethport, Pennsylvania storm that produced world record rainfall at 4- to 5-hr time scale (Michaelis & Lackmann, 2013; see Smith et al., 2011b, for similar analyses).

Measurement error is an important element of flood frequency analysis, especially for the upper tail of flood distributions (see, e.g., Costa & Jarrett, 2008; Potter & Walker, 1981). The Meyers Canyon and Lane Canyon flood peak estimates have generated considerable controversy (Costa & Jarrett, 2008; Levish & Ostenaa, 1996)

Table 1
U.S. Flood Events With Upper Tail Ratio Greater Than 20

Name	Date / USGS ID	Upper tail ratio	Area km^2	S_i
1. Teton Dam	5 June 1976			
Teton River	13055000	319	2,295	NA
2. Heppner	14 June 1903			
Balm Fork	14034480	200	68	1.288
Willow Creek	14034500	80	252	1.560
3. Republican	31 May 1935			
Republican R.	06837000	71	31,752	3.002
Republican R.	06843500	35	37,413	3.369
Republican R.	06828500	32	21,007	2.962
Red Willow Ck.	06837500	25	1,808	2.202
4. Washita	3 April 1934			
Washita R.	07324200	57	3,539	0.772
5. Front Range	16–17 June 1965			
Plum C.	06709500	55	783	1.769
Jimmy Camp Ck.	07105900	34	170	2.193
Cimmaron R.	07207000	27	708	1.976
Rayado Ck.	07208500	21	159	1.470
6. Rapid City	9 June 1972			
Rapid Ck.	06412500	49	973	4.095
Rapid Ck.	06414000	34	1,075	3.673
Boxelder Ck.	06422500	32	244	3.737
Spring Ck.	06408500	26	531	3.949
Battle Ck.	06404000	21	152	4.961
7. Prairie Dog	28 May 1953			
Prairie Dog Ck.	06847900	36	1,536	1.998
8. Sybille Ck.	20 Aug 1990			
Sybille Ck.	06664400	35	505	1.433
9. San Felipe CA	21 Feb 1980			
San Felipe Ck.	10255700	27	255	1.131
10. Ashland, OR	15 Jan 1974			
EF Ashland Ck.	14353500	25	20.9	0.860
WF Ashland Ck.	14353000	21	25.5	0.856
11. San Antonio CA	2 March 1938			
San Antonio Ck.	11073000	24	43	1.653
12. Sand Creek CO	19 July 1977			
Sand Ck.	06659580	22	79.1	0.919
13. Costilla NM	22 July 1954			
Costilla Ck.	08252500	22	65.6	2.082
14. Medio TX	22 Sept 1967			
Medio Ck.	08189300	21	527	1.155
15. Teton River	8 June 1964			
Teton River	06102500	21	287	11.20
16. San Emigdio CA	6 Aug 1961			
San Emigdio Ck.	11195500	20	127	0.659

and reflect the important role that measurement error plays in examining the nature of the upper tail of flood peak distributions.

There are 16 flood events, including the June 1903 Heppner flood, that produced upper tail ratios larger than 20. These provide a catalog of events (Table 1) that we use to examine upper tail properties of floods in the United States. Not surprisingly, this catalog includes some of the most devastating floods in U.S. history. Analyses of these events augment and expand analyses based on the entire USGS flood peak record, especially through insights concerning the role of thunderstorms and mountainous terrain in determining upper tail properties of flood peaks.

We introduce a statistical framework based on the Generalized Extreme Value (GEV) distribution to ground our analyses of the upper tail ratio. The GEV has been widely used for flood frequency analysis (see Katz et al., 2002), based on arguments from extreme value theory and the representation of the annual maximum flood peak as the maximum of a number of flood peaks over the course of a year. The GEV framework that we develop yields a scaling theory for record floods in which the upper tail ratio does not depend on drainage area. Analyses of upper tail ratios provide insights to upper tail thickness and, more generally, the nature of the upper tail of flood peaks. Robert Horton, along with many of his contemporaries, was convinced that rainfall and floods were bounded, writing "It is not difficult to show from sound meteorological reasoning, and aside from any statistical proof, that there is a natural limitation to rain intensity for any given duration" (letter to F. E. Schmidt, Associate Editor for Engineering News Record; 18 November 1927). We use the upper tail ratio to examine whether bounded flood distributions or unbounded thick-tailed distributions are more consistent with the USGS annual peak record. The literature on envelope curves and record floods (see, e.g., Douglas & Vogel, 2006; Enzel et al., 1993; Gaume et al., 2009; Halbert et al., 2016; Vogel et al., 2001) provides important background for these analyses. Zorzetto et al. (2016) develop ideas for characterizing the upper tail of hydrologic extremes in which extremes are represented as emergent properties of more common events.

Quantifying the upper tail of flood magnitudes is critical for flood hazard assessment, making strange floods a topic of significant practical importance. Many of these destructive, strange floods led to construction of flood control dams and levees to reduce loss of life and catastrophic damages. Our analyses are designed to advance the understanding of extreme floods and point to new methods for characterizing the upper tail of flood distributions. The United States is an important study region both for the societal importance of flood hazards and due to the exceptional observational resources provided by the USGS stream gaging record. Our analyses, however, are designed to provide general insights on extreme flooding and to provide results that can be usefully compared to extreme flooding around the world (see, e.g., Costa, 1987b; Delrieu et al., 2005; Gaume et al., 2009; Hosking et al., 1985; Li et al., 2013; Llasat et al., 2016; Nguyen et al., 2013; Parajka et al., 2010; Salinas et al., 2014; Smith et al., 2015; Yang et al., 2017).

The objective of this paper is to characterize the *nature* of the upper tail of flood peak distributions in the conterminous United States. To address this objective, we will focus on the following questions: (1) What are the flood-generating mechanisms that control the upper tail of flood distributions in the United States? (2) What is the relationship between strange floods and the convective intensity of flood-producing storms? (3) What are the characteristic features of watersheds with the largest upper tail ratios in the United States? (4) Are strange floods produced by strange storms or by extremes of more common storms? (5) How do upper tail ratio values scale with drainage area and what are the implications for assessing distributional properties of the upper tail of flood peak distributions? (6) Why does eastern Oregon have the strangest floods in the United States?

2. Data and Methods

We use USGS stream gaging records, and especially the annual maximum flood peak record (see Ryberg et al., 2017, for a recent description), to examine strange floods in the conterminous United States (i.e., we exclude Hawaii, Alaska, and Puerto Rico). Measurements of many extreme floods are made by *indirect discharge* methods, involving field measurements of peak water surface profiles and channel cross-sections, combined with hydraulic computations (Costa & Jarrett, 2008; Koenig et al., 2016). Indirect measurements are made for floods at stream gaging sites when the gage is destroyed or fails to operate properly. They are also made at miscellaneous sites, that is, sites that do not have stream gaging stations, typically for the most extreme floods. The envelope curve of flood peaks for the United States is controlled by indirect discharge measurements from miscellaneous sites (Costa, 1987a). Estimates of the July 1956 Meyers Canyon and July 1965 Lane Canyon

flood peaks in eastern Oregon, which are based on indirect discharge methods at miscellaneous sites, have generated considerable controversy (as discussed in section 4).

The sample of flood peak observations consists of annual maximum instantaneous discharge from USGS stream gaging stations. For a single stream gaging station, let $X_1, \ldots, X_n$ denote the sample of annual flood peaks and $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ denote the order statistics of the annual peak record with n years of observations. With this notation, $X_{(n)}$ is the flood of record. For stations that satisfy the stationarity assumption, we can represent the flood peak distribution at a site by the cumulative distribution function

$$F(x) = P\{X_i \leq x\} \quad , \quad x > 0. \tag{1}$$

An alternative representation is through the quantile function:

$$Q(p) = F^{-1}(p) \quad , \quad p \quad \in [0, 1]. \tag{2}$$

The GEV distribution has been widely used for hydrologic extremes, based on both theoretical and practical considerations (El Adlouni et al., 2007; Martins & Stedinger, 2000; Smith et al., 2015; Villarini & Smith, 2010). From a theoretical perspective, the GEV represents a limiting distribution that results from the operation of taking the maxima of independent and identically distributed or weakly dependent random variables (Leadbetter et al., 1982). The quantile function of the GEV takes the form

$$Q(p) = \mu - \sigma \left\{ \frac{(1 - [-ln(p)]^{-\xi})}{\xi} \right\} , \ \xi \neq 0, \tag{3}$$

$$= \mu - \sigma \, ln\{-ln(p)\} \ , \ \xi = 0, \tag{4}$$

where $\mu \in (-\infty, +\infty)$ is the location parameter, $\sigma \in (0, +\infty)$ is the scale parameter, and $\xi \in (-\infty, +\infty)$ is the shape parameter. For $\xi > 0$, the distribution is unbounded above. For $\xi < 0$, the distribution is bounded above with an upper bound of $\mu - \sigma/\xi$. The Gumbel distribution is the special case for $\xi = 0$ and corresponds to unbounded, but *thin*, upper tails. The GEV provides a useful tool for examining tail properties of flood peak distributions. We will use the shape parameter ξ as an index of tail properties (see, e.g., Morrison & Smith, 2002).

We examine the stationarity of annual maximum peak discharge data using the Mann-Kendall test (see Morrison & Smith, 1993). The test has been widely used for hydrologic time series, including annual peak records (see, e.g., Villarini et al., 2009). Other methods, including change point analyses, can be used to identify the nature and timing of nonstationarities (see, e.g., Helsel & Hirsch, 1993; Villarini et al., 2009). In this study, we will focus on the presence or absence of nonstationarities in a flood record and restrict analyses to those based on the Mann-Kendall test.

We estimated the GEV parameters for stationary annual peak records (5,546 USGS stations) using the maximum likelihood method (see Coles, 2001). The estimated location and scale parameters $\{\hat{\mu}_i, \hat{\sigma}_i\}$ are linearly related (correlation of 0.91), with $\hat{\mu} = c \times \hat{\sigma}$ and c approximately equal to 2. The estimated location and scale parameters also have significant correlation with drainage area (0.57 for both). The estimated shape parameters $\{\hat{\xi}_i\}$ are independent of the estimated location and scale parameters and drainage area. We examined dependence of the estimated shape parameter on a range of other covariates contained in the USGS Gages-II database (Falcone et al., 2010) and found no significant results.

The upper tail ratio, Z, is the ratio of the flood of record to the sample 10-year flood, that is,

$$Z = \frac{X_{(n)}}{\hat{Q}(.9)}, \tag{5}$$

where $\hat{Q}(.9)$ is computed in the standard way as a linear combination of the order statistics $X_{(j)}$ and $X_{(j+1)}$, with $\frac{j}{n+1} < 0.9 \leq \frac{j+1}{n+1}$. We will use the upper tail ratio as a principal tool for examining the upper tail of all USGS flood records. For nonstationary records, $\hat{Q}(.9)$ cannot be interpreted as a 10-year flood for any particular time period, but provides an aggregate characterization of the 1 in 10 probability of occurrence over the changing flood regimes represented in a particular record of a stream gaging station. Of particular importance for our

analyses will be stations for which extreme floods provided the stimulus for river regulation, stream gaging, or both.

The upper tail ratio provides a nonparametric approach to analyze the upper tail of flood peak distributions. Parametric assumptions on the parent flood peak distributions from extreme value theory provide guidance on the distribution of the upper tail ratio for stationary flood records. Under the assumptions that (1) the parent flood peak distribution for a stream gaging station has a GEV distribution and the shape parameter is linearly related to the location parameter, with $\mu = c\,\sigma$ (see discussion above) and (2) the sample 10-year flood is close to the theoretical 10-year flood, then the upper tail ratio has a GEV distribution that depends only on the shape parameter and sample size (details are provided in the Appendix A). This result provides a characterization of the control of the shape parameter and sample size on upper tail properties of the GEV distribution (see Koutsoyiannis & Montanari, 1999, for related scaling analyses of extreme precipitation). Under the assumption that the GEV provides a good representation of annual flood peaks, the result shows that upper tail ratio and the shape parameter of the GEV are intricately linked. Because the shape parameter does not depend on basin scale, it follows that the same holds for the upper tail ratio.

The form of the GEV distribution for Z does not depend on the selection of the 10-year flood as the normalizing variable. The same form of the distribution (equation (A3)) holds for any quantile $Q(p)$, with 0.9 replaced by the new value of p (equation (A4)). More generally, we could define Z as the ratio of the record flood to the sample T-year flood for values of T different from 10 years. Selecting shorter return intervals could provide sample T-year floods with lower variability, but at the cost of reduced relevance to the upper tail of flood distributions. Our selection of the 10-year flood is motivated in part by DuMouchel's admonition to *let the tails speak for themselves* (DuMouchel, 1983).

3. Strange Floods—The USGS Stream Gaging Record

3.1. Upper Tail Ratio for USGS Stream Gaging Stations

We examine the estimated upper tail ratio for 8,911 stream gaging stations that are included in the USGS GAGES-II data set (Falcone et al., 2010). For each stream gaging station, we have a vector of covariates which includes the number of annual peak observations n_i, drainage area A_i, the time $T_i \in [0, 1]$ of the record flood (with time 0 denoting the beginning of the day on 1 January and time 1 denoting the end of the day on 31 December), the mean annual precipitation, mean basin elevation, and a wide range of land use/land cover and basin morphology variables taken from the Gages-II data set. We principally use time T_i to assess seasonality of flood occurrences. Synoptic analyses of USGS flood records have assessed flood frequency for the United States from diverse perspectives (see, e.g., Costa, 1987a, 1987b; Michaud et al., 2001; O'Connor & Costa, 2004a, 2004b; O'Connor et al., 2002; Patton & Baker, 1976; Smith & Smith, 2015).

The locations of flood peaks with upper tail ratios greater than 5, 10, and 20 are concentrated in diverse regions across the United States (Figure 2), but the largest upper tail ratio floods are found in the western United States (Table 1 and Figure 2). The spatial distribution of large upper tail ratio flood peaks reflects many of the features identified by O'Connor and Costa (2004a) in their examination of large unit discharge flood peaks. Analyses based on unit discharge flood peaks place a stronger spotlight on somewhat different regions (notably Texas and the Central Appalachians; see O'Connor & Costa, 2004a) than analyses based on upper tail ratio (see additional discussion below).

An important distributional property of the upper tail ratio is the dependence on record length (Figure 3). The median and 0.95 quantile values of upper tail ratio for the second decile (annual peak records of 14–22 years) are 1.42 and 3.94. For the ninth decile (75–86 years of annual peak records), the median and 0.95 quantile values are 2.13 and 6.37. Dependence of upper tail ratio on sample size is most pronounced for stations in the lower half of the distribution of sample size. For the upper half, that is, stations with more than 46 years of record, there is only a modest change in the distribution of upper tail ratio with increasing record length.

The GEV scaling theory for upper tail ratio implies that the distribution of upper tail ratio values should depend on the GEV shape parameter, in addition to sample size. This is supported by analyses based on estimated values of the GEV shape parameter (Figure 4). For the analyses in Figure 4, we use stations with sample size greater than 46 (the median value of sample size) to focus on the dependence of upper tail ratio on the shape parameter and to mitigate the impacts of sampling variability for estimators of the shape parameter. Median, interquartile range, and maximum values of upper tail ratio increase systematically with deciles of estimated shape parameter. Variability of upper tail ratio values increases sharply for the two largest deciles of estimated

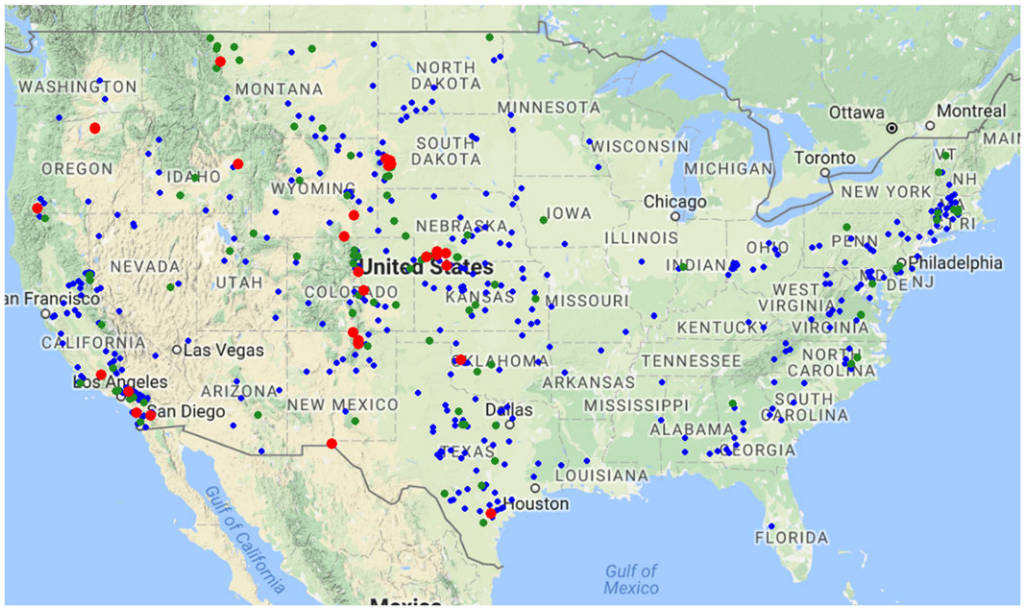

Figure 2. Locations of U.S. Geological Survey gaging stations with upper tail ratios exceeding 5 and less than 10 (blue), between 10 and 20 (green), and greater than 20 (red).

GEV shape parameters (the lower bound of the estimated shape parameter for the ninth decile is 0.5) which is consistent with prior studies of variability of maximum likelihood estimators of the shape parameter (see Morrison & Smith, 2002).

For the entire sample of 8,911 stream gaging stations, the distribution of upper tail ratio magnitudes increases with basin scale, as represented by drainage area, A_i (Figure 5 top). This result appears to be at odds with the GEV scaling theory for which the distribution of upper tail ratio depends only on the shape parameter and sample size. As noted in the previous section, the estimated shape parameter for stationary flood records does not depend on drainage area, or any other basin characteristic. There is, however, a modest positive correlation between basin area and sample size (figure not shown), which is sufficient to skew the distribution of upper tail ratio. If we restrict consideration to stations with more than 46 years of record (the median value of record length), the dependence of upper tail ratio on basin area disappears (Figure 5 bottom). The distribution of upper tail ratio magnitudes is uniform across the range of basin scales, consistent with the GEV scaling theory.

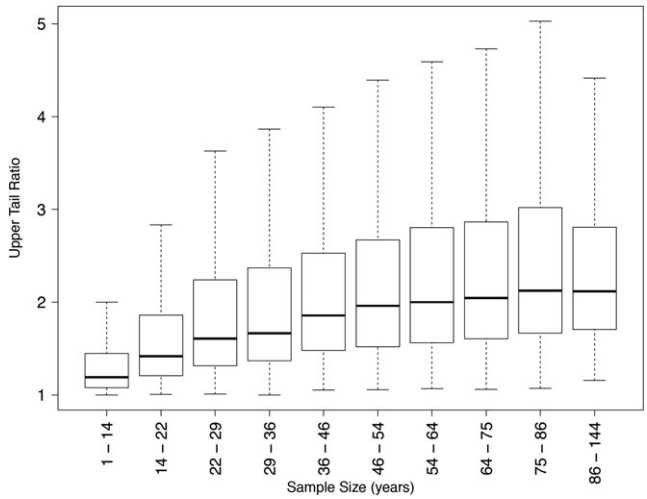

Figure 3. Boxplots of upper tail ratio organized by deciles of sample size in years. For each decile, the minimum and maximum values of sample size are shown.

Record floods, and especially the most extreme record floods, are concentrated during the peak of warm season convection (Figure 6). The distribution of record flood peaks has a maximum around mid-June, with a secondary maximum around the beginning of September, corresponding to the peak of extreme flooding in the eastern United States from landfalling tropical cyclones. The distribution sharpens when we restrict consideration to stations with upper tail ratio greater than 5, with the summer maximum becoming more prominent. The tropical cyclone peak shifts earlier in the season, due largely to a handful of hurricanes (Diane in August of 1955, Alberto in July of 1994, and Agnes in June of 1972), which have a disproportionate impact on the seasonal distribution of upper tail ratios greater than 5.

The contrasts in seasonal distribution of record floods and record floods with upper tail ratio greater than 5 are small, compared with the broad seasonal distribution of annual flood peaks at USGS stream gaging stations (Figure 6), which has winter/spring and summer maxima and a diminished contribution from tropical cyclones. The seasonal distribution of extreme

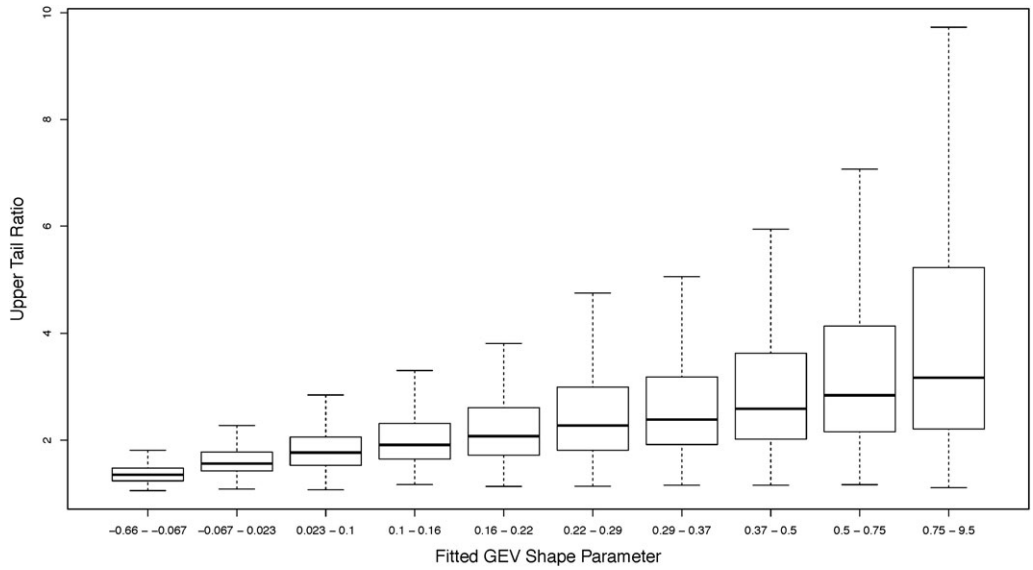

Figure 4. Boxplots of upper tail ratio organized by deciles of estimated GEV shape parameter. For each decile, the minimum and maximum values of sample size are shown. GEV = Generalized Extreme Value.

floods becomes even more skewed toward warm season convection and tropical cyclones when attention is restricted to small watersheds (less than 200 km^2 in drainage area; figure not shown).

We also examine strange floods through the record flood seasonality, S_i, which characterizes the seasonal anomaly of the record flood for station i:

$$S_i = \hat{f}_i(T_i), \tag{6}$$

where $\hat{f}_i(x)$, $x \in [0, 1]$, is the sample probability density function of the time of occurrence of flood peaks in basin i. We compute S_i as the sample probability density function of the seasonal occurrence of annual flood peaks at station i, evaluated at the time of occurrence T_i of the flood of record. We estimate the seasonal probability density function using the R Kernel Density Estimator (Venables & Ripley, 2002). If the times of occurrence are uniformly distributed over the year, the function $f_i(x)$ takes the value 1. Values of S_i less than 1 indicate record floods that have anomalous components of seasonality. Values that are significantly greater than 1 indicate record floods that are produced by common flood-generating mechanisms. Record flood seasonality between stations with upper tail index greater than 5 versus those with upper tail index less than 5 differ only on the high end of record flood seasonality. For upper tail index less than 5, the median value of S_i is 1.43, compared with 1.46 for upper tail index greater than 5. The 0.95 quantile of record flood seasonality is 7.83 for upper tail index less than 5 and 4.93 for upper tail index greater than 5, indicating a larger percentage of basins with very strong seasonality for upper tail index less than 5. These results reflect the prominence of winter/spring flood peaks, including record floods, for many stream gaging stations in the eastern United States.

Arid and semiarid regions have upper tail ratios that are somewhat larger than those in more humid regions (Figure 7). Upper tail ratios for the lowest two deciles of mean annual basin precipitation are systematically larger than those for basins with greater precipitation. The upper bound of the second decile of precipitation is approximately 650 mm. There is relatively little variation in the distribution of upper tail ratio with precipitation for the upper 8 deciles. As highlighted in Figures 2 and 7, arid/semiarid regions play an important role in controlling the extremes of upper tail ratio.

Jarrett's conjecture (Jarrett & Costa, 1988; Jarrett & Tomlinson, 2000) that 2,300 m elevation marks a significant break in extreme flood records is supported by analyses of the upper tail ratio for the entire USGS annual peak record (Figure 8). There is a pronounced decrease in median, interquartile range, and extremes of upper tail ratio for the last decile of basin elevations, which has a lower bound at 2,300 m. The sharp decrease in upper tail ratio basins with elevations greater than 2,300 m is paired with a general increase in upper tail ratio for deciles 6–9 (elevations ranging from approximately 395 to 2,300 m). This feature reflects the role of orographic precipitation mechanisms in amplifying flood peak extremes.

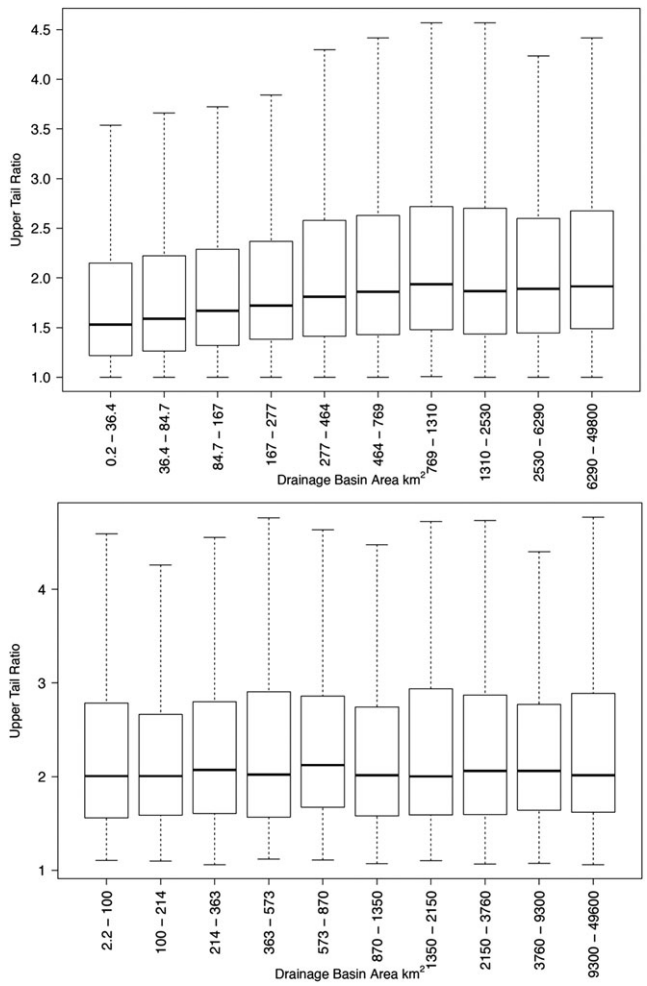

Figure 5. Boxplots of upper tail ratio organized by deciles of basin area in square kilometers (top). For each decile, the minimum and maximum values of basin area are shown. In the bottom panel of the figure, the sample of stations is conditioned on sample size exceeding 46 years, which is the median sample size (see Figure 3). Consequently, deciles of basin area in the bottom plot are for the smaller set of stations with sample size greater than 46 years.

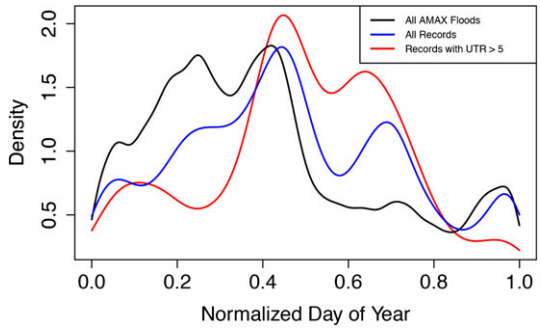

Figure 6. Sample probability density functions (based on R Kernel Density Estimator) of occurrence times of record floods (blue line), records floods with UTR greater than 5 (red line) and of all flood peaks from the 8,911 stream gaging stations (black line; denoted AMAX floods). Time 0 corresponds to 1 January and time 1 to 31 December. UTR = upper tail ratio.

For more than a century, there has been debate over the impact of forests on floods (see, e.g., Jones & Grant, 1996). There is little evidence of forest cover impacts on upper tail ratio (Figure 9), suggesting that forest cover does not have a major impact on the most extreme floods. This result is consistent with previous studies showing that the impacts of forest cover on flooding diminishes in the upper tail of flood events (Wood et al., 1990). These studies point to the relative importance of soil moisture decreasing as return interval (and storm total rainfall accumulations) becomes large. There are caveats to the inferences concerning forests and floods. The Gages-II forest cover data represent current conditions and do not capture changes over time in forest cover. There may also be impacts on more common floods that are not present for the most extreme events. For the smallest decile of forest cover, which includes urbanized watersheds and arid watersheds, there is an elevated distribution of upper tail ratio.

Upper tail ratio decreases as impervious fraction increases, implying that increases in urbanization result in lower values of upper tail ratio. This result is consistent with increases in frequency of flooding (which leads to elevated values of 10-year flood) and capacity constraints on peak flood magnitude imposed by the urban drainage network (Smith & Smith, 2015). Capacity constraints in urban watersheds result from design of components of the urban drainage network, including pipe capacities in the storm drain network and bridge culverts. The decreases in upper tail ratio with impervious fraction do not reflect the nature of flood hazards, which increase in urban regions due to the vulnerability of large flood-impacted populations (see DiBaldassarre et al., 2013, for discussion and review of human-flood interactions).

The pronounced nonstationarities in flood records of the United States (see, e.g., Hirsch & Archfield, 2015; Mallakpour & Villarini, 2015; Milly et al., 2008; Villarini et al., 2009) do not translate into large differences in upper tail ratio. We examine the distributions of upper tail ratio for stations that pass the Mann-Kendall test for stationarity and those that do not. The median value for the nonstationary stations, 1.99, is only slightly larger than the median value for stationary stations, 1.88. There are similarly modest changes for interquartile range and extremes. River regulation by dams is an important element of nonstationarities in flood records. There is a link between large upper tail ratio floods and dam building, as we illustrate in the following section. The occurrence of a large flood in a particular drainage basin marks that watershed as a target for flood control, especially if there is extensive loss of life or damage from the flood.

The sample properties of upper tail ratio for USGS stream gaging stations are most consistent with a GEV world in which the shape parameter is positive (Figures 10 and 11), implying unbounded, thick-tailed flood distributions. In Figure 10, we illustrate the dependence of the distribution of upper tail ratio on record length for GEV distributions with shape parameters of 0.25 (unbounded and thick-tailed), 0 (Gumbel), and −0.25 (bounded), with sample sizes taken from the 8,911 USGS stations. The bounded case produces upper tail ratios that are too small at all quantiles and the extremes in the Gumbel case are too small. Only the unbounded, thick-tailed case can produce the extremes that are present in the USGS upper tail ratio samples. In Figure 11, we show the distribution of upper tail ratio for 8,000 simulated flood records of length 60 years with shape parameter ranging from −0.5 to 0.5. Again, we conclude that bounded and Gumbel distributions are not consistent with the large upper tail ratios for

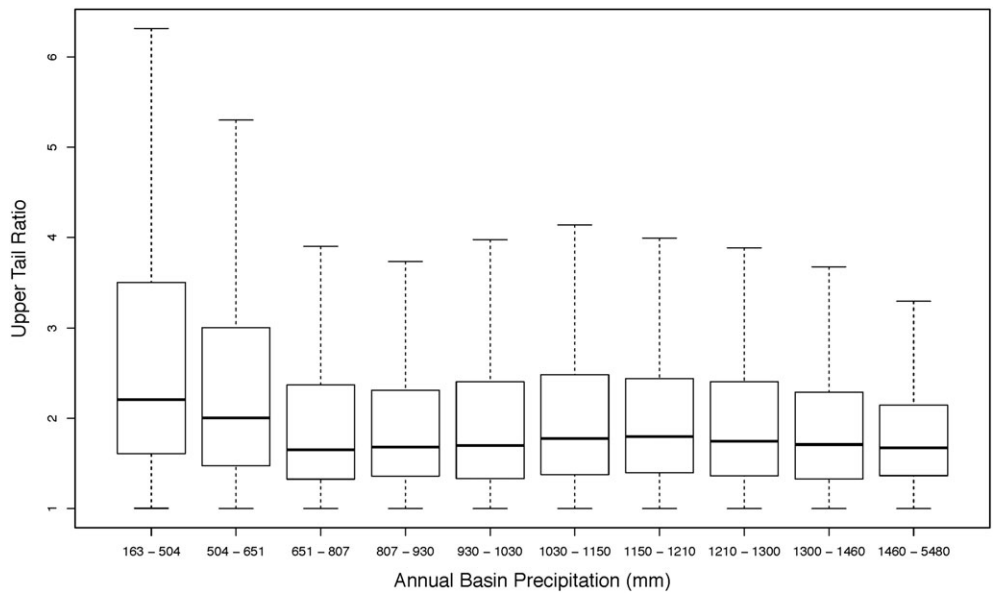

Figure 7. Boxplots of upper tail ratio organized by deciles of mean annual precipitation for the drainage basin (mm). For each decile, the minimum and maximum values of precipitation are shown.

USGS gaging stations. Large upper tail ratios in the GEV world are tied to unbounded thick-tailed distributions. We have also explored mixtures of GEV families and come to a similar conclusion (figures not shown). These results, along with relationships between estimated GEV shape parameters and upper tail ratio (Figure 4), point to thick-tailed distributions as a key element of large upper tail ratio flood records. Future studies will examine these issues in more detail.

3.2. The Largest Upper Tail Ratio Floods in the United States

In this section, we examine the largest upper tail ratio flood peaks in the USGS stream gaging record (Table 1 and Figure 2) to expand the interpretation of results presented in the previous section. The 14 June 1903 flood peak in Balm Fork (Table 1 and Figure 1) translates to an upper tail ratio of 198, the second largest in the USGS stream gaging record (Table 1). Balm Fork enters Willow Creek immediately upstream of Heppner, Oregon. The Willow Creek flood peak from the 1903 flood ranks third, with an upper tail ratio of 80 (Table 1 and Figure 1). The flood peaks for Balm Fork and Willow Creek in 1903 were determined from indirect discharge estimates (see section 2). Years after the Heppner flood, USGS stream gaging stations were installed in Willow Creek (1949) and Balm Fork (1983), providing the observations to assess the distribution of the upper tail ratio; the record length for the Willow Creek station is 67 years and the record length for the Balm Fork station is 33 years. The Heppner flood is the quintessential strange flood. It has the largest upper tail ratio for a meteorological flood in the United States and it was produced by an unusual flood agent for the region, that is, a hailstorm in complex terrain. As Cooper (2006) notes, flood peaks in the Blue Mountains region of eastern Oregon are largely from snowmelt.

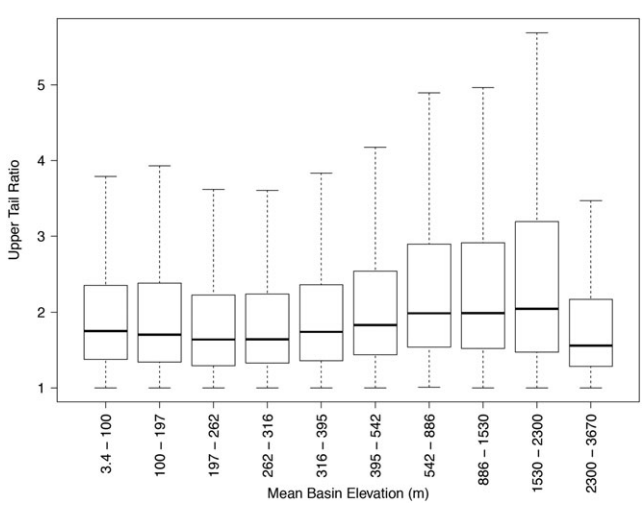

Figure 8. Boxplots of upper tail ratio organized by deciles of mean basin elevation (meters above mean sea level). For each decile, the minimum and maximum values of elevation are shown.

The largest upper tail ratio in the USGS flood record, 319 from the 5 June 1977 flood peak of 48,140 m^3/s on the Teton River near St. Anthony, Idaho (2,295 km^2), highlights the extreme nature of the June 1903 Heppner flood. The Teton River flood resulted from the failure of the Teton Dam, which was located a short distance upstream from the gaging station. One of the most notable dam failures in U.S. history (see O'Connor et al., 2002, for discussion of dam-break floods), the resulting flood devastated downstream communities. The value of the Balm Fork upper tail ratio is closer to the Teton River upper tail ratio than it is to the Willow Creek upper tail ratio (Table 1).

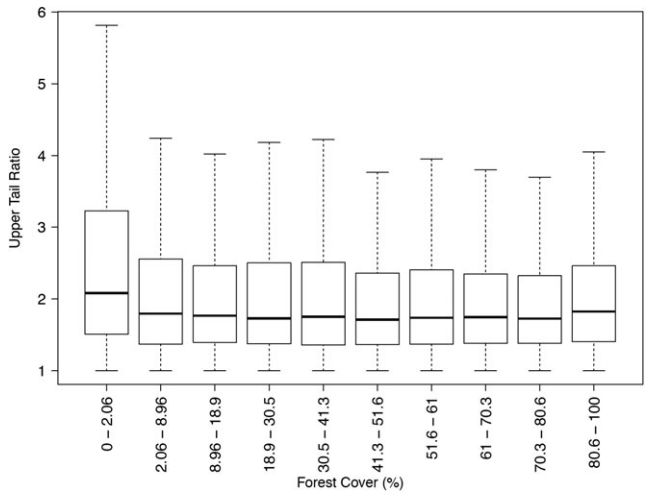

Figure 9. Boxplots of upper tail ratio organized by deciles of percent forest cover. For each decile, the minimum and maximum values of percent forest cover are shown.

The 31 May 1935 Republican River Flood, which produced an upper tail ratio of 71 for the McCook, Nebraska, gaging station (6,838 m^3/s peak; see Table 1), is notable for the relatively large drainage area, 31,700 km^2, and for the combined occurrence of extreme flooding and severe weather (Follansbee & Spiegel, 1937). In addition to the upper tail ratio of 71 at the McCook gaging station, four other USGS stream gaging stations reported record flood peaks with upper tail ratios greater than 20 (Table 1).

The Republican River basin is located in a corridor of high frequency occurrence of extreme rainfall events during the May–July period. The corridor extends from the Front Range of the Rocky Mountains through Nebraska into Iowa. This feature is linked to the climatological maximum of rainfall from mesoscale convective complexes (see Maddox, 1980; McAnelly & Cotton, 1989; Tuttle& Davis, 2006; Zhang et al., 2001), and more generally, the climatology of mesoscale convective systems (Schumacher & Johnson, 2005). Heavy rainfall on 30 May 1935 was initiated along the eastern margin of the Rocky Mountains near Colorado Springs, Colorado, around the middle of the day and the path of heavy rainfall was extended northeast through eastern Colorado, northwestern Kansas and into Nebraska on 31 May; peak rainfall accumulations exceeded 500 mm. The record flood seasonality index, S_i for the Republican River flood of 3.0 (Table 1), is large, reflecting the prominence of mesoscale convective systems in the flood record for the region.

Like the Heppner flood, the principal flood agents for the Republican River flooding were from the most intense spectrum of thunderstorms. Multiple bucket survey reports noted that precipitation accumulations included rain and hail (Follansbee & Spiegel, 1937). As storms passed McCook, Nebraska, a tornado was reported (Follansbee & Spiegel, 1937). Unlike the Heppner flood, the 1935 Republican River flood is not strange in terms of the flood agent at play, but in terms of the magnitudes of rainfall and flood peaks. Extreme rainfall and flooding extended down the High Plains from Nebraska to Texas. The world record rainfall at 2.75-hr time scale is the 559 mm rainfall accumulation in D'Hannis, Texas, on 31 May 1935. The D'Hannis storm was a tornadic supercell thunderstorm (Smith et al., 2000) that produced an envelope curve flood peak in Seco Creek (based on indirect discharge measurements; see Costa, 1987a and Costa & Jarrett, 2008).

A common consequence of the largest floods in the USGS flood record has been to stimulate development of flood control systems, typically through dam building. The May–June 1935 flooding was a major impetus for the 1944 Pick-Sloan Plan (Billington & Jackson, 2006), which resulted in a system of dams in the Republican River basin, and the Flood Control Act of 1936, which accelerated dam building throughout the United States. The annual peak records for the Republican River stream gaging stations all exhibit nonstationarities and these are tied to dam building after implementation of the Pick-Sloan Plan.

The High Plains region of the central United States is an area with a concentration of large upper tail ratios extending from Texas to North Dakota (Figure 2 and Table 1). Like the Republican River flood of 1935, these events are dominated by mesoscale convective systems. The Washita flood of 3 April 1934 in Oklahoma and Prairie Dog Creek flood of 28 May 1953 in Kansas (listed in the USGS annual peak records as 23 May 1953) are two of the most extreme (Table 1). The Washita flood, which ranks just behind the Republican River flood in upper tail ratio (Table 1), resulted in 17 fatalities and stimulated development of a flood control system for the Washita River basin. Newspaper accounts of the storm refer to the frequent occurrence of lightning. The Prairie Dog Creek flood had an upper tail ratio of 37 (seventh on the USGS list; Table 1) and was produced by rains of 8 to 11 in., which resulted in a *phenomenal rise in Prairie Dog Creek* (Wells, 1959). Extreme rainfall for the Prairie Dog Creek storm was associated with a tornadic thunderstorm (based on National Centers for Environmental Information [NCEI] Storm Reports).

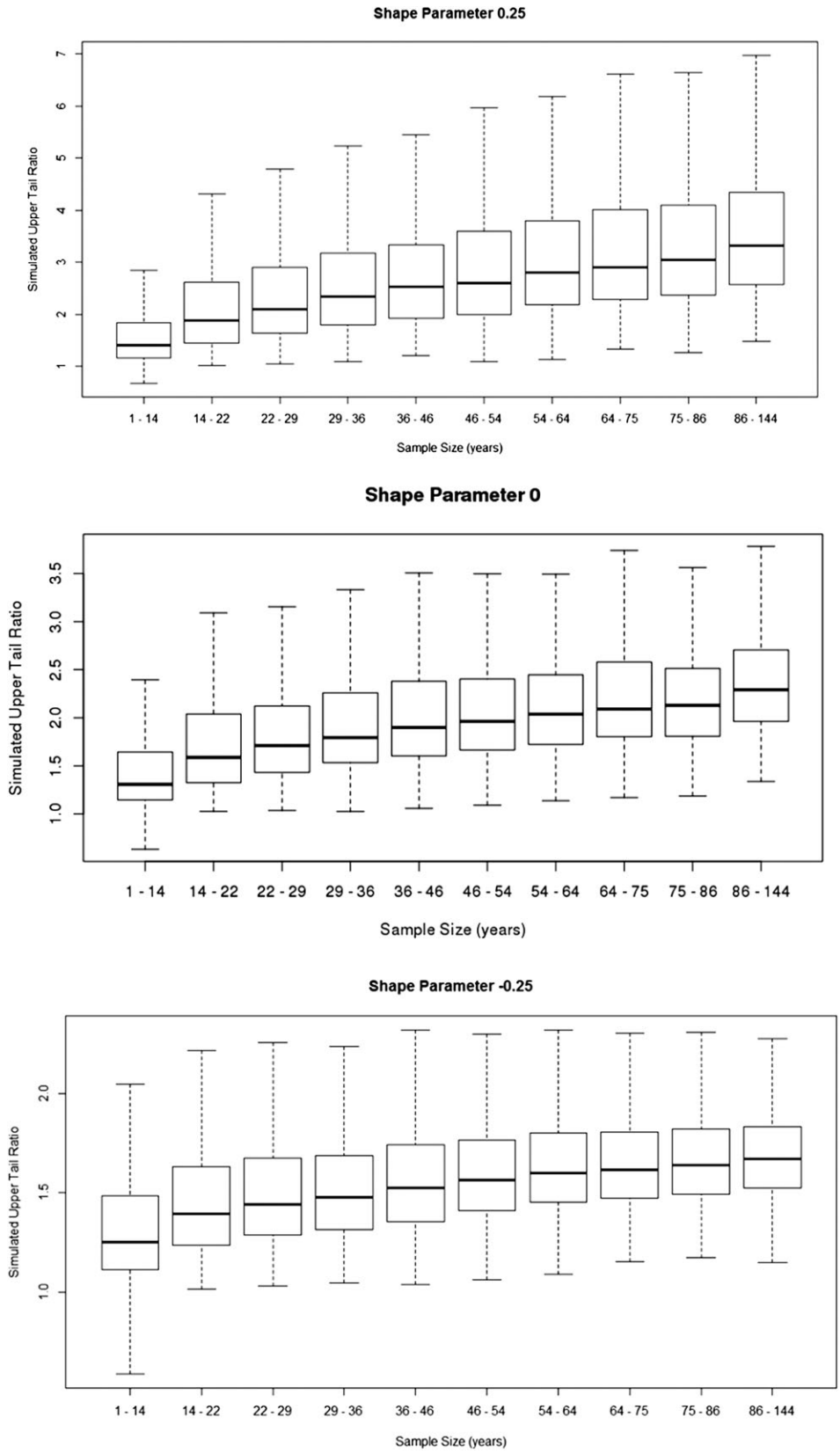

Figure 10. Boxplots of simulated upper tail ratio from Generalized Extreme Value distributions with shape parameter 0.25 (top; unbounded, thick-tailed), 0 (middle; Gumbel), and −0.25 (bottom; bounded). The location parameter is 0 and shape parameter is 1 for all simulations. Boxplots are organized by deciles of sample size.

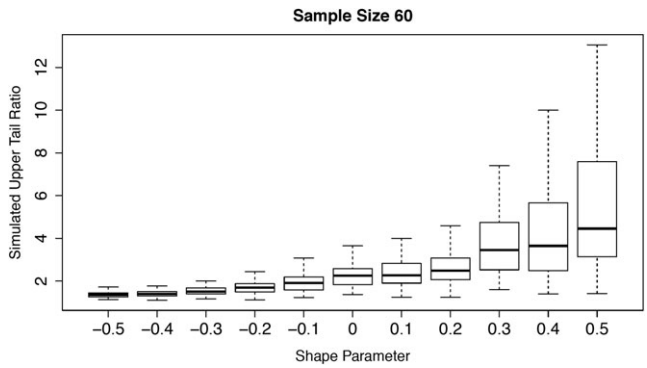

Figure 11. Boxplots of simulated upper tail ratio from Generalized Extreme Value distributions with shape parameters ranging from −0.5 to 0.5 with sample size of 60 years.

The most striking expression of mountainous terrain on the occurrence of large upper tail ratio floods is in the Front Range of the Rocky Mountains (Figure 2; see Follansbee & Jones, 1922; Jarrett & Costa, 1988; Javier et al., 2007; Maddox et al., 1978; Mahoney et al., 2012; Ogden et al., 2000; Petersen et al., 1999). By many metrics, including the upper tail ratio (Table 1), the most extreme flooding in the Front Range region resulted from a series of storms during the period from 16 to 18 June 1965 (in terms of fatalities, however, the 21 deaths in 1965 rank well behind the more than 140 fatalities in Big Thompson on 31 July 1976 and the *hundreds* of fatalities in Pueblo on 4 June 1921; see England et al., 2014; Follansbee & Jones, 1922). The Plum Creek flood peak in suburban Denver on 16 June 1965 resulted in an upper tail ratio of 55 at a drainage area of 783 km^2 (and, not surprisingly, development of flood control for Plum Creek, Bear Creek, and the South Platte River). The Jimmy Camp Creek flood peak the following day produced an upper tail ratio of 34 at a drainage area of 170 km^2 and an envelope curve flood peak for the United States (Costa, 1987b). The storms were severe thunderstorms, especially the Jimmy Camp Creek storms, which were associated with large hail and F1 tornadoes. The only overlap between Costa's U.S. flood peak envelope curve (Costa, 1987a) and the strange floods in the systematic record (Table 1) is the Jimmy Camp Creek flood.

The 16–18 June 1965 period was a major flood outbreak along the Front Range from Colorado south into New Mexico. The Cimarron River near Cimarron, New Mexico, on 17 June (708 km^2) reported a record flood with an upper tail ratio of 27. Rayado Creek near Cimarron, New Mexico, (159 km^2) reported a record flood peak with an upper tail ratio of 21. Tornado reports (NCDC Storm Events database) suggest that the Rayado Creek flood was produced by extreme rainfall from a supercell thunderstorm. The record flood seasonality for the Front Range watersheds experiencing record floods during the 16–18 June 1965 period range from 1.5 to 2.2 (Table 1), reflecting the prominence of warm season thunderstorms in the flood records of the Front Range region. The June 1965 flood was not strange in terms of the flood-generating agent, but in the magnitude of rainfall.

The combination of mountainous terrain and organized thunderstorms resulted in one of the most deadly floods in U.S. history in Rapid City, South Dakota, on 9 June 1972. The Rapid City flood (also known as the Black Hills Flood) was not just one of the most extreme floods in U.S. history in terms of upper tail ratio (49 at a drainage area of 973 km^2, with five stations exceeding an upper tail ratio of 20; see Table 1), but also in terms of fatalities, with at least 238 deaths that were the direct result of the flood. The flood was produced by a nearly stationary system of thunderstorms, with orographic mechanisms playing an important role in storm structure and evolution (Maddox et al., 1978; see also Soderholm et al., 2014). The Black Hills region has an even more pronounced seasonal concentration of flood peaks in June than the Front Range, with record flood seasonality for the five stations ranging from 3.7 to 5.0 (Table 1). Paleoflood studies (Harden et al., 2011) have shown that the June 1972 floods were not unprecedented; during the past 2000 years, multiple flood peaks as large or larger than the June 1972 flood occurred in the watersheds most heavily impacted by the June 1972 storms.

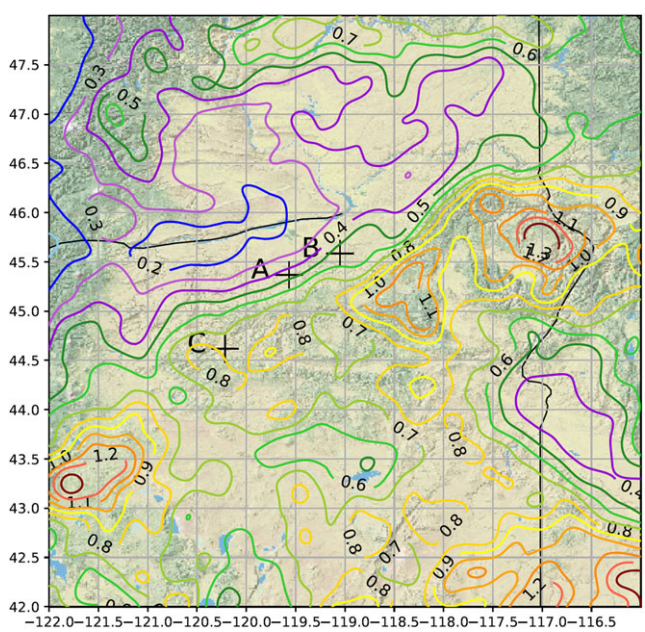

Figure 12. Mean annual cloud-to-ground lightning flash density (cloud-to-ground strikes km^{-2}) for the Blue Mountains region of eastern Oregon. Locations of the Heppner, Lane Canyon, and Meyers Canyon floods are denoted by "A", "B" and "C", respectively.

Although Jarrett's 2,300-meter hypothesis is supported by results in Figure 8, there are exceptions to the 2,300-meter rule, with one ranking among the largest upper tail ratio flood events. The Costilla Creek flood of 22 July 1954 (Table 1) occurred in a watershed with an elevation of 2,880 m at the basin outlet. Costilla Creek is located in the Sangre de Christo Mountains of northern New Mexico. The 22 July 1954 flood peak occurred during the North American Monsoon season. Snowmelt is the dominant flood agent in the watershed, with 70% of annual flood peaks occurring during the April–June snowmelt season. The third and fourth largest peaks, however, are North American Monsoon events occurring in August 1957 and

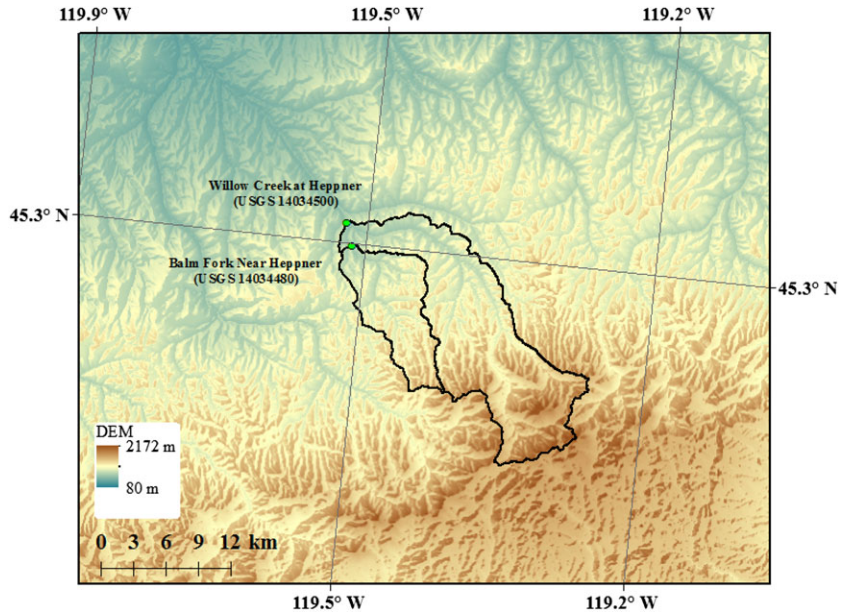

Figure 13. Basin boundaries of Balm Fork and Willow Creek, with topography. USGS = U.S. Geological Survey; DEM = digital elevation model.

August 1982. The record flood seasonality of 2.1 for the 22 July 1954 peak (Table 1) reflects the bimodal distribution of flood seasonality in the watershed. Snowmelt dominates, but there is a significant contribution to the flood record from North American Monsoon thunderstorms. The magnitude of the 22 July 1954 flood peak is unprecedented, not the occurrence of a flood during the North American Monsoon season.

The Sybille Creek flood of 20 August 1990 in Wyoming, which produced an upper tail ratio of 35 (Table 1), reprises themes from the Heppner flood (see G. F. Ritz's summary in Jordan & Combs, 1996). The Sybille Creek flood, like the Heppner flood and many of the large upper tail ratio events, was the product of an intense hailstorm in which extreme rainfall and exceptional hail falls were in close proximity. Accumulations of hail reached several feet in portions of the Sybille Creek watershed.

The Teton River of Montana was the setting for a strange hydrologic flood, with a record peak on 8 June 1964 producing an upper tail ratio of 21 (Table 1; Boner & Stermitz, 1967). Although geographically close to the Teton River of Idaho (Table 1), the Montana river is tributary to the Missouri and Mississippi, flowing to the Gulf of Mexico; the Idaho Teton River is a tributary to the Snake and Columbia, flowing to the Pacific Ocean. In a supplement to Boner and Stermitz (1967), R. A. Dightman notes that "an item of more than passing interest was the lack of thunderstorms in the heavy precipitation region." This observation reflects the view of the hydrometeorological community in the U.S. Weather Bureau and the flood hydrology community of the USGS from the 1930s starting with Robert Follansbee and Nathan Grover through the 1960s that thunderstorms play an important role as agents of extreme flooding. In the 1920s and 1930s, USGS hydrologists documented extreme rainfall and flood runoff from *cloudburst* thunderstorms (see, e.g., Follansbee & Jones, 1922; Woolley, 1946). The USGS view of floods during the 1950s and 1960s was clearly influenced by Luna Leopold (Chief Hydrologist from 1956 to 1966). Leopold's background in meteorology and his early research on cloudburst floods in the southwestern United States (Leopold, 1942, 1946) gave him a deep appreciation for the role of thunderstorms as flood agents in the United States.

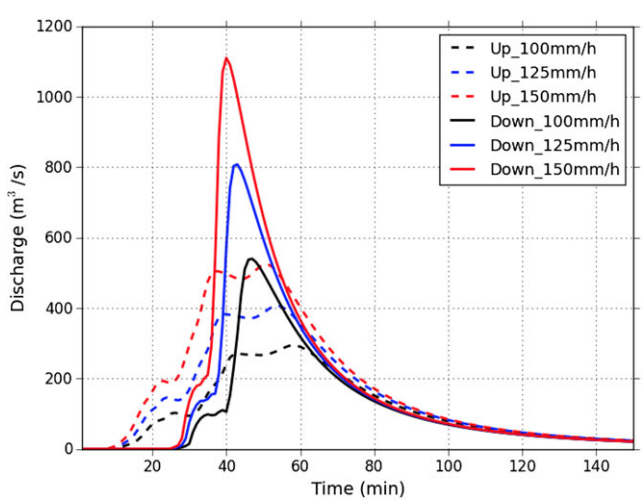

Figure 14. KINEROS-2 model reconstruction of 14 June 1903 flood in Balm Fork. Results for rainfall rates of 100, 125, and 150 mm/hr for downstream storm motion (solid lines) and upstream storm motion (dashed lines).

The Ashland Creek floods in the mountains of southern Oregon (Figure 2 and Table 1) on 15 January 1974 reflect a prominent setting for extreme

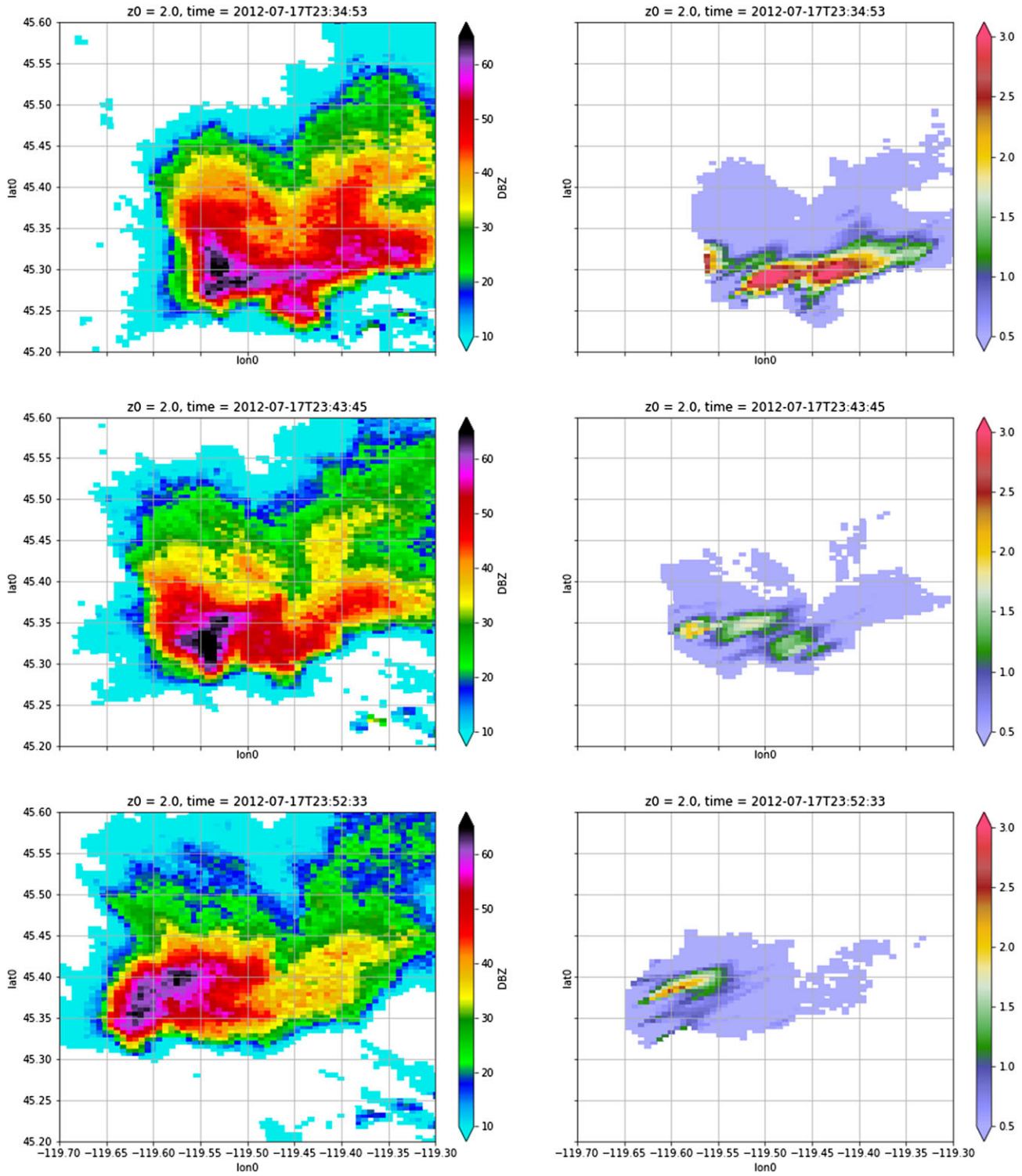

Figure 15. Reflectivity fields (dBZ; left) and specific differential phase shift fields (degrees km^{-1}; right) at 2 km above ground level for the 17 July 2012 storm at 23:34, 23:43, and 23:52 UTC (prepared using the Python ARM Radar Toolkit; see Helmus & Collis, 2016).

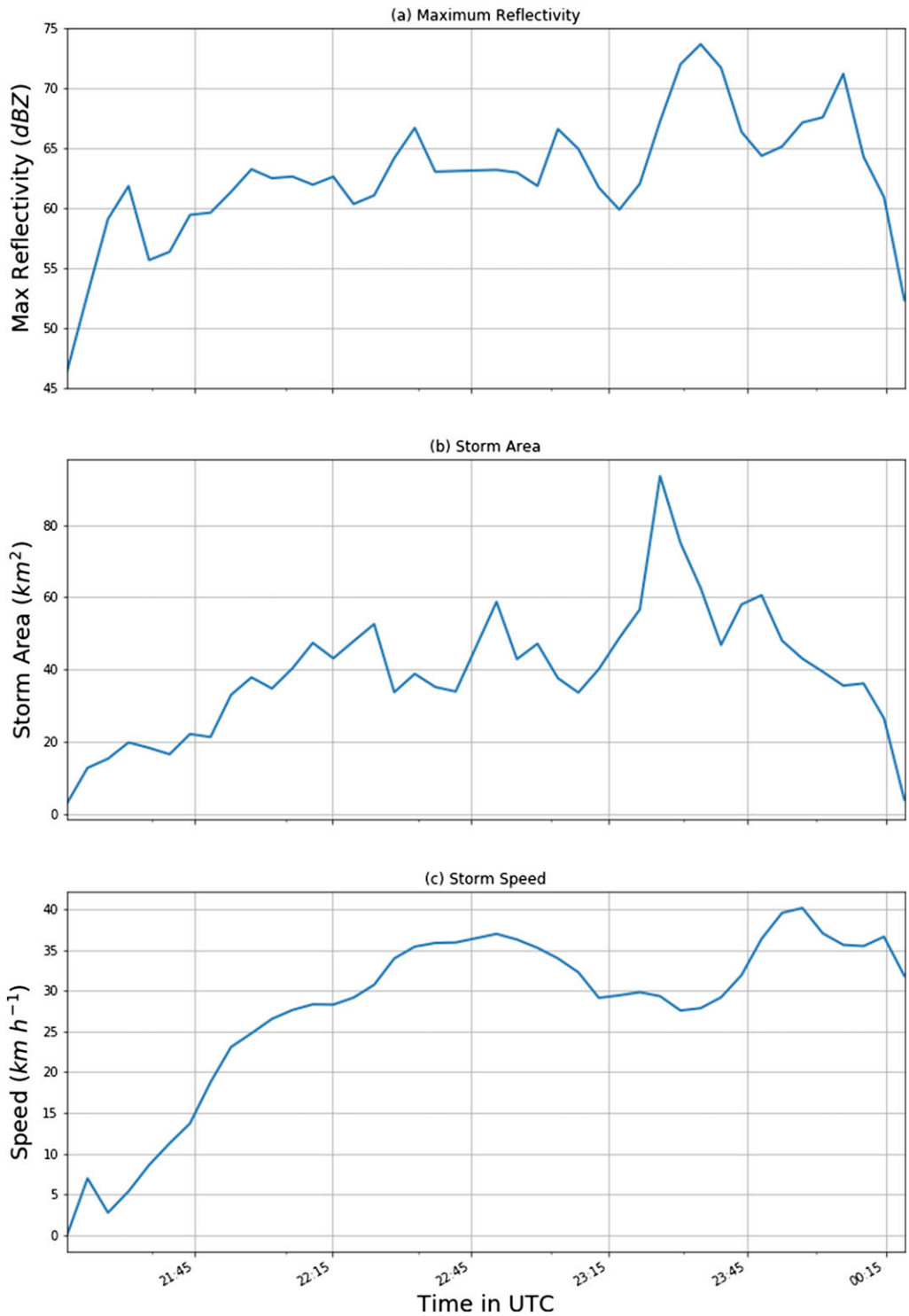

Figure 16. Time series of maximum reflectivity (top; dBZ), storm size (middle; km^2); and storm speed (bottom; km/hr) for the 17 July 2012 storm (see Figure 15).

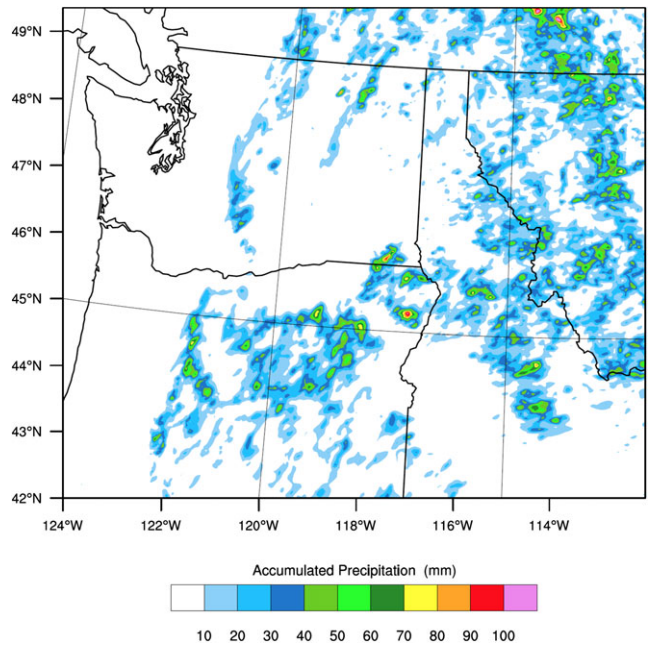

Figure 17. Weather Research and Forecasting storm total rain (mm) for the 14 June 1903 Heppner storm.

flooding in northern California and the Pacific Northwest. Heavy orographic rainfall from extratropical systems, combined with rapid snowmelt associated with strong warm air advection (Colle & Mass, 2000), can produce exceptionally large flood peaks in the region. These events are linked to concentrated low-level transport of water vapor in *atmospheric rivers* (in contrast to the Blue Mountain floods in eastern Oregon, as discussed below).

A cluster of stations with upper tail ratio greater than 5 is located in southern California along the San Gabriel Mountains north of Los Angeles (Figure 2). This cluster is dominated by the 3–4 March 1938 flood for which nine stations reported record floods with upper tail ratios greater than 5; the San Antonio Creek station has an upper tail ratio of 24 (Table 1). The flood devastated portions of Los Angeles County and resulted in more than 100 fatalities. The March 1938 storms provide additional insights to the 2,300-m hypothesis for extreme rainfall and flooding. Rainfall at Kelly's Kamp at an altitude of 2,530 m in the headwaters of San Antonio Creek was 810 mm during the period from 27 February to 4 March. The March 1938 flood, like the January 1974 flood in southern Oregon and many of the major floods along the west coast of the United States, was produced by major Pacific storms during the winter season (see, e.g., Barth et al., 2017; Dettinger et al., 2011; Neiman et al., 2011).

The San Emigdio Creek flood on 5 August 1961 along the north slope of the San Emigdio Mountains north of Los Angeles, California, (Table 1) is both strange and mysterious. The seasonal distribution of flood peaks in San Emigdio Creek is dominated by winter storms, resulting in the smallest value of the record flood seasonality S_i, 0.66, among the events with upper tail ratio greater than 20. Exceptional floods from summer cloudbursts in San Emigdio Creek have, however, been described by Douglas (1908), who notes that a high-gradient tributary is named Cloudburst Canyon.

Texas has a well-recognized history of extreme floods and it is a region with a high concentration of large upper tail ratio floods (Figure 2) and a diverse population of flood mechanisms (Baker, 1975; Nielsen-Gammon et al., 2005; O'Connor & Costa, 2004a; Patton & Baker, 1976; Smith et al., 2000; Zhu & Quiring, 2013), including severe thunderstorms as noted above for the May 1935 Seco Creek flood. The Medio Creek flood of 22 September 1967 is the only Texas flood and the only tropical cyclone flood to produce an upper tail ratio greater than 20 (Table 1). Hurricane Beulah produced extreme rainfall in south Texas, with a maximum accumulation of 695 mm in 36 hr in Pettus, Texas, which is located in the upper portion of the Medio Creek watershed. More extreme rainfall accumulations have been recorded in Texas. The 24-hr record rainfall for the United States of 1,090 mm occurred at Alvin, Texas, on 25–26 July 1979 (Tropical Storm Claudette), returning to southeastern Texas the record that was held for many years by the 965 mm at Thrall, Texas, on 10 September 1921 (Hurricane 1 of 1921) near the Balcones Escarpment (Caracena & Fritsch, 1978; Smith et al., 2000). Hurricane Harvey produced 5-day accumulations (25–29 August 2017) of 1,250 mm in Houston, the largest for a single storm in the conterminous United States (see Emanuel, 2017, for climatological assessments of tropical cyclone rainfall frequency in Texas).

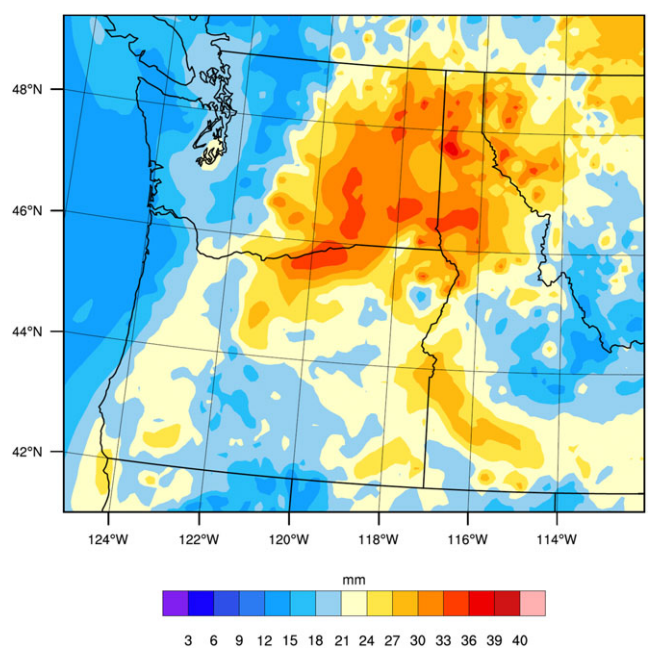

Figure 18. Weather Research and Forecasting precipitable water (mm) at 20:00 UTC on 14 June 1903.

Floods with large upper tail ratio in the United States east of the Mississippi River are dominated by tropical cyclones (Figure 2). Five hurricanes account for more than 40% of the 123 stations east of 90.2° west with upper tail ratios greater than 5. Hurricane Diane in August 1955 accounted for 19 of the stations, the Great New England Hurricane of September 1938 accounted for 9, Hurricane Floyd in September 1999 and Hurricane Alberto

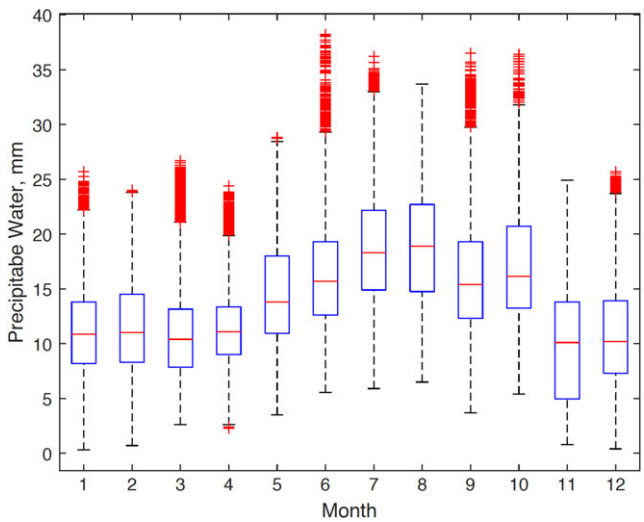

Figure 19. Monthly boxplots of precipitable water (mm) for Heppner, Oregon, based on GPS precipitable water measurements. Whiskers represent 0.1 and 0.9 quantiles of precipitable water. Outliers are denoted by red plus signs.

in July 1994 accounted for 8, and Hurricane Agnes in June 1972 accounted for 7. Tropical cyclones account for two thirds of the 123 record flood events east of the Mississippi River with upper tail ratio greater than 5.

There is large diversity in rainfall and flooding from this collection of tropical cyclones. Diane produced abnormally large rainfall rates and retained tropical characteristics throughout its passage through the northeastern United States. Most tropical cyclones that reach the mid-Atlantic and New England experience extratropical transition through interaction with mid-latitude extratropical systems (Atallah & Bosart, 2003; Hart & Evans, 2001). Hurricanes Agnes (DiMego & Bosart, 1982), Floyd (Atallah & Bosart, 2003; Colle, 2003), and Irene (Liu & Smith, 2016) reflect rainfall distributions characteristic of extratropical transition events, producing extensive regions of heavy rainfall. For Hurricane Agnes, there were 125 stations reporting record floods, compared with 68 stations from Hurricane Diane.

Orographic thunderstorms in the central Appalachians provide another class of strange floods (Giordano & Fritsch, 1991; Konrad, 2001; Smith et al., 2011b). Villarini and Smith (2010) show that orographic thunderstorms are largely unrepresented in the upper tail of flood peaks from USGS stream gaging stations in the central Appalachians, but they dominate the envelope curve of flood peaks (Hicks et al., 2005; Miller, 1990; Smith et al., 1996, 2011b), based on indirect discharge estimates. Warm season thunderstorms account for a handful of the flood events with upper tail ratio greater than 5 in the eastern United States (Figure 2), including the Smethport flood of 18–19 July 1942, the Rapidan flood of 27 June 1995, and the northern Ohio flood of 4 July 1969 (see Smith et al., 2011b). The Smethport storms of 18–19 July 1942 in northwestern Pennsylvania (Eisenlohr, 1952; Smith et al., 2011b) produced the world record rainfall of 780 mm in 4.75 hr and flood peaks (based on indirect discharge measurements) that control the eastern U.S. envelope curve at basin scales less than 100 km^2. They were exceptionally intense thunderstorms; in transmitting rainfall analyses for the (Eisenlohr, 1952) report, J. E. Stewart notes that "everyone contacted emphasizes that this was by far the worst electrical storm they ever witnessed, not only for length of time involved, but particularly for intensity at the time any one of the thunderstorms were passing" (J. E. Stewart, letter to Merill Bernard, Supervising Hydrologist, U. S. Weather Bureau, 18 September 1942).

4. The Strangest Floods: The Blue Mountains of Eastern Oregon

In this section, we examine the hydrology, hydrometeorology, and hydroclimatology of extreme floods in and adjacent to the Blue Mountains of eastern Oregon. The region, which is located along the southern boundary of the Columbia Plateau, is the setting for the largest upper tail ratio flood in the USGS stream gaging record (excluding the Teton dam failure flood peak) and two miscellaneous flood peaks which control the U.S. and world envelope curves at 10–40 km^2 scale (see Figure 12 for locations). The Columbia Plateau is an apt setting for strange floods. The largest floods on Earth, the Lake Missoula floods, shaped the landscape of eastern Washington and Oregon 15,000–13,000 years ago (see, e.g., O'Connor & Baker, 1992). In a paper that examines the properties of strange floods, this section is designed to highlight key features of the strangest.

4.1. Heppner Flood

What is known about rainfall and streamflow from the Heppner flood comes largely from USGS hydrologist John Whistler (Whistler, 1903; a synopsis of Whistler's report appears in Murphy, 1904). Whistler introduced the event by noting that "On June 14 there occurred at Heppner, Oregon, one of those storms so peculiar to arid regions and generally referred to as 'cloud-bursts.' A knowledge of the amount of precipitation and the area covered, seemed to be of special interest in the consideration of irrigation projects … I have made an investigation to determine more or less approximately the facts as to run-off and area of storm, … together with other facts which I have thought may be of general interest."

The 1903 flood in Willow Creek reached Heppner (see Figure 13) at 5:30 P.M. local time and resulted in more than 250 fatalities in less than 1 hr (Whistler, 1903). The time of *first water* and flood crest were almost simultaneous for Willow Creek at Heppner. The floodwaters came principally from Balm Fork, a 68-km^2 tributary of Willow Creek. The 14 June 1903 flood peak in Balm Fork (Figure 1) was estimated to be 36,000 cfs (1,019 m^3/s),

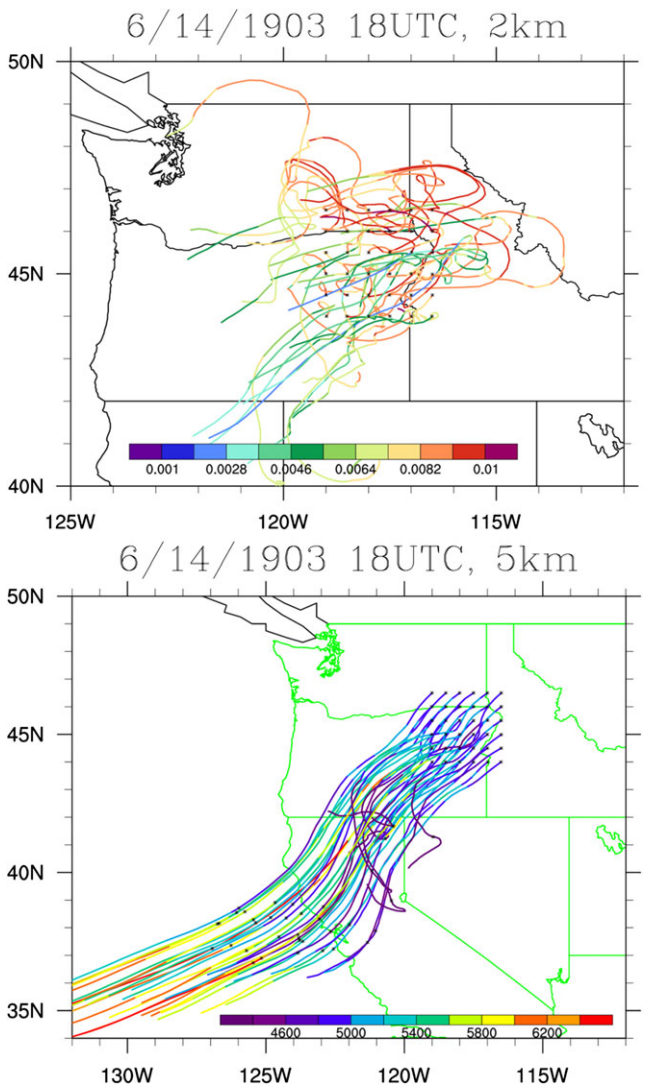

Figure 20. Lagranto back-trajectories for 5 (elevation) and 2 km (specific humidity) ending at 18:00 UTC on 14 June 1903 at 2-km elevation (top) and 5-km elevation (bottom). Color coding for the 2-km analyses (top) is specific humidity of the parcel (kg/kg); color coding for the 5-km analyses (bottom) is elevation (m).

198 times larger than the sample 10-year flood magnitude (Table 1). The unit discharge for the Balm Fork peak is 15 $m^3 \cdot s^{-1} \cdot km^{-2}$. The estimated flood peak in Willow Creek at Heppner (Figure 13), which has a drainage area of 251 km^2, was also 36,000 cfs (1,019 m^3/s), 78 times larger than the sample 10-year flood magnitude. The peak discharge values for Balm Fork and locations downstream on Willow Creek were based on slope-area computations (Whistler, 1903; see Levish & Ostenaa, 1996, and Byrd, 2009, for additional discussion and photographs).

The Heppner flood was the product of a hailstorm that moved down the valley of Balm Fork. Whistler summarized the Heppner storm as follows: "residents of the valley and other eyewitnesses agree that the storm moved down the valley of Balm Fork … and had entirely passed away in less than a half-hour. Accounts agree that it extended from the foot of the Blue Mountains, 8 to 10 miles distant, to and a short distance across Willow Creek valley, covering a strip, variously said to be from 2 to 4 miles wide." Whistler concluded that rainfall at any location was no more than 30 min in duration.

We use the KINEROS-2 model, which has been applied to flood modeling for arid/semiarid watersheds (see Goodrich et al., 2011, and Morin et al., 2006, for additional details), to examine the dependence of flood peaks on structure, motion, and rainfall rates from the Heppner storm. We use a 30-m USGS digital elevation model for implementing KINEROS-2 for Balm Fork (see Figure 13), with the 68 km^2 watershed partitioned into 117 overland flow planes and 77 channel elements. Whistler describes the hillslopes of Balm Fork as *rocky and barren hills which carry off water very rapidly*. Based on Whistler's analyses, we assume low values of saturated hydraulic conductivity (0.001 mm/hr) and assume channel roughness of 0.04, in line with Whistler's assumptions for peak discharge computations. Channel geometry is also based on Whistler's analyses. We use the Automated Geospatial Watershed Assessment algorithms (Goodrich et al., 2011) with conventional geographic information systems data layers to specify model parameters that are not fixed by observations from the 1903 flood (see Goodrich et al., 2011, and Kampf & Burges, 2007, for discussion of errors inherent in modeling analyses with KINEROS and similar hydrologic models).

As noted above, Whistler's report provides an idealized depiction of storm size and motion for the June 1903 Heppner storm. From these, we develop rainfall rate fields for the June 1903 storm, representing the storm as a circular region of uniform rainfall rate and radius 5 km, which moves down the Balm Fork watershed, that is, from southeast to northwest, at a speed resulting in 30-min rainfall durations for locations in Balm Fork. A model simulation based on rainfall fields consistent with Whistler's characterizations of storm structure and motion and with a rainfall rate of 135 mm/hr reproduces the peak discharge and timing of peak response (Figure 14). In Figure 14, we also show sensitivity of model analyses to assumptions on rainfall rate and storm motion. Rainfall rates at the 30-min time scale have return intervals that are large for eastern Oregon (the 100-year 30-min rainfall rate for Heppner is approximately 40 mm/hr, based on Oregon Department of Transportation analyses), but are not uncommon for hailstorms producing extreme flooding in small watersheds (Doswell et al., 1996; Smith et al., 2001). As will be shown below, 30-min rainfall rates needed to produce peak flood response in Balm Fork are smaller than those required to produce estimated flood peaks for the Meyers Canyon and Lane Canyon events, based on similar modeling analyses.

Model simulations suggest that storm motion played a significant role in determining flood peak magnitude in Balm Fork. For a circular storm of radius 5 km (approximately 50 km^2 in area) moving down Balm Fork with a rain rate of 100 mm/hr, the peak discharge at the outlet is 931 m^3/s (Figure 14). For a 5-km radius storm moving up the basin at the same speed, the peak discharge is 485 m^3/s. In both cases, the storm total rainfall

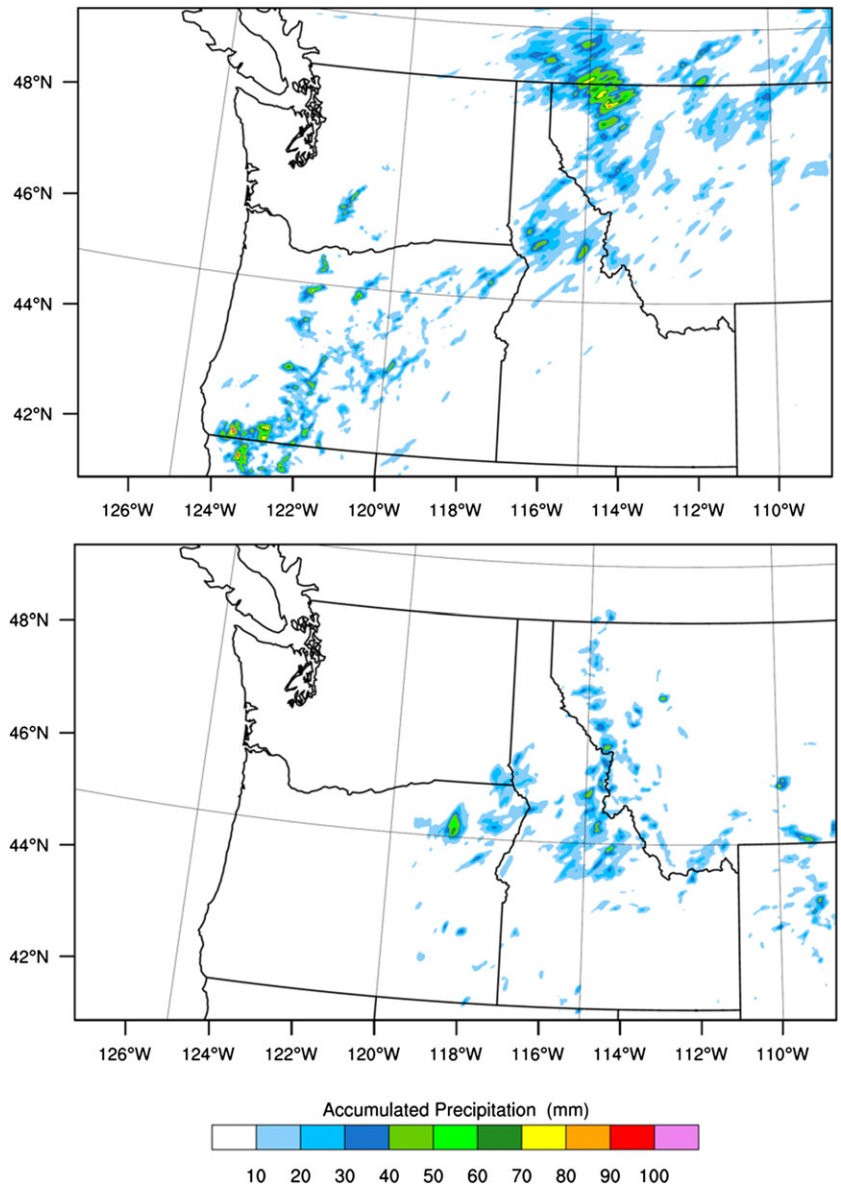

Figure 21. Weather Research and Forecasting storm total rain (mm) for the 13 July 1956 Meyers Canyon storm (top) and 26 July 1965 Lane Canyon storm (bottom).

and distribution of instantaneous rain rates over the watershed are the same. We note below that for typical thunderstorm environments in eastern Oregon, *steering winds* have a southerly component. The combination of a preferred pattern of storm motion and sensitivity of extreme flood peaks to storm motion relative to the drainage network suggests that topographic heterogeneity can play an important role in determining regional variation of the upper tail of flood distributions. Not only are the Blue Mountains a distinctive setting for extreme floods, but locations like the Willow Creek watershed will have enhanced potential for flooding due to properties of storm motion relative to the drainage network structure (see Soderholm et al., 2014, for related analyses of the Black Hills region of South Dakota). *Hots pots* of catastrophic flooding in mountainous terrain have been examined in geomorphic studies of extreme floods and debris flows (Eaton et al., 2003; see also Hack& Goodlett, 1960) and in paleoflood reconstructions (Harden et al., 2011).

The June 1903 storm is not entirely unprecedented for the Willow Creek watershed. Major flooding occurred on 25 May 1971 from a storm in the Shobe Creek tributary to Willow Creek. Shobe Creek enters Willow Creek near Heppner downstream of the confluence of Balm Fork (Figure 13). A USGS indirect discharge measurement for the 25 May 1971 flood peak in Shobe Creek puts the peak discharge at 172 m^3/s from a drainage area

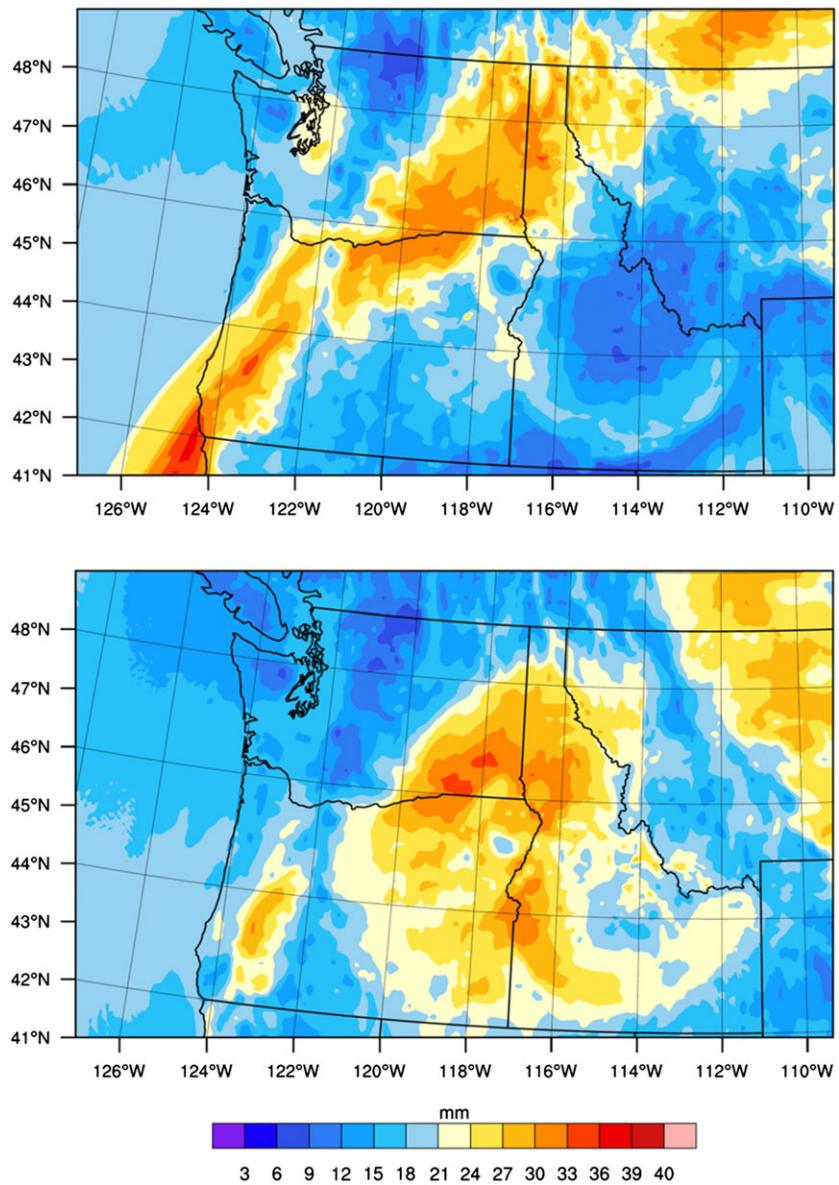

Figure 22. Weather Research and Forecasting pre–storm precipitable water (mm) for Meyers Canyon storm (13 July 1956 at 18:00 UTC; top) and Lane Canyon storm (26 July 1965 at 180:0 UTC; bottom).

of 16 km^2. Like the Heppner storm, the 1971 Shobe Creek storm was a hailstorm with large hail and heavy rain. The Heppner Gazette-Times of 27 May 1971 reported that "a sudden hard torrent of rainfall with a brief dash of lightning struck about 3:15 PM … The alfalfa and grain crops were badly damaged by hail [accumulations of four inches] … This was followed by a hard rain."

The 1903 Heppner storm was an unusually intense hailstorm. Whistler notes that rainfall was preceded by a "phenomenal fall of hail" and that "grim evidence of the amount of hail is that, while most of the bodies being recovered on the fifth day were already badly decomposed, one was occasionally found almost perfectly preserved in a large drift of hail" (Whistler, 1903). Eyewitness accounts of the storm suggest that it may have been a supercell thunderstorm, with one observer describing the storm as "a broad black streak almost straight north of us which seemed to be coming straight down from the sky" (Byrd, 2009). Supercells are the most intense storms in the spectrum of thunderstorms and have produced some of the largest short-term rainfall rates in the United States (Bunkers & Doswell, 2016; Doswell et al., 1996; Giordano & Fritsch, 1991; Nielsen et al., 2015; Rogash & Racy, 2002; Smith et al., 2001). Recent studies are shedding light on both the dynamics

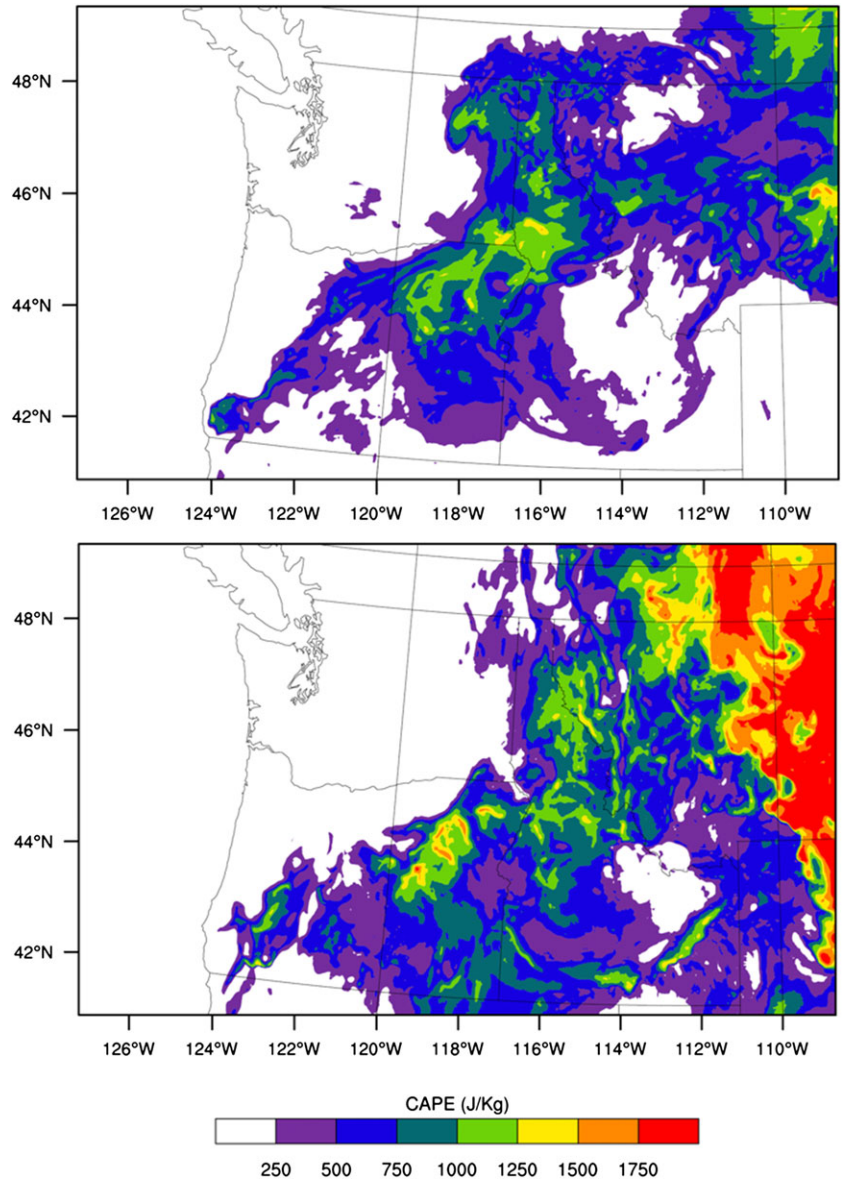

Figure 23. Pre–storm CAPE fields (J/kg) for the 1956 Meyers Canyon storm (top) and 1965 Lane Canyon storm (bottom). CAPE = Convective Available Potential Energy.

and climatology of supercells in mountainous terrain (see, e.g., Bosart et al., 2006; Markowski & Dotzek, 2011; Scheffknecht et al., 2017).

The NCDC Storm Events database and USGS annual peak discharge record contain no events in the Blue Mountains region that are comparable to the Heppner storm during the period of modern observing records, which we take to begin with deployment of the WSR-88D radar network in the mid-1990s. One of the closest analogs is a hailstorm which passed to the west of Heppner on 17 July 2012. Like the Heppner storm, the 17 July 2012 storm was an intense hailstorm embedded in a multiday period of thunderstorm activity. Polarimetric radar analyses of the 17 July 2012 storm (Figures 15 and 16) illustrate structure, motion, and evolution of severe hailstorms. The 17 July 2012 storm initiated along the Blue Mountains, as the Heppner storm did, and generally moved from southeast to northwest as the Heppner storm did, passing along the western margin of Balm Fork and producing an (small) annual peak discharge at the Balm Fork stream gaging station. The NCDC Storm Events database includes reports of heavy rain, more than 1 in. in 1 hr as the storm passed Shobe Canyon (approximately 23:35 UTC; see Figure 15), and multiple reports of 1- to 2-in. hail after 23:15 UTC.

The 17 July 2012 storm was anything but uniform in its structure, motion, and evolution (Figures 15 and 16; storm tracking analyses are based on the Thunderstorm Identification, Tracking, Analysis, and Nowcasting algorithms of Dixon & Wiener, 1993). The peak value of reflectivity at 23:35 UTC was 73 dBZ (Figure 16) and the paired value of differential reflectivity was close to 0, indicative of large hail. The storm decayed rapidly around 00:15 UTC, as it moved north of Heppner toward the Columbia River. Merger of two cells around 23:20 UTC resulted in an increase in storm size and the spike in maximum reflectivity around 23:30 UTC (Figure 16). Storm area (determined by 45 dBZ threshold) ranged from 20 to 80 km^2, with a sharp increase in size after merger of the two storm elements. Storm motion was determined by storm propagation and advection by steering winds, with storm speed fluctuating around 30 km/hr. Polarimetric radar measurements of specific differential phase shift (Figure 15; see Romine et al., 2008, and Kumjian, 2013), which is approximately linearly related to rainfall rate and immune to the effects of hail, illustrate the rapidly varying temporal and spatial structure of rainfall. Polarimetric measurements indicate short periods and small areas with rainfall rates of approximately 100 mm/hr, corresponding to specific differential phases shift values larger than 3°/km (Figure 15). Analyses of the 17 July 2012 storm based on modern observing capabilities point to the large uncertainties in characterizing rainfall from the Heppner storm (as described above) and from the Meyers Canyon and Lane Canyon storms (as detailed below).

4.2. Meyers Canyon and Lane Canyon Floods

The Meyers Canyon and Lane Canyon floods in eastern Oregon (see Figure 12 for locations) control the U.S. and world envelope curves at basin scales ranging from 10 to 40 km^2. Both the Meyers Canyon and Lane Canyon flood peaks resulted in larger unit discharge peaks than the Balm Fork flood of 1903. Perhaps more than any other floods in the USGS record, they highlight the uncertainties in measurement of extreme flood magnitudes (see Costa & Jarrett, 2008, for discussion of the accuracy of both the Meyers Canyon and Lane Canyon flood peak measurements). Unlike Balm Fork, no systematic stream gaging has been established by the USGS at either the Meyers Canyon or Lane Canyon sites.

The estimated peak discharge for the Meyers Canyon flood of 13 July 1956 (Hendricks, 1964; see Figure 12 for the location) is 1,543 m^3/s (54,500 cfs) from a drainage area of 32.9 km^2 (12.7 mi^2), resulting in a unit discharge of 46.9 m$^3\cdot$s$^{-1}\cdot$km^{-2} (4,290 cfs per square mile). A rough estimate of the upper tail ratio for the Meyers Canyon flood peak is 600, which we obtain using the regression equations developed by Cooper (2006) to estimate the 10-year flood. The accuracy of the flood peak estimate has been, and continues to be, a topic of debate (see, e.g., Levish & Ostenaa, 1996). Following site investigation and review of previous analyses, Costa and Jarrett (2008) concluded that the "54,500 cfs peak should remain as published, but the rating should be downgraded from fair to estimate", which implies that the uncertainty in the estimate is larger than 25% and potentially as large as 50%. Even if peak discharge were at the lower end of the uncertainty range, it would rank with the June 1903 Balm Fork peak as one of the most extreme flood measurements in the United States.

An observer reported that rainfall in Meyers Canyon lasted for approximately 2 hr (4:30 P.M. local time to 6:30 P.M. local time), but a large part of the precipitation occurred in approximately 30 min with rainfall in *cloudburst proportions* (Hendricks, 1964). There are no reports of lightning or hail for the storm, but the area is remote and there was only a single report from the area of heavy rainfall in Meyers Canyon. Downscaling simulations with WRF (as detailed below) indicate that the storm environment, like the Heppner and Lane Canyon settings, would support the development of thunderstorms.

KINEROS-2 simulations were carried out for the Meyers Canyon flood, emulating the procedures used for the Heppner flood. A rainfall rate of 250 mm/hr from a storm moving down the Meyers Canyon drainage network at 32 km/hr produces a 30-min period of rainfall for locations in the watershed and a peak discharge close to 1,543 m^3/s at the basin outlet. The peak discharge that results from the same scenario, except for motion up the watershed, is 637 m^3/s at the basin outlet.

The Lane Canyon flood on 26 July 1965 (see the summary by D. D. Harris in Rostvedt, 1970) produced a flood peak of 807 m^3/s (28,500 cfs) from a drainage area of 13.1 km^2 (5.04 mi^2), resulting in a unit discharge of 61.6 m$^3\cdot$s$^{-1}\cdot$km^{-2} (5,650 cfs per square mile). Using the regression equations of Cooper (2006) to estimate the 10-year flood magnitude in Lane Canyon, it yields an upper tail ratio of approximately 800 for the 1965 flood. Accuracy of the estimated peak discharge for the Lane Canyon flood, like the Meyers Canyon peak, has sparked controversy. Costa and Jarrett (2008) concluded that "the peak discharge for the flood in Lane Canyon was likely 28,500 cfs, but the rating should be changed from fair to poor." Also, like the Meyers Canyon flood,

the Lane Canyon flood peak would be one of the most extreme in the United States even if peak discharge were at the low end of discharge values associated with measurement uncertainty.

The Lane Canyon flood was caused by a *severe cloudburst*, with observers reporting that most of the rain fell in 30 min (Rostvedt, 1970). Observers also reported that large accumulations of hail accompanied the storm. The spatial extent of erosional features indicated that the entire watershed produced heavy runoff. KINEROS-2 simulations for a storm moving down the watershed with a rain rate of 360 mm/hr produce a peak discharge close to the observed maximum of 807 m^3/s. The same storm moving up the watershed produces a peak discharge of 330 m^3/s.

Using comparable model implementations, we conclude that the rainfall rates needed to produce the estimated peak discharge values for the 1956 Meyers Canyon and 1965 Lane Canyon flood peaks are larger than those required to produce the Balm Fork peak. The rainfall rates are not unprecedented for thunderstorms in the United States (Engelbrecht & Brancato, 1959; Smith et al., 2011b), but they are exceptionally rare. These analyses highlight the large uncertainties in discharge estimates for the Meyers Canyon and Lane Canyon flood peaks.

4.3. Hydrometeorology of Blue Mountain Thunderstorms

In this section we examine the hydrometeorology of extreme rainfall in the Blue Mountains, principally through analyses of WRF simulations of the June 1903 Heppner storm, the July 1956 Meyers Canyon storm, and the July 1965 Lane Canyon storm. We first note that the climatology of thunderstorms, as represented by cloud-to-ground (CG) lightning observations from National Lightning Detection Network observations over eastern Oregon and Washington state, exhibits pronounced spatial heterogeneity, with a peak occurrence over the Blue Mountains (Figure 12). Mean annual CG flash density ranges from 0.8 strikes km^{-2} to 1.2 strikes km^{-2} along the Blue Mountains. Mean flash density decreases sharply in the Columbia Plateau region to the northwest, with a mean flash density of 0.5 km^{-2} in Heppner and values less than 0.2 km^{-2} over the Columbia River. For the WRF simulations of thunderstorms over the Blue Mountains region, we use three one-way nested domains of 27-, 9-, and 3-km resolution. Model physics options include (1) Noah land surface model, (2) WSM6 microphysics scheme, (3) Yonsei planetary boundary layer scheme, (4) Dudhia shortwave radiation scheme, and (5) the Rapid Radiative Transfer Model scheme for longwave radiation (see Smith & Baeck, 2015, for similar implementations). For the 1903 and 1965 simulations, Twentieth Century Reanalysis fields were used for initial and boundary conditions. For the 1956 Meyers Canyon storm, National Centers for Environmental Prediction-National Center for Atmospheric Research reanalysis fields were used for initial and boundary conditions.

The simulation for the June 1903 storm captures the occurrence of heavy rainfall along the Blue Mountains (Figure 17). Peak rainfall accumulations of 80–100 mm are located in the higher elevations of the Blue Mountains east of Heppner, with rainfall near Heppner exceeding 50 mm. Broad features of the simulated rainfall are consistent with the available evidence on rainfall distribution. The Daily Weather Map for 14 June 1903 shows an area of rainfall extending through eastern Oregon from southern Idaho into Washington state. Simulated rainfall is organized into small convective elements (median size of 50 km^2 based on Thunderstorm Identification, Tracking, Analysis, and Nowcasting storm tracking analyses of simulated reflectivity fields) that have a south-to-north component of storm motion. The synoptic scale environment of the Heppner storm was characterized by an upper level trough in the eastern Pacific off the coast of Canada and a high pressure ridge extending from Mexico through the Rocky Mountain region into Canada.

Water vapor for the Heppner storm was exceptionally large for the Blue Mountains region. The simulated precipitable water field at 18:00 UTC on 14 June 1903 shows a large area of extreme precipitable water values centered on the Columbia Plateau (Figure 18). Precipitable water values approached 40 mm, which is close to the upper bound of precipitable water observations from the GPS precipitable water station near Heppner, Oregon (Figure 19). Paired with elevated precipitable water, the WRF simulation indicates that the convective potential was large over a significant portion of the Columbia basin margin, with peak values of Convective Available Potential Energy (CAPE) approaching 1,000 J/kg adjacent to the highest elevation region of the Blue Mountains in eastern Oregon (figure not shown). The 0.9 quantile of CAPE for June at the nearest radiosonde site in Spokane, Washington, is less than 300 J/kg and the record CAPE for June is less than 2,000 J/kg based on observations extending back to 1948. Like the Meyers Canyon and Lane Canyon events, the Heppner storm occurred during a multiday period of elevated humidity and convective potential. A tornado was reported the day after the Heppner storm (15 June 1903) approximately 57 mi northeast of Heppner (Byrd, 2009).

A central issue surrounding record flooding in the arid region of eastern Oregon is the source of water vapor that fueled the storms. Back-trajectory analyses using the WRF simulation and the LAGRANTO algorithm (Figure 20; see Sprenger & Wernli, 2015, for details on back-trajectory computations) illustrate a complex pattern of water vapor transport. Back-trajectory analyses ending at 18:00 UTC on 14 June and at 5-km elevation show parcel transport from southwest to northeast from the Pacific to eastern Oregon. Similar analyses terminated at 2-km elevation over eastern Oregon show diverse flow paths to the Blue Mountains flood region, including transport from Idaho and Montana. Unlike major flood events in the Pacific Northwest from cold-season extratropical systems, the Heppner storm was not the product of a well-defined *river* of atmospheric water vapor transport (Colle and Mass ; 2000; Neiman et al. 2011).

Simulated rainfall for the July 1965 Lane Canyon storm includes a local maximum (70 mm) in rainfall on the northern margin of the Blue Mountains (Figure 21). The U.S. Weather Bureau surface map shows an area of rainfall that matches the pattern of rain extending from northeastern Oregon into northwestern Montana. Like the Heppner storm, the Lane Canyon storm environment was characterized by extreme values of precipitable water (Figure 22) and elevated CAPE (Figure 23) along the margins of the Columbia Plateau. The synoptic environment for the Lane Canyon storm was similar to the Heppner storm, with comparable trough/ridge positions.

The WRF simulation of the 1956 Meyers Canyon storm (Figure 21) captures key elements of rainfall for the day, as represented by the U.S. Weather Bureau surface map, which indicated areas of rainfall extending along the California-Oregon border northeastward into Oregon. A local maximum in the WRF simulation is located southwest of Meyers Canyon. The largest observed rainfall for the day was located along the Montana-Canada border in the area of the local maximum from the WRF simulation. Precipitable water and CAPE fields for the Meyers Canyon storm are similar to the Heppner and Lane Canyon storm environments (Figures 22 and 23). The synoptic environment for the Meyers Canyon storm was characterized by an upper level trough that extended along the northern California coast into Oregon and Washington.

The synoptic environment of the 17 July 2012 storm, the modern analog of the Heppner storm, was similar to the Meyers Canyon storm, with an upper level trough that extended from north to south along the Pacific Northwest during the period from 16 to 18 July 2012. Like the Heppner, Lane Canyon, and Meyers Canyon environments, elevated precipitable water characterized a 4-day period of thunderstorms in eastern Oregon and Washington, with peak values of precipitable water at Heppner reaching 35 mm. The period from 15 to 18 July 2012 contains 4 of the top 10 thunderstorm days in the Blue Mountain region since 1992, based on National Lightning Detection Network CG flash density observations. In addition to copious lightning, storm reports for the period include large hail, funnel clouds, heavy rain, flash floods, debris flows, and landslides for eastern Oregon and Washington (NCDC Storm Reports).

Downscaling simulations with WRF provide insights to common elements of storm environment for the record floods of Eastern Oregon, reducing some of the strangeness of these events. The Heppner, Meyers Canyon, and Lane Canyon storms exhibited similar structure of steering level winds, precipitable water, and CAPE fields, which play an important role in determining the potential for extreme rainfall rates (see Bunkers & Doswell, 2016; Nielsen et al., 2015; and Nielsen & Schumacher, 2018, for additional discussion). Model simulations also suggest that storm size and motion were similar for the three events. Storm size and motion dictate that extreme floods in the Blue Mountains region will concentrate at basin scales smaller than 100 km^2. The Heppner, Meyers Canyon, and Lane Canyon floods are rare, but the environmental conditions for multiday thunderstorm outbreaks in the Blue Mountains region are recurring phenomena, as illustrated by the July 2012 storm period. The WRF simulations suggest that some aspects of these major thunderstorm outbreaks are predictable, based on information contained in conventional reanalysis fields. A mystery that remains is how some storms that form during major thunderstorm outbreaks can produce rainfall rates that result in Heppner-like flood peaks.

5. Summary and Conclusions

We examine the upper tail of flood peaks in the United States, principally through analyses of the upper tail ratio for annual flood peak observations from more than 8,000 USGS stream gaging stations. Our major findings are the following:

- With a few caveats, the 14 June 1903 Heppner, Oregon, flood is the strangest flood in the systematic stream gaging record of the United States. The upper tail ratio for Balm Fork at a drainage area of 68 km^2 is 200, implying that the June 1903 flood peak was 200 times larger than the 10-year flood for the watershed. The 1903 Heppner flood is paired with two other extreme floods in the Blue Mountain region of eastern Oregon. Based on indirect discharge estimates in the miscellaneous flood record of the USGS, the July 1956 Meyers Canyon flood and the July 1965 Lane Canyon flood control the U.S. and world flood peak envelope curves at basin scales between 10 and 40 km^2 (see Costa, 1987a, and Herschy, 2003).
- A fundamental distinction between record floods in the United States and the broader population of annual flood peaks is in their seasonal distribution. Annual flood peaks in the United States have winter/spring and warm season maxima in the seasonal rate of occurrence. For record floods, there is a striking shift toward warm season flooding and an increase in occurrence probability during the tropical cyclone season. This feature becomes more pronounced as we focus on floods with larger upper tail ratio. These analyses point to warm season thunderstorms and tropical cyclones as major flood agents for record floods, in contrast to the broader population of annual flood peaks, which are dominated by snowmelt and winter/spring storm systems.
- The strangest floods in the United States, that is, those producing the largest upper tail ratios (Table 1), are heavily weighted toward intense thunderstorms during the warm season. Furthermore, the most intense forms of convection, hailstorms and supercell thunderstorms, are prominent agents of strange floods. These include some of the most deadly and devastating floods in the U.S. history, notably the June 1903 Heppner flood, the May 1935 Republican River flood, the April 1934 Washita flood, the June 1965 Front Range floods, and the June 1972 Rapid City flood. Severe thunderstorms in and adjacent to mountainous terrain are prominent in the largest upper tail ratio floods (Table 1). Strange floods from intense convection are not restricted to the smallest basin scales; the May 1935 Republican River flooding extended to basin scales exceeding 30,000 km^2. The largest flood peak in the USGS stream gaging record resulted from the 1927 flooding in the Lower Mississippi River at a basin scale of 3,000,000 km^2 and was the product of multiple episodes of mesoscale convective systems, which produced severe thunderstorms and tornadoes (Smith & Baeck, 2015).
- Strange floods can result from flood agents that are uncommon for the watershed, or they can result from extremes of typical flood agents. The eastern Oregon floods, which are produced by severe thunderstorms in a region dominated by snowmelt, are quintessential examples of the former. The Front Range and High Plains floods in Table 1 (May 1935 Republican, April 1934 Washita, June 1965 Front Range, June 1972 Rapid City, and 28 May 1953 Prairie Dog Creek), all the product of mesoscale convective systems during late spring and early summer, are examples of the latter. We introduce the record flood peak seasonality as a tool for examining anomalies in the seasonal occurrence of record flood peaks.
- We show that under the assumption that the annual flood peak distribution for a site is GEV with location μ, scale σ, and shape ξ, the upper tail ratio has a GEV distribution that depends only on the shape parameter ξ and sample size. Estimates of the location and scale parameters for stationary USGS flood records are linearly related and they are linearly related to drainage basin area. Estimated shape parameters show no dependence on basin area or any other basin property. The distribution of upper tail ratio does not depend on drainage area, provided that sample size is accounted for, supporting a scaling theory for the upper tail ratio that is dependent only on the shape parameter of the parent GEV distribution and sample size. Upper tail ratio analyses for stationary USGS stations are consistent with an unbounded, thick-tailed GEV world. Surprises, in the form of unprecedented flood peaks, are a prominent feature of a thick-tailed GEV world. Error bounds in estimates of extreme flood quantiles will increase with increasing values of the GEV shape parameter.
- We examine the geography of record floods by conditioning the upper tail ratio on basin properties. Arid/semiarid watersheds have elevated values of upper tail ratio. The distribution of upper tail ratio values decreases with impervious fraction. Urban watersheds have a high frequency of large unit discharge flood peaks, but the maximum flow can be limited by capacity constraints in urban drainage networks. Jarrett's hypothesis that large floods do not occur above 2,300 m is an important element of the hydroclimatology of upper tail ratio values. There is no evidence that forest cover affects the distribution of upper tail ratios in watersheds of the United States.
- Nonstationarities are a prominent feature of some stream gaging stations with large upper tail ratio values. The Republican River flood of May 1935 provides a prototypical example. Devastating flooding lead to river regulation with flood control dams and levees and a coupled program of stream gaging. This pattern and

closely related variants of this pattern characterize the flood series for a number of the USGS stream gaging stations in Table 1.

- Large upper tail ratio flood events in the eastern United States are dominated by tropical cyclones. More than two thirds of the record flood peaks with upper tail ratios greater than 5 east of the Mississippi River were produced by tropical cyclones. Hurricanes Diane (1955), the New England Hurricane of 1938, Hurricane Floyd (1999), Hurricane Alberto (1994), and Hurricane Agnes (1972) account for 40% of the upper tail ratios larger than 5 in the eastern United States.
- Record floods in June 1903, July 1956, and July 1965 in the Blue Mountains of Oregon reflect two of the most important aspects of strange floods in the United States: they are located in and adjacent to mountainous terrain and intense thunderstorms are the principal agents of extreme rainfall. The storm environments for the three floods include similar patterns of precipitable water, CAPE, and steering winds around the Blue Mountains, with near-record values of precipitable water for each event. Transport of water vapor to the Blue Hills region occurs over 3–5 days and can involve diverse and complex paths to the region of heavy rainfall. Each of the flood events was embedded in a multiday period of thunderstorms in eastern Oregon and Washington. The 15–18 July 2012 thunderstorm outbreak provides a modern analog for the Blue Mountains flood events.
- Hydrologic modeling analyses show that motion down the drainage network of Balm Fork, which was documented by Whistler (1903) based on observer accounts, results in much larger peak discharge at Heppner than for a similar storm that moves up the drainage network. Modeling analyses for Meyers Canyon and Lane Canyon result in extreme rainfall rates, highlighting the uncertainty of the peak discharge estimates for these sites. The magnitudes of peak rainfall rates are unknown for the three storms and we have a poor understanding of the physical processes that in rare circumstances produce exceptionally large rainfall rates.

Eastern Oregon is an interesting setting for examining the regional variation of extreme floods, because it includes the location of the largest upper tail ratio flood in the United States (excluding the flood peak resulting from the Teton Dam failure) and the locations of two miscellaneous USGS flood peak measurements that control the U.S. and world flood peak envelope curves at basin scales ranging from 10 to 40 km^2. One reading of the hydrometeorological and hydrologic analyses of the Heppner, Meyers Canyon, Lane Canyon, and July 2012 events is that spatial heterogeneity of upper tail flood properties is extremely large and that each watershed will have unique flood properties. This point of view is largely driven by hydrometeorological properties of storms that control extreme floods, including the environmental conditions that create the distinctive patterns of precipitable water, CAPE, and steering winds for the episodic thunderstorm outbreaks in the region. Extreme floods in the Blue Mountains reflect an interplay between storm motion and drainage network structure, providing strong linkage between hydrometeorological and hydrologic controls of upper tail flood properties.

Regionalization is one of the core elements of many techniques used for rainfall and flood frequency analysis, including procedures used for assessing Probable Maximum Precipitation and Probable Maximum Flood. A conclusion that extreme floods in watersheds of the Blue Mountains exhibit large spatial heterogeneity does not alter the need for, and utility of, regionalization procedures. But, the Blue Mountain analyses do suggest that major advances in regional assessments of upper tail flood properties require advances in understanding and modeling hydrometeorological processes that control extreme rainfall from thunderstorms. The early ideas underlying hydrometeorological procedures for Probable Maximum Precipitation analysis were driven by the goal of developing modeling capabilities that would be coupled with appropriate observations to provide local assessments of maximum precipitation (Myers, 1967). A solid foundation for regionalization will require advances in hydrometeorology along the lines envisioned many decades ago.

The themes that emerge from examination of regional variation of extreme floods in the Blue Mountains region of Oregon are likely prominent features of many (if not most) settings in the United States. Landfalling tropical cyclones in Texas produce extreme floods which are strongly dependent on amplification of rainfall at the land-water boundary (as was the case for Hurricane Harvey in August of 2017) and at inland locations where complex terrain has contributed to record rainfall accumulations for the United States. Similar features characterize tropical cyclone flood hazards in coastal and inland regions along the remainder of the Gulf Coast and the east coast of the United States. The mountainous regions with large upper tail ratio flood peaks (section 3.2) are also notable settings in which spatial variation of upper tail flood properties has received considerable attention; these include the Front Range of the Rocky Mountains, the Black Hills, the North American Monsoon region of the southwestern United States, central Appalachians, and the San

Gabriel Mountains of California. Geomorphic and paleohydrologic studies have provided insights to hotspots of extreme flood occurrences and are key tools for advancing our understanding of the geography of upper tail flood properties.

One of the biggest challenges in developing a deeper understanding of strange floods centers on the physical processes controlling rainfall rates in thunderstorms. We do not know how large the rainfall rates were from the Heppner, Meyers Canyon, and Lane Canyon storms. The rain rates inferred from flood peak values and hydrologic modeling for the Meyers Canyon and Lane Canyon storms, which are roughly in the range from 250 to 400 mm/hr, certainly seem too large, but it is difficult to constrain rainfall rates from thunderstorms based on rain gage or even conventional radar reflectivity measurements for any setting, especially sparsely monitored locations like those in eastern Oregon. There have been larger rainfall rates documented for thunderstorms in the United States. The 18–19 July 1942 Smethport, Pennsylvania, storm not only produced the world record rainfall rate measurement of 164 mm/hr for 4.75 hr (780 mm accumulation), but also rain rates at 5- to 10-min time scale of 1,000 mm/hr (Eisenlohr, 1952). How does that happen? Polarimetric radar measurements (Figure 15) hold significant potential for examining physical processes that control rainfall rate in thunderstorms (Kumjian, 2013), because they can circumvent problems associated with hail and provide observational constraints on the interplay between storm dynamics and microphysics. Considerable progress has been made in understanding the dynamics of hailstorms, especially tornadic supercell storms, but relatively little attention has been focused on understanding the physical processes that control extreme rainfall rates, despite the fact that flash flooding is a comparable hazard to tornadoes in terms of loss of life. The quasi-idealized simulations of rainfall from supercell thunderstorms in Nielsen and Schumacher (2018) provide an important avenue for advancing the predictive understanding of extreme rainfall rates.

Our analyses of USGS flood records suggest that the nature of flood peak distributions in the conterminous United States is unbounded and thick-tailed; we certainly find little support for Horton's argument that rainfall and flood peaks are bounded. There is a substantial population of gaged watersheds for which the upper tail ratio is greater than 10 (Figure 2) and the Heppner flood provides evidence of upper tail ratios greater than 100. Comparisons between distributional properties of upper tail ratio values in the United States with other settings around the world will provide additional insights to the nature of the upper tail of flood peaks.

Appendix A

The GEV family of distributions satisfies the max-stable property (Leadbetter et al., 1982) implying that the maximum flood $X_{(n)}$ has a GEV distribution that is related to the distribution of the annual peaks X_i, through a transformation of the scale a_n and location b_n:

$$X_{(n)} \sim a_n X_i + b_n. \tag{A1}$$

For $\xi \neq 0$, $a_n = n^{\xi}$ and $b_n = (\mu - \frac{\sigma}{\xi})(1 - n^{\xi})$ and for the Gumbel case the sequences simplify to $a_n = 1$ and $b_n = \mu + \sigma\, ln(n)$.

For $X_i \sim GEV(\mu, \sigma, \xi)$,

$$a X_i + b \sim GEV(a\,\mu + b,\ a\,\sigma,\ \xi). \tag{A2}$$

The distribution of the record flood is thus

$$X_{(n)} \sim GEV(\mu - \frac{\sigma}{\xi}(1 - n^{\xi}),\ n^{\xi}\,\sigma,\ \xi). \tag{A3}$$

We assume that the sample quantile function provides a good approximation to the quantile function and evaluate the ratio of the record flood $X_{(n)}$ and the 10-year flood Q(.9). The 10-year flood under the GEV assumption is

$$Q(.9) = \mu - \frac{\sigma}{\xi}(1 - [-ln(0.9)]^{-\xi}). \tag{A4}$$

As noted in section 2, estimated location and scale parameters for annual peak flood records are linearly related (and they have an intercept that is approximately 0), so we can assume

$$\mu = c\,\sigma. \tag{A5}$$

Combining these results, we conclude that in a GEV world the upper tail ratio depends only on the shape parameter and sample size. In particular

$$Z \sim GEV\left(\frac{c - g(\xi)}{c - h(\xi)}, \frac{n^{\xi}}{c - h(\xi)}, \xi\right), \tag{A6}$$

where

$$h(\xi) = \frac{(1 - [-ln(0.9)]^{-\xi})}{\xi} \tag{A7}$$

and

$$g(\xi) = \frac{(1 - n^{-\xi})}{\xi}. \tag{A8}$$

Acknowledgments
The authors gratefully acknowledge constructive comments from Robert Mason, Jonathan Gourley, and two anonymous reviewers; constructive comments were also provided by Jim O'Connor, William Asquith, Eric Gaume, Stephen Burges, Chris Milly, and Gabriele Villarini on earlier versions of the paper. We also thank Dan Turner-Gillespie for the encouragement to examine the Heppner Flood and acknowledge the assistance of personnel from multiple USGS offices in sorting out details of a number of flood peak measurements. This research was supported by the National Science Foundation (NSF; Grants EAR-1632048, AGS-1522492, and CBET-1444758), NASA (Grant NNX10AI46G), NOAA (Grant NA14OAR4830101), and the Moore Charitable Foundation. Annual flood peak data used in this study can be obtained from the USGS National Water Information System. NLDN data were provided by the NASA Lightning Imaging Sensor (LIS) instrument team and the LIS data center via the Global Hydrology Resource Center (GHRC) located at the Global Hydrology and Climate Center (GHCC), Huntsville, Alabama, through a license agreement with Global Atmospherics (now Vaisala), Inc. (GAI). Paul Bates was supported by a Leverhulme Research Fellowship and a Royal Society Wolfson Research Merit award.

References

Atallah, E., & Bosart, L. F. (2003). The extratropical transition and precipitation distribution of Hurricane Floyd (1999). *Monthly Weather Review*, *131*, 1063–1081.
Ayalew, T. B., Krajewski, W. F., & Mantilla, R. (2014). Connecting the power-law scaling structure of peak-discharges to spatially variable rainfall and catchment physical properties. *Advances in Water Resources*, *71*, 32–43.
Baker, V. R. (1975). Flood hazards along the Balcones Escarpment in central Texas: Alternative approaches to their recognition, mapping and management (Tech. Rep.) *Bureau of Economic Geology University of Texas Circ. 75-5*. Texas: University of Texas.
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*, 257–269. https://doi.org/10.1002/2016WR019064
Billington, D. P., & Jackson, D. C. (2006). *Big dams of the New Deal era: A confluence of engineering and politics* (369 pp.). Cambridge: University of Oklahoma Press.
Blöschl, G., & Sivapalan, M. (1997). Process controls on regional flood frequency: Coefficient of variation and basin scale. *Water Resources Research*, *33*(12), 2967–2980.
Boner, F. C., & Stermitz, F. (1967). Floods of June 1964 in northwestern Montana (*Water Supply Paper 1840-B, U.S*). Washington D. C: Geological Survey.
Bosart, L. F., Seimon, A., LaPenta, K. D., & Dickinson, M. J. (2006). Supercell tornadogenesis over complex terrain: The Great Barrington, Massachusetts tornado on 29 May 1995. *Weather and Forecasting*, *21*, 897–921.
Bunkers, M. J., & Doswell, C. A. (2016). Comment on "Double impact: When both tornadoes and flash floods threaten the same place and time". *Weather and Forecasting*, *31*, 1715–1721.
Byrd, J. G. (2009). *Calamity: The Heppner Flood of 1903* (p. 202). Seattle, Washington: The University of Washington Press.
Caracena, F., & Fritsch, J. M. (1978). Focusing mechanisms in the Texas Hill Country flash FIOMS of 1978. *Monthly Weather Review*, *111*, 2319–2332.
Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer.
Colle, B. A. (2003). Numerical simulations of the extratropical transition of Floyd (1999): Structural evolution and responsible mechanisms for the heavy rainfall over the Northeast United States. *Monthly Weather Review*, *131*, 2905–2926.
Colle, B. A., & Mass, C. F. (2000). The 5–9 February 1996 flooding event over the Pacific Northwest: Sensitivity studies and evaluation of the MM5 precipitation forecasts. *Monthly Weather Review*, *128*(3), 593–618.
Cooper, R. M. (2006). Estimation of peak discharges for rural, unregulated streams in Eastern Oregon (Open File Report SW 06-002). Salem, Oregon: State of Oregon.
Costa, J. E. (1987a). Hydraulics and basin morphometry of the largest flash floods in the conterminous United States. *Journal of Hydrology*, *93*(3–4), 313–338.
Costa, J. E. (1987b). A comparison of the largest rainfall-runoff floods in the United States with those of the People's Republic of China and the world. *Journal of Hydrology*, *96*, 101–115.
Costa, J. E., & Jarrett, R. D. (2008). An evaluation of selected extraordinary floods in the United States reported by the US Geological Survey and implications for future advancement of flood science (vol. Scientific Investigations Report 2008–5164). Reston, VA: U.S. Geological Survey.
Cotton, W. R., Bryan, G. H., & van den Heever, S. C. (2010). *Storm and cloud dynamics*, 2nd edition. (p. 775). Oxford: Academic Press.
Delrieu, G., Ducrocq, V., Gaume, E., Nicol, J., Payrastre, O., Yates, E., et al., (2005). The catastrophic flash-flood event of 8–9 September 2002 in the gard region, France: A first case study for the Cèvennes Vivarais Mediterranean Hydrometeorological Observatory. *Journal of Hydrometeorology*, *6*, 34–52.
Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., & Cayan, D. R. (2011). Atmospheric rivers, floods and the water resources of California. *Water*, *3*, 445–478.
DiBaldassarre, G., Viglione, A., Carr, G., Salinas, J. L., & Bloeshl, G. (2013). Socio-hydrology: Conceptualizing human-flood interactions. *Hydrology and Earth System Sciences*, *17*, 3295–3303.
DiMego, G. J., & Bosart, L. F. (1982). The transformation of Tropical Storm Agnes into an extratropical cyclone. Part I: The observed fields and vertical motion computations. *Monthly Weather Review*, *110*(5), 385–411.
Dixon, M., & Wiener, G. (1993). TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, *10*(6), 785–797.

Doswell, C. A., Brooks, H. E., & Maddox, R. A. (1996). Flash flood forecasting: An ingredients-based methodology. *Weather and Forecasting*, *11*(4), 560–581.

Douglas, J. S. (1908). A California cloudburst. *Monthly Weather Review*, *36*(9), 631–632.

Douglas, E. M., & Vogel, R. M. (2006). Probabilistic behavior of floods of record in the United States. *Journal of Hydrologic Engineering*, *11*(5), 482–485.

DuMouchel, W. H (1983). Estimating the stable index α in order to measure tail thickness: A critique. *The Annals of Statistics*, *11*(4), 1019–1031.

Ducrocq, V., ad D. Ricard, O. N., Lebeaupin, C., & Thouvenin, T. (2008). A numerical study of three catastrophic precipitation events over southern France. II: Mesoscale triggering and stationarity factors. *Quarterly Journal of the Royal Meteorological Society*, *134*, 131–145.

Eaton, L. S., Morgan, B. A., Kochel, R. C., & Howard, A. D. (2003). Quaternary deposits and landscape evolution in the central Blue Ridge of Virginia. *Geomorphology*, *56*, 139–154.

Eisenlohr, W. S. (1952). Floods of July 18, 1942 in north-central Pennsylvania (U. S. Geological Water Supply Paper 1134-B). United States: USGS.

El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., & Bobeé, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, *43*, W03410. https://doi.org/10.1029/2005WR004545

Emanuel, K. (2017). Assessing the present and future probability of Hurricane Harvey's rainfall. *Proceedings of the National Academy of Sciences*, *114*(48), 12,681–12,684.

Engelbrecht, H. H., & Brancato, G. N. (1959). World record one-minute rainfall at Unionville, Maryland. *Monthly Weather Review*, *87*(8), 303–306.

England, J. F., Julien, P. Y., & Vellleux, M. L. (2014). Physically-based extreme flood frequency with stochastic storm transposition and paleoflood data on large watersheds. *Journal of Hydrology*, *510*, 228–245.

Enzel, Y., Ely, L. L., House, K., & Baker, V. R. (1993). Paleoflood evidence for a natural upper bound to flood magnitudes in the Colorado River Basin. *Water Resources Research*, *29*(7), 2287–2297.

Falcone, J. A., Carlisle, D. M., Wolock, D. M., & Meador, M. R. (2010). GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology*, *91*(2), 621.

Follansbee, R., & Jones, E. E. (1922). Arkansas River flood of June 3–5, 1921 (U.S. Geological Water Supply Paper 487). United States: USGS.

Follansbee, R., & Spiegel, J. B. (1937). Floods on the Republican and Kansas Rivers May and June 1935 (U.S. Geological Water Supply Paper 796-B). United State: USGS.

Fundel, F., & Zappa, M. (2011). Hydrological ensemble forecasting in mesoscale catchments: Sensitivity to initial conditions and value of reforecasts. *Water Resources Research*, *47*, W09520. https://doi.org/10.1029/2010WR009996

Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., et al. (2009). A compilation of data on European flash floods. *Journal of Hydrology*, *367*, 70–78.

Giordano, L. A., & Fritsch, J. M. (1991). Strong tornadoes and flash-flood-producing rainstorms during the warm season in the mid-Atlantic region. *Weather and Forecasting*, *6*, 437–455.

Goodrich, D. C., Guertin, D. P., Burns, I. S., Nearing, M. A., Stone, J. J., Wei, H., & Heilman, P. (2011). AGWA: The automated geospatial watershed assessment tool to inform rangeland management. *Rangelands*, *33*(4), 41–47.

Hack, J. T., & Goodlett, J. C. (1960). Geomorphology and forest ecology of a mountain region in the central Appalachians (U.S. Geological Prof. Paper 347). United State: USGS.

Halbert, K., Nguyen, C. C., Payrastre, O., & Gaume, E. (2016). Reducing uncertainty in flood frequency analyses: A comparison of local and regional approaches involving information on extreme historical floods. *Journal of Hydrology*, *541*, 90–98. https://doi.org/10.1016/j.jhydrol.2016.01.017

Hamada, A., Takayabu, Y. N., Liu, C., & Zipser, E. J. (2015). Weak linkage between the heaviest rainfall and tallest storms. *Nature Communications*, *6*(6213). https://doi.org/10.1038/ncomms7213

Harden, T. M., O'Connor, J. E., Driscoll, D. G., & Stamm, J. F. (2011). Flood-frequency analyses from paleoflood investigations for Spring, Rapid, Boxelder, and Elk Creeks, Black Hills, western South Dakota (Scientific Investigations Report 2011 - 5131). United State: U.S. Geological Survey.

Hart, R. E., & Evans, J. L. (2001). A climatology of the extratropical transition of Atlantic tropical cyclones. *Journal of Climate*, *14*, 546–564.

Helmus, J. J., & Collis, S. M. (2016). The Python ARM Radar Toolkit (Py-ART), a library for working with weather radar data in the python programming language. *Journal of Open Research Software*, *4*(1), e25.

Helsel, D. R., & Hirsch, R. M. (1993). *Statistical methods in water resources*, (p. 522). Amsterdam: Elsevier.

Hendricks, E. L. (1964). Summary of floods in the United States during 1956 (USGS Water-Supply Paper 1530). Washington, DC: U.S. Geological Survey.

Herschy, R. (2003). *World catalogue of maximum observed floods*, IAHS-AISH Publication 284 320. Wallingfor: International Association of Hydrological Sciences.

Hicks, N. S., Smith, J. A., Miller, A. J., & Nelson, P. A. (2005). Catastrophic flooding from an orographic thunderstorm in the central Appalachians. *Water Resources Research*, *41*(W12428). https://doi.org/10.1029/2005WR004129

Hirsch, R. M., & Archfield, S. A. (2015). Flood trends: Not higher but more often. *Nature Climate Change*, *5*, 198–199.

Hirschboeck, K. K. (1987). Hydroclimatically-defined mixed distributions in partial duration flood series. In V. P. Singh (Ed.), *Hydrologic frequency modelling*. Boston: D. Reidel Publishing Company (pp. 199–212).

Hirschboeck, K. K. (1988). Flood hydroclimatology. In V. R. Baker, R. C. Kockel, & P. C. Patton (Eds.), *Flood geomorphology* (pp. 27–49). Hoboken, NJ: John Wiley & Sons.

Hitchens, N. M., & Brooks, H. E. (2013). Preliminary investigation of the contribution of supercell thunderstorms to the climatology of heavy and extreme precipitation in the United States. *Atmospheric Research*, *123*, 206–210.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*. Cambridge; New York: Cambridge University Press.

Hosking, J., Wallis, J. R., & Wood, E. F. (1985). An appraisal of the regional flood frequency procedure in the UK Flood Studies Report. *Hydrological Sciences Journal*, *30*(1), 85–109.

Jarrett, R. D., & Costa, J. E. (1988). Evaluation of the flood hydrology in the Colorado Front Range using precipitation, streamflow and paleoflood data for the Big Thompson River basin (USGS Water Resources Investigation Report 87-4177). Washington, DC: U.S Geological Survey.

Jarrett, R. D., & Tomlinson, E. M. (2000). Regional interdisciplinary paleoflood approach to assess extreme flood potential. *Water Resources Research*, *36*(10), 2957–2984.

Javier, J. R. N., Smith, J. A., England, J., Baeck, M. L., Steiner, M., & Ntelekos, A. A. (2007). Climatology of extreme rainfall and flooding from orographic thunderstorm systems in the upper Arkansas River basin. *Water Resources Research*, *43*, W10410. https://doi.org/10.1029/2006WR005093

Jones, J. A., & Grant, G. E. (1996). Peakflow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon. *Water Resources Research*, *32*, 959–974.

Jordan, P. R., & Combs, L. J. (1996). Summary of floods in the United States during 1990 and 1991 (Water Supply Paper 2474). United State: U.S. Geological Survey.

Kampf, S. K., & Burges, S. J. (2007). A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water Resources Research*, *43*, W05423. https://doi.org/10.1029/2006WR005

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*, 1287–1304.

Koenig, T. A., Bruce, J. L., O'Connor, J. E., McGee, B. D., Holmes, R. R., Hollins, R., et al. (2016). Identifying and preserving high-water mark data (Surface-Water Techniques Chapter 24). United State: U.S. Geological Survey.

Konrad, C. E. (2001). The most extreme precipitation events over the eastern United States from 1950 to 1996: Considerations of scale. *Journal of Hydrometeorology*, *2*, 309–325.

Koutsoyiannis, D., & Montanari, A. (1999). A probabilistic view of Hershfield's method for estimating probable maximum precipitation. *Water Resources Research*, *35*(4), W05429. https://doi.org/10.1029/2006WR005592 1313–1322.

Kumjian, M. R. (2013). Principles and applications of dual-polarization radar. *Journal of Operational Meteorology*, *1*(19-21), 226–274.

Leadbetter, M. R., Lindgren, G., & Rootzen, H. (1982). *Extremes and related properties of random sequences and processes*. New York: Springer Verlag.

Leopold, L. B. (1942). Areal extent of intense rainfalls, New Mexico and Arizona. *Transactions of the American Geophysical Union*, *23*, 558–563.

Leopold, L. B. (1946). Two intense local floods in New Mexico. *Transactions of the American Geophysical Union*, *27*(4), 535–539.

Levish, D. R., & Ostenaa, D. A. (1996). Applied paleoflood hydrology in north-central Oregon (Guidebook for Fieldtrip 2, April 19-21, 1996 Cordilleran Section 92nd Annual Meeting). Denver, CO: GSA, Bureau of Reclamation.

Li, C., Wang, G. A., & Li, R. R. (2013). Maximum observed floods in China. *Hydrological Sciences Journal*, *58*, 728–735.

Liu, M., & Smith, J. A. (2016). Extreme rainfall from landfalling tropical cyclones in the eastern United States: Hurricane Irene (2011). *Journal of Hydrometeorology*, *17*, 2883–2904.

Llasat, M. C., Marcos, R., Turco, M., Gilabert, J., & Llasat-Botija, M. (2016). Trends in flash flood events versus convective precipitation in the Mediterranean region: The case of Catalonia. *Journal of Hydrology*, *541*(A), 24–37.

Maddox, R. A. (1980). Mesoscale convective complexes. *Bulletin of the American Meteorological Society*, *61*(11), 1374–1387.

Maddox, R. A., Hoxit, L. R., Chappell, C. F., & Caracena, F. (1978). Comparison of meteorological aspects of the Big Thompson and Rapid City flash floods. *Monthly Weather Review*, *106*, 375–389.

Mahoney, K., Alexander, M. A., Thompson, G., Barsugli, J. J., & Scott, J. D. (2012). Changes in hail and flood risk in high-resolution simulations over Colorado's mountains. *Nature Climate Change*, *2*, 125–131.

Mallakpour, I., & Villarini, G. (2015). The changing nature of flooding across the central United States. *Nature Climate Change*, *5*, 250–254.

Markowski, P. M., & Dotzek, N. (2011). A numerical study of the effects of orography on supercells. *Atmospheric Research*, *100*(4), 457–478.

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*(3), 737–744.

McAnelly, R. L., & Cotton, W. R. (1989). The precipitation life cycle of mesoscale convective complexes over the central United States. *Monthly Weather Review*, *117*, 784–808.

Michaelis, A. C., & Lackmann, G. M. (2013). Numerical modeling of a historic storm: Simulating the Blizzard of 1888. *Geophysical Research Letters*, *40*, 4092–4097. https://doi.org/10.1002/grl.50750

Michaud, J. D., Hirschboeck, K. K., & Winchel, M. (2001). Regional variations in small-basin floods in the United States. *Water Resources Research*, *37*, 1405–1416.

Miller, A. J. (1990). Flood hydrology and geomorphic effectiveness in the central Appalachians. *Earth Surface Processes and Landforms*, *15*, 119–134.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science*, *319*, 573–574.

Morin, E., Goodrich, D. C., Maddox, R. A., Gao, X. G., Gupta, H. V., & Sorooshian, S. (2006). Spatial patterns in thunderstorm rainfall events and their coupling with watershed hydrological response. *Advances in Water Resources*, *29*(6), 843–860. https://doi.org/10.1016/j.advwatres.2005.07.014

Morrison, J. E., & Smith, J. A. (2002). Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water Resources Research*, *38*(12), 1305. https://doi.org/10.1029/2001WR000502

Murphy, E. C. (1904). Destructive floods in the United States in 1903 (Water Supply and Irrigation Paper 96). Washington, DC: U. S Geological Survey.

Myers, V. A. (1967). Meteorological estimation of extreme precipitation for spillway design floods (Tech. Rep. WBTM Hydro-5). Washington, DC: U. S. Weather Bureau Office of Hydrology.

Neiman, P. J., Schick, L. J., Ralph, F. M., Hughes, M., & Wick, G. A. (2011). Flooding in western Washington: The connection to atmospheric rivers. *Journal of Hydrometeorology*, *12*(6), 1337–1358.

Nguyen, C. C., Gaume, E., & Payrastre, O. (2013). Regional flood frequency analyses involving extraordinary flood events at ungauged sites: Further developments and validations. *Journal of Hydrology*, *508*, 385–396. https://doi.org/10.1016/j.jhydrol.2013.09.058

Nielsen, E. R., Herman, G. R., Tournay, R. C., Peters, J. M., & Schumacher, R. S. (2015). Double impact: When both tornadoes and flash floods threaten the same place and time. *Weather and Forecasting*, *30*, 1673–1693.

Nielsen, E. R., & Schumacher, R. S. (2018). Dynamical insights into extreme short-term precipitation associated with supercells and mesovortices. *Journal of the Atmospheric Sciences*, *75*, 2983–3009. https://doi.org/10.1175/JAS-D-17-0385.1

Nielsen-Gammon, J. W., Zhang, F., Odins, A. M., & Myoung, B. (2005). Extreme rainfall in Texas: Patterns and predictability. *Physical Geography*, *26*(5), 340–364.

O'Connor, J. E., & Baker, V. R. (1992). Magnitudes and implications of peak discharges from glacial Lake Missoula. *Geological Society of America Bulletin*, *104*(3), 267–279.

O'Connor, J. E., & Costa, J. E. (2004a). Spatial distribution of the largest rainfall-runoff floods from basins between 2.6 and 26,000 km^2 in the United States and Puerto Rico. *Water Resources Research*, *40*, W01107. https://doi.org/10.1029/2003WR002247

O'Connor, J. E., & Costa, J. E. (2004b). The world's largest floods, past and present: Their causes and magnitudes (*Circular 1254*). United State: U.S. Geological Survey.

O'Connor, J. E., Grant, G. E., & Costa, J. E. (2002). The geology and geography of floods. *Ancient Floods, Modern Hazards*, *5*, 359–385.

Ogden, F. L., Sharif, H. O., Senarath, S. U. S., Smith, J. A., Baeck, M. L., & Richardson, J. R. (2000). Hydrologic analysis of the Fort Collins, Colorado, flash flood of 1997. *Journal of Hydrology*, *228*, 82–100.

Parajka, J., Kohnová, S, Bálint, G., Barbuc, M., Borga, M., Claps, P., et al. (2010). Seasonal characteristics of flood regimes across the Alpine-Carpathian range. *Journal of Hydrology*, *394*, 78–89.

Patton, P. C., & Baker, V. R. (1976). Morphometry and floods in small drainage basins subject to diverse hydrogeomorphic controls. *Water Resources Research*, *12*(5), 941–952.

Petersen, W. A., Carey, L. D., Rutledge, S. A., Knievel, J. C., Doesken, N. J., Johnson, R. H., et al. (1999). Mesoscale and radar observations of the Fort Collins flash flood of 28 July 1997. *Bulletin of the American Meteorological Society*, *80*(2), 191–216.

Potter, K. W., & Walker, J. F. (1981). A model of discontinuous measurement error and its effects on the probability distribution of flood discharge measurements. *Water Resources Research*, *17*(5), 1505–1509.

Robinson, J. S., & Sivapalan, M. (1997). An investigation into the physical causes of scaling and heterogeneity of regional flood frequency. *Water Resources Research*, *33*(5), 1045–1059.

Rogash, J. A., & Racy, F. (2002). Some meteorological characteristics of significant tornado events occurring in proximity to flash flooding. *Weather and Forecasting*, *17*, 155–159.

Romine, R. S., Burgess, D. W., & Wilhelmson, R. B. (2008). A dual-polarization-radar-based assessment of the 8 May 2003 Oklahoma City area tornadic supercell. *Monthly Weather Review*, *136*, 2849–2870.

Rossi, F., Fiorentino, M., & Versace, P. (1984). Two-component extreme value distribution for flood frequency analysis. *Water Resources Research*, *20*(7), 847–856.

Rostvedt, J. O. (1970). Summary of floods in the United States during 1965 (USGS Water-Supply Paper 1850-E). Washington, DC: U.S Geological Survey.

Ryberg, K. R., Goree, B. B., Williams-Sether, T., & Mason, Jr. R. R. (2017). The U.S. Geological Survey peak-flow file data verification project (Scientific Investigations Report 2017-5119). United State: U. S. Geological Survey.

Salinas, J. L., Castellarin, A., Viglione, A., Kohnova, S., & Kjeldsen, T. R. (2014). Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent? *Hydrology and Earth System Sciences*, *18*, 4381–4389.

Scheffknecht, P., Serafin, S., & Grubisic, V. (2017). A long-lived supercell over mountainous terrain. *Quarterly Journal of the Royal Meteorological Society*, *143*, 2973–2986.

Schumacher, R. S., & Johnson, R. H. (2005). Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Monthly Weather Review*, *133*(4), 961–976. https://doi.org/10.1175/MWR2899.1.

Smith, J. A., & Baeck, M. L. (2015). "Prophetic vision, vivid imagination": The 1927 Mississippi River flood. *Water Resources Research*, *51*, 9964–9994. https://doi.org/10.1002/2015WR017927

Smith, J. A., Baeck, M. L., Morrison, J. E., & Sturdevant-Rees, P. (2000). Catastrophic rainfall and flooding in Texas. *Journal of Hydrometeorology*, *1*, 5–25.

Smith, J. A., Baeck, M. L., Ntelekos, A. A., Villarini, G., & Steiner, M. (2011b). Extreme rainfall and flooding from orographic thunderstorms in the central Appalachians. *Water Resources Research*, *47*, W04514. https://doi.org/10.1029/2010WR010190

Smith, J. A., Baeck, M. L., Steiner, M., & Miller, A. J. (1996). Catastrophic rainfall from an upslope thunderstorm in the central Appalachians: The Rapidan storm of June 27, 1995. *Water Resources Research*, *32*(10), 3099–3113.

Smith, J. A., Baeck, M. L., Zhang, Y., & Doswell, C. A. III. (2001). Extreme rainfall and flooding from supercell thunderstorms. *Journal of Hydrometeorology*, *2*, 469–489.

Smith, A., Sampson, C., & Bates, P. (2015). Regional flood frequency analysis at the global scale. *Water Resources Research*, *51*, 539–553. https://doi.org/10.1002/2014WR015814

Smith, B. K., & Smith, J. A. (2015). The flashiest watersheds in the contiguous United States. *Journal of Hydrometeorology*, *16*(12), 2365–2381.

Smith, J. A., Villarini, G., & Baeck, M. L. (2011a). Mixture distributions and the climatology of extreme rainfall and flooding in the eastern United States. *Journal of Hydrometeorology*, *12*(2), 294–309.

Soderholm, B., Ronalds, B., & Kirshbaum, D. N. (2014). The evolution of convective storms initiated by an isolated mountain ridge. *Monthly Weather Review*, *142*, 1430–1451.

Sprenger, M., & Wernli, H. (2015). The LAGRANTO Lagrangian analysis tool: Version 2.0. *Geoscientific Model Development*, *8*, 2569–2586.

Tuttle, J. D., & Davis, C. A. (2006). Corridors of warm season precipitation in the central United States. *Monthly Weather Review*, *134*(9), 2297–2317. https://doi.org/10.1175/MWR3188.1

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer.

Villarini, G. (2016). On the seasonality of flooding across the continental United States. *Advances in Water Resources*, *87*, 80–91.

Villarini, G., Serinaldi, F., Smith, J. A., & Krajewski, W. F. (2009). On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, *45*, W08417. https://doi.org/10.1029/2008WR007645

Villarini, G., & Smith, J. A. (2010). Flood peak distributions for the eastern United States. *Water Resources Research*, *46*, W06504. https://doi.org/10.1029/2009WR008395

Villarini, G., Smith, J. A., Baeck, M. L., Marchock, T., & Vecchi, G. A. (2011). Characterization of rainfall distribution and flooding associated with U.S. landfalling tropical cyclones: Analyses of Hurricanes Frances, Ivan, and Jeanne (2004). *Journal of Geophysical Research*, *116*, D23116. https://doi.org/10.1029/2011JD016175

Vogel, R. M., Zafirakou-Koulouris, A., & Matalas, N. C. (2001). Frequency of record-breaking floods in the United States. *Water Resources Research*, *37*(6), 1723–1731.

Wells, J. V. (1959). Summary of floods in the United States During 1953 (USGS Water-Supply Paper 1320-E). Washington, DC: U.S. Geological Survey.

Whistler, J. T. (1903). The Heppner disaster. *Engineering News Record*, *L*(3), 53–54.

Wood, E. F., Sivapalan, M., & Beven, K. (1990). Similarity and scale in catchment storm response. *Reviews of Geophysics*, *28*(1), 1–18. https://doi.org/10.1029/RG028i001p00001

Woolley, R. R. (1946). Cloudburst floods in Utah 1850–1938 (Water Supply Paper 994). Washington, DC: U.S. Geological Survey, US Government Printing Office.

Yang, L., Liu, M., Smith, J. A., & Tian, F. (2017). Typhoon Nina and the August 1975 flood over central China. *Journal of Hydrometeorology*, *18*, 451–472.

Zhang, Y., Smith, J. A., & Baeck, M. L. (2001). The hydrology and hydrometeorology of extreme floods in the Great Plains of eastern Nebraska. *Advances in Water Resources*, *24*, 1037–1049.

Zhu, L., & Quiring, S. M. (2013). Variations in tropical cyclone precipitation in Texas (1950 to 2009). *Journal of Geophysical Research: Atmospheres*, *118*, 3085–3096. https://doi.org/10.1029/2012JD018554

Zipser, E. J., Cecil, D. J., Liu, C., Nesbitt, S. W., & Yorty, D. P. (2006). Where are the most intense thunderstorms on Earth? *Bulletin of the American Meteorological Society*, *87*, 1057–1071.

Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophysical Research Letters*, *43*, 8082–8076. https://doi.org/10.1002/2016GL069445