An Extreme Value Model for United States Hail Size

John T. Allen*

Department of Earth and Atmospheric Sciences, Central Michigan University, Mt Pleasant, Michigan, USA.

International Research Institute for Climate and Society, Columbia University, Palisades, New York, USA.

Michael K. Tippett

Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, USA.

Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia.

Yasir Kaheil

Center for Property Risk Solutions, FM Global, Research Division, Norwood, Massachusetts, USA.

Adam H. Sobel

Department of Applied Physics and Applied Mathematics and

Department of Earth and Environmental Sciences, Columbia University, New York, New York, USA.

Chiara Lepore

Lamont Doherty Earth Observatory, Columbia University, Palisades, New York, USA.

Shangyao Nong, Andreas Muehlbauer

Center for Property Risk Solutions, FM Global, Research Division, Norwood, Massachusetts, USA.

[*] *Corresponding author address:* Brooks Hall 326, Central Michigan University, Mt Pleasant, MI, 48858, USA

E-mail: JohnTerrAllen@gmail.com

ABSTRACT

The spatial distribution of return intervals for U.S. hail size is explored within the framework of extreme value theory using observations from the period 1979-2013. The center of the continent has experienced hail in excess of 5 inches in the past 30 years, whereas hail in excess of 1 inch is more common in other regions, including the West Coast. Observed hail sizes show heavy quantization toward fixed diameter reference objects and are influenced by spatial and temporal biases similar to those noted for hail occurrence. Recorded hail diameters have been growing in recent decades due to improved reporting. These data limitations motivate exploration of extreme value distributions to represent the return periods for various hail diameters. The parameters of a Gumbel distribution are fit to dithered observed annual maxima on a national 1° x 1° grid at locations with sufficient records. Gridded and kernel smoothed return sizes and quantiles up to the 200-year return period are determined for the fitted Gumbel distribution. These are used to illustrate return levels for hail greater than a given size for at least one location within each 1° x 1° grid box for the U.S.

1. Introduction

Large hail ($\geq$ 25 mm or 1 in.) can produce significant damage to property and agriculture. However, little is known about the hazard posed by or incidence of the largest hail diameters. Large hail is the greatest contributor to insured losses from thunderstorms in both the U.S. and globally (Gunturi and Tippett 2017), producing cumulative and single event losses that often total in excess of USD $1 billion (Changnon 2008; Sander et al. 2013; Munich RE 2015). Cumulative losses typically arise as the result of a single or several days of damaging hail events of smaller magnitude (e.g., St Louis, Missouri 2012, USD $1.6 billion total) or impacts on a number of rural centers, in addition to agricultural losses. Large catastrophic single-event losses typically occur when a larger urban center is impacted with hail at or exceeding golf ball diameter (45mm or 1.75in), when damage to structures, windows and vehicles become more frequent (Brown et al. 2015). Recent examples of such catastrophic hail storms include a USD $4 billion hail event in Phoenix, Arizona (100 mm or 4 in. maximum diameter hailstones), a USD $900 million loss hail event impacting Dallas-Fort Worth in 2012 (2.75-3.5 in. stones; Brown et al. (2015)), and two hail storms in Texas during the spring of 2016 (including one in San Antonio that produced a combined USD $4.7 billion loss due to hail (Swiss RE 2017). To understand the hazard and damage potential posed by large hail events, there are several important quantities that need to be explored. The likelihood of hail occurrence at a given location provides some guidance in determining this hazard (e.g., Allen and Tippett 2015; Allen et al. 2015). However, it is not only the likelihood of occurrence but also the size, velocity of impact, and spatial extent of these stones that determines the scale and nature of damage (Changnon Jr 1966; Morgan Jr and Towery 1975; Changnon Jr 1977; Nelson and Young 1979; Cox and Armstrong 1981; Cheng et al. 1985; Sánchez et al. 1996; Heymsfield et al. 2014; Brown et al. 2015). These elements present an important part of the potential for economic losses to agriculture and property. The significance of large hail to the country motivates an analysis of just how large hailstones can get over the U.S., leveraging both climatology and extrapolation of the likelihood of large hail.

To explore the spatial risk, occurrence, and magnitude of hail, previous work in the U.S. has leveraged a mixture of insurance data, National Weather Service spotter observations, field campaigns and weather station data (e.g., Changnon Jr 1977; Cox and Armstrong 1981). Most challenging to many of these studies was the limited spatial distribution of observed hail, which generally provided insufficient resolution to determine hail swathes and corresponding loss characteristics (Morgan Jr and Towery 1975; Nelson and Young 1979). Obtaining a picture of the hazard over larger parts of the continental U.S. is challenging, as insurance data are often non-specific in their spatial extent or combined with other hazards (Changnon 1999; Brown et al. 2015), and spotter observations and field campaigns are relatively few and far between (Strong and Lozowski 1977; Ortega et al. 2009; Blair and Leighton 2012; Heymsfield et al. 2014; Blair et al. 2017). In contrast, the abundance of hail reports for the U.S. in Storm Data (Schaefer and Edwards 1999) have led to a number of climatologies exploring hail occurrence and the hazard posed (Kelly et al. 1985; Changnon 1999; Changnon and Changnon 2000; Schaefer et al. 2004; Doswell et al. 2005; Changnon 2008; Allen et al. 2015; Allen and Tippett 2015). Despite these efforts contributing greatly to our understanding of spatial hail occurrence, the temporal and spatial limitations of hail size observations have made the hazard posed to property by large hail unclear (Doswell et al. 2005; Allen and Tippett 2015). Illustrative examples of these problems include the tendency for clustering and duplication of hail reports towards more heavily populated areas, concentration of the early reports in the record in the Great Plains, and a sensitivity to quantization as hail size approaches the arbitrary criteria used to define severe thunderstorms (Schaefer et al. 2004; Doswell et al. 2005; Allen and Tippett 2015).

There are also several challenges introduced by specifically considering hail size, rather than occurrence. Arbitrary methods of hail size measurement and a text-based observation system that trusts observers to make estimations increases the difficulty of subsequent analysis of size/occurrence distributions (Allen and Tippett 2015). Hail size observations are heavily quantized by the use of reference objects when recording their occurrence, rather than direct measurements (Doswell et al. 2005; Blair and Leighton 2012; Blair et al. 2017). This methodology leads to both the over- and under-reporting of maximum size as hail is skewed toward reference objects, further emphasizing issues with hail size observations (Heymsfield et al. 2014). It is also questionable whether largest point observation of hail size correctly reflects the largest hail that occurs in a storm (Bardsley 1990; Blair and Leighton 2012; Blair et al. 2017). This is exacerbated by the rarity of large hail (Fraile et al. 1992), as well as storms being likely to produce fewer large stones, or large volumes of hail, but not both (Cheng et al. 1985). A non-observational complication is posed by the relative importance of the size of hail to different economic sectors. To agriculture, a hail stone of 12.5mm (0.5 in.) or larger could be extremely damaging (Changnon Jr 1977; McMaster 1999; Doswell 2001). In contrast, for structures or vehicles, hailstones of 45mm (1.75 in.) or greater are typically necessary to cause large amounts of damage (Cox and Armstrong 1981; Heymsfield et al. 2014; Brown et al. 2015; Allen and Tippett 2015). Thus estimating the hazard posed by larger hail can pose challenges, and can depend heavily on the targeted group exposed to the hazard.

In this paper, we focus on the likelihood of hail in excess of the U.S. severe thunderstorm criterion (25 mm or 1 in.). In particular, we explore the characteristics of the larger diameter hail that produces the greatest degree of damage to infrastructure and property, and look to statistically estimate the probability of occurrence. We do this by applying extreme value theory methods to hail size observations. Extreme Value Theory (Fisher and Tippett 1928; Gumbel 1958; Frechet 1927) has changed the way engineers and scientists quantify the hazard associated with rare but extreme events. In particular, the Generalized Extreme Value (GEV) distribution (Jenkinson 1955) has seen widespread application in fields such as hydrology (e.g., rainfall extremes, streamflow), extreme wind speeds, finance and, more generally, in earth and atmospheric science (for more complete reviews, see Palutikof et al. 1999; Coles 2001). However, this approach has only been applied rarely to hail due to limitations in data availability and insufficient record length to provide an estimate of return size (Cox and Armstrong 1981; Smith and Waldvogel 1989; Bardsley 1990; Fraile et al. 2003). This study leverages the growing temporal extent and quality of the hail observations dataset to explore the likelihood of seeing given hail sizes using the Gumbel distribution.

The paper is structured as follows; Section 2 describes the hail observations dataset and the selection of an appropriate distribution to model extreme hail sizes. Section 3 outlines the characteristics of U.S. hail size data and approaches to negate the limitations of the data. Section 4 describes the fitted extreme value model developed from these data, while Section 5 discusses the estimated return intervals for hail of various sizes and the stability of the fitting approach. In Section 6, we interpret these results in the context of providing an analysis of the hazard posed by hail over the U.S.

2. Datasets and Approach

a. U.S. Hail Observations

U.S. hail reports were taken from the National Centers for Environmental Information archive (Schaefer and Edwards 1999) for the period 1979 to 2013. While these data are available for a longer period (1955-2015), changes in reporting of events influence the dataset in a more pronounced way between 1955 and 1979, and there are several years with no reported hail for many locations (Allen and Tippett 2015). Hail reports were gridded to a 1° x 1° grid to smooth the otherwise noisy observational dataset, which has larger point-to-point variations over the domain considered, but it is possible that in some locations choosing higher resolutions might be appropriate. These reports were gridded for 3-hour periods (0Z-3Z, 3Z-6Z, etc.), with assignment of the largest hail reported for that grid point in each 3-hourly period. This aggregation choice prevents repetitive inclusion from a single thunderstorm, limits biases that would occur due to a higher reporting frequency over cities, and reduces the limitations associated with the small spatial distribution of large hail and sporadic report data. Despite the dataset being likely the longest and most complete national hail record (Allen and Tippett 2015), there are significant non-meteorological inhomogeneties as described above, and thus gridded or point results should be carefully interpreted.

b. Modeling Extremes - Generalized Extreme Value (GEV) Distribution

The GEV distribution (Jenkinson 1955) is a continuous probability distribution which combines the Gumbel, Frechet and Weibull families, also known as type I, II and III extreme value distributions. The relationship is usually presented in the form:

$$F(x) = exp\left\{-\left[1+k\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/k}\right\} \tag{1}$$

where, in this application, $F(x)$ represents the probability of occurrence of a given hail size, and where k, σ and μ are known as the shape, scale and location parameters, respectively. For $k=0$, equation (1) reduces to the Gumbel (EV1) distribution, whereas for positive and negative k, the distribution is respectively Frechet (EV2) and Weibull (EV3).

The three EV limiting behaviors depend on the type of the distribution from which the maxima (or minima) are extracted. Since these parent distributions are often unknown, the GEV flexibility is particularly appealing, allowing all three parameters (including the shape parameter, k) to vary. This flexibility has some drawbacks as well, with limited data making the estimation of the parameters (in particular k) difficult (i.e., Hosking et al. 1985; Martins and Stedinger 2000). Common applications of the Weibull distribution include intense tornadoes and wind speed analyses (e.g., Pavia and O'Brien 1986; Dotzek et al. 2003). Typical Frechet applications have included, among many, rainfall maxima and streamflow data (Coles 2001).

Where there is insufficient information about the extreme tail of a dataset, a popular first order solution is to set $k = 0$, and consider the simplest Gumbel (Type I) distribution (Hosking et al. 1985). Thus, Eqn. 1 in the Type 1 case simplifies to:

$$F(x) = exp\{exp[(x-\mu)/\sigma]\} \tag{2}$$

The location (μ) parameter summarizes the location or shift of the body of extremes (in this case the mean annual maximum hail size), while the scale (σ) parameter describes its statistical dispersion (interannual variability of the annual maximum hail size).

Typical estimation procedures are Maximum Likelihood (MLE), L-moments (also known as Probability Weighted Moments, PWM) and more recent hybridized methods such as Generalized Maximum Likelihood (GMLE), and Generalized PWM (GPWM), with the latter three performing better with small samples (Hosking et al. 1985; Martins and Stedinger 2000; Coles 2001).

For the purposes of this investigation, we considered a Gumbel (Type I) distribution with the MLE and L-moments estimation methods. This decision was made based on testing the value of the shape parameter over the continent, which revealed only small variations from zero and non-significant likelihood-ratio tests for all but seven grid points over the continental U.S. (not shown), suggesting that, when combined with the difficulty in fitting three-parameter models, the Gumbel approach was preferable given the characteristics of the data. Both MLE and L-

Moments approaches have been applied to limited areas for hail in the past (Cox and Armstrong 1981; Smith and Waldvogel 1989; Bardsley 1990; Fraile et al. 2003). We focus here on using gridded annual maxima, which show a more limited spurious temporal trend in the frequency of observations compared to higher frequency data (Allen and Tippett 2015). An implicit assumption of the Gumbel estimation technique is that the data do not exhibit a trend. Otherwise, there is a need for the trend to be accounted for separately, and thus this aspect of the data record was explored. Where data are missing, or less than 30 years of observations are available, the model is not fitted, as this was identified to lead to overly wide confidence intervals, particularly for long return periods.

3. Results

a. U.S. Hail Size Observations

Assessing the characteristics of the hail size record in the past 35 years (1979-2013), the majority of the U.S. east of the Rockies has experienced at least one hail event where the maximum observed hail produced was between 75 and 100mm (3-4 in.), and many places had hail of 112-125mm (4.5-5 in.) diameter (Fig. 1a). Within this area, isolated hail events between 150 and 200mm (6-8 in.) are scattered from southern Texas into South Dakota. The largest differences between the 1979-2013 period and the full hail record (Fig. 1b) are in areas where severe thunderstorms producing large hail are less frequent (Allen et al. 2015; Allen and Tippett 2015), over the northern Plains and the Southeast (Fig. 1a). This extension of the record by 24 years also yields a considerable increase to gridded maxima above 125 mm (5 in.), and numerous sites with at least 175 mm (7 in.) hail, suggesting that much of the Great Plains, Midwest, and Northeast are susceptible to extremely large hail events. Instances of large hail are less common into the Southeast and in general further east where thermodynamic energy (e.g., CAPE) is reduced, owing to decreased lapse rates from repeated diurnal mixing (Allen et al. 2015). In addition, over the Southeast the veracity of large hail size reports has been questioned (Cintineo et al. 2012; Allen and Tippett 2015).

The overall number of observed hail reports has increased remarkably over the past 58 years (Allen and Tippett 2015). Maximum hail size displays less of a trend than the number of reports, however, as many large hail events occurred between 1955-1979 (Fig. 1b). However in the past decade, the largest ten hailstones on record for the entire continental U.S. have changed on several occasions (Blair and Leighton 2012; Blair et al. 2017), suggesting that local maximum possible hail sizes may change as the record extends. This variability is perhaps a result of the increased number of active observers in past decades. Also unsurprising is the incompleteness of the record, as at any one location, large hail size events occur on a rare subset of hail days, which are again a small subset of days in any given year. Thus without a sufficiently long record, there is potential for significant instability in estimations of maximum size of the hazard. The impact of this uncertainty can be considered by comparing the overall maximum hail size on a 1° x 1° grid for the period 1955-2013 to values for 1979-2013 (Figure 1a,b). The largest diameter hail reported for the U.S. occurred in Vivian, South Dakota, and was 200 mm (8 in.). This however may not reflect the upper bound for hail size, as it is plausible that individual stones in a storm may have exceeded this value (Blair and Leighton 2012; Blair et al. 2017). There is likely an upper limit to the maximum hail size suspended by any updraft depending on the updraft speed. This upper limit in turn is controlled by environmental parameters such as the maximum value of CAPE and the strength of vertical wind shear in a storms formative environment, but this value might not be captured by available observations (Ziegler et al. 1983; Nelson 1983, 1987). Other potential limiting factors to maximum hail size include the availability of supercooled liquid water, the ambient temperature, as well as other microphysical effects.

To evaluate the year-to-year consistency in observations of large hail sizes, the mean annual maximum was explored, which is partly dependent on the observational record of years with an annual maxima (Fig. 1c). As the record is sparse spatially and temporally for the period 1955-1978 (Allen and Tippett 2015), we focus on the period 1979-2013. For much of Oklahoma, Kansas, Colorado, Nebraska and the nearby states, the mean annual maximum hail size is 50 mm (2 in.) diameter or larger, with 70-75 mm (2.75 to 3 in.) being more common in both Oklahoma and Texas. The annual mean hail size for much of the eastern U.S. is between 25 and 50 mm (1-2 in.), suggesting that for longer return periods, considerably damaging hail is certainly possible, and can be expected to be likely. The number of years with at least one non-zero hail observation is also examined on a grid point basis, illustrating that for most of the Plains, Midwest and Southeast, more than 30 of the last 35 years meet this criterion (Fig. 1d). West of the Rocky Mountains however, most locations have fewer than 20 annual maxima, and thus are not fitted.

Seasonally, maximum hail size shifts northward in the summer months (Fig. 2a), consistent with the occurrence climatology and the seasonal cycle of CAPE (Allen et al. 2015). However, despite this shift, the incidence of the largest hail sizes is not uncommon through the entire central U.S. during the summer, reflecting climatologically rare events with extreme CAPE and some degree of vertical wind shear (Fig. 2b,c,d). These events introduce localized peaks in the maximum hail size, but the relative fraction leads to a smaller mean maximum in the summer, reflecting the fact that environmental conditions favorable to larger hail are more infrequent during the summer months (Brooks 2013).

An important consideration of the overall hail record highlighted by Allen and Tippett (2015) is its consistency through time. To evaluate this, the maximum size and mean annual maximum size are broken into two segments 1979-1996 and 1997-2013 (Fig. 3a). The magnitude of differences for the maximums suggests that changes are not large or systematic, particularly in the central

U.S. Performing a similar comparison to Fig. 3(a) for 1955-1978 as compared to 1979-2013 provides an overall similar pattern, with isolated larger maxima reflecting rarely occurring events being captured by the longer record (not shown). This similarity suggests that the maximum hail size has a greater sensitivity to record length than the changes in reported size between the two segments, which is in contrast to the finding that many of the largest observed hailstones have occurred in the most recent decade (Blair and Leighton 2012; Blair et al. 2017). The inconsistency between these two characteristics can be resolved as any of the individual stones noted by Blair and Leighton (2012) would influence only a small number of the grid boxes used in the current study. In the southeast U.S., there is a slightly greater change in maximum observed hail size over a large area, which can be explained by a regional trend in environment, or potentially a bias in the reported maximum size arising from recent increases in reports in these regions (Schaefer et al. 2004; Allen and Tippett 2015). Considering the mean annual maximum (Fig. 3b), there is a noticeable contribution from the increasing number of reports of hail of 25mm to 50mm (1-2 in.) diameter. There is also suggestion of increases in mean annual maximum hail size over the Southeast and High Plains reflecting a greater diligence in collecting hail reports to verifying warnings, though these increases are generally small, at 12.5-25 mm (0.5-1 in.). Analyzing this change using a Wilcoxon signed rank test for the difference between the medians (Wilks 2006), a substantial number of points, especially in the Southeast show a significant change at the p-value of ≤ 0.05. This reflects the large increase in the number of observations in this region (where zeroes occur in the first period) in the most recent two decades rather than a trend in size (Allen and Tippett 2015), suggesting the data are stationary and thus the trend does not need to be included in the fitting procedure. The results from this analysis of hail size characteristics suggest that while there are considerable pitfalls with the record over the continental U.S., there are also sufficient data to warrant development of a hail size model.

b. Model Fitting

Examining the empirical cumulative distribution function (CDF; Fig. 4a), the distribution of hail size is heavily quantized as a result of reference objects, most notably in the 19 to 25 mm (0.75 to 1.00 in.) range, reflecting the minimum thresholds for severe hail reports (19 mm or 0.75 in. 1979-2010, 25 mm or 1.00 in. 2010-2013) and for golfball sized hail (45 mm or 1.75 in.) and to other extents for other reference objects (e.g., baseball, 70 mm or 2.75 in.). These characteristics suggest that care needs to be taken in subsequent model fitting. As the desired model is a Gumbel distribution of the annual maximum hail size over each 1 degree cell, several approaches are needed to reduce the sensitivity to quantization of the data and limited sample size. To address the quantization, the data were dithered, where-by a small random uniform amount is added to, or subtracted from, the observed value before the whole set of observations is used to determine the sample annual maxima (Fig. 4b). To avoid overly large biases at small hail diameters, a linear fitted random uniform correction was developed, which uses a dithering process of the form: $y_{new} = y_{old} + y_{dithered}$, where a random value is sampled between $y_{dithered} = \pm(0.247 y_{old} + 0.0279)$, capped at ± 0.5 in. following testing of a range of values and fitting lines of regression to ensure minimal influence on the overall size distribution. This results in a hail size error range at 19 mm (0.75 in.) of $\pm$6.3 mm (0.25 in.), and at 45 mm (1.75 in.) or greater bounded at $\pm$13 mm (0.5 in.). This dithering equivalent to a fuzzy-error in hail size for parameter estimation serves two purposes: to reduce the natural quantization of the data, and to offset issues with size estimation errors that bias the hail record, such as parallax in measurements and low hail size bias (Allen and Tippett 2015; Blair et al. 2017). This random variation results in a preserved distribution, but overall smoother empirical CDF (Fig. 4a) and better representation of the fitted distribution on a quantile-quantile plot. The second step of quality control is only to fit data grid points with 30 or more observations, to ensure that a sufficient sample exists to accurately and reliably determine the Gumbel location and scale parameters while minimizing the parameter errors and increasing fit confidence.

Performing this fitting of the Gumbel distribution using the MLE procedure, we obtain a gridded set of location and scale parameters over the continental U.S. (Fig. 5a,b). Generally higher values for scale are found across the Great Plains states, particularly over Texas, Oklahoma and Kansas, reflecting more regular return rates of larger hail sizes. To a lesser extent this is also found over the Southeast U.S. The differences between neighboring grid points are considerable over the domain with parameter estimation errors of 15-20% (Fig. 5c), reflecting the difficulty in estimating the scale parameter with limited observation sets, and its sensitivity to outliers. This variability between the nearby grid points is particularly noticeable for locations with significant urban population (e.g., Dallas-Fort Worth, Amarillo, Lubbock, Wichita, Oklahoma City) that increases the likelihood of large hail size reports. In contrast to the relatively limited area with high scale parameters, the location parameter is higher over a larger area, including the Plains and through the Midwest and Southeast, with the highest values from central Texas to the Dakotas. The standard error in the location parameter estimates is between 2.5 and 5% over much of the domain except in locations which receive fewer hail reports, suggesting a greater confidence on expected maximum sizes from the sample available (Fig. 5d). A test of a random set of 30 dithered fits shows minimal to negligible contributions to the standard error in using the dithering procedure (not shown). As another test of performance, the mean of the Gumbel distribution $(Gumbel_{mean} = 0.5772 scale + location)$ is compared to the mean of the annual maxima to which it was fitted (Fig. 6). This revealed that the Gumbel mean values were close to the expected result, but somewhat higher than the observations, potentially reflecting the limitations of the record, or a tendency of the Gumbel fitted model to overestimate the hail size.

Other fitting approaches can also be used to determine the point Gumbel distribution that may be able to leverage greater confidence from the limited observations (e.g., Probability Weighted Moments or L-moments). To evaluate whether this difference in fitting procedure influences the result compared to the MLE approach, identical data were fitted using L-Moments, which suggests that there is little to be gained by using the second procedure given the existing limitations of the data (Fig. 5e,f). This lack of distinction between the two methods is consistent with the prior analysis of hail-pad return levels by Fraile et al. (2003), and thus here we focus on results from the MLE approach.

c. Return Levels and Stability Analysis

To assess the suitability of the models to produce realistic return periods, several evaluations of performance were needed. First, the return levels and the confidence intervals for four regional locations which are co-located with highly populated observational records were analyzed (Fig. 7). Individual grid data show a larger spread in the potential regional fits for the surrounding grid boxes, reflecting variations in the sample size and relatively infrequent returns of larger hail sizes with decreasing confidence at longer return intervals. The grid box encompassing Oklahoma City, Oklahoma is chosen as the long-term station representing the Great Plains (Figs. 7a). Observed annual maxima in the area range between 25 and 127 mm (1 and 5 in.), with heavy quantization toward both golfball (45 mm or 1.75 in.) and baseball diameters (70 mm or 2.75 in.) over the full 35 year record. Exploring the return levels yields a 127 mm (5 in.) stone at a 40-year return interval, with the remainder of observations pointing to a stably fitted model. The bounds of the surrounding points indicate that, at the 2-year interval, 51mm (2 in.) hail is expected, while at the 10-year return period, hail is expected to be within 64-114mm (2.5-4.5 in.) with relatively strong confidence based on the relatively narrow confidence interval of fit results. As would be expected given the paucity of the largest of hail observations, the greatest range in both regions and confidence intervals is seen outside of the 50-year return levels, with the regional spread as large as between 101 and 178mm (4 and 7 in.) at the 200-year interval. To explore performance over the Northern Plains, the point nearest Pierre, South Dakota was examined (Fig. 7b). This point includes, in the observed record, the largest verified hail size observation in the U.S. of 203 mm (8 in.), and thus can be used to explore whether a single outlying observation heavily skews the distribution. As for Oklahoma, there is heavy quantization in the golfball category, with the three largest stones found to be 114, 114 and 200mm (4.5, 4.5 and 8 in.) in diameter. Bounds from the surrounding grid points are tighter than for the Oklahoma case, with a range at the 200-year return level of 89-197mm (3.5-7.75 in.), which encompasses the record size observed near Vivian, South Dakota. As a third evaluation point to explore performance over the Southeast U.S., we consider the return levels around Atlanta, Georgia. There is extreme quantization at the golfball level, with the largest observed size of 82.5mm (3.25 in.). The concentration at smaller hail sizes over a wider area is reflected by the narrower range at longer return levels over the surrounding region and tighter corresponding confidence intervals, with a maximum 200-year return level between 76 and 152 mm (3 and 6 in.). Golfball (45 mm or 1.75 in.) hail has a 2-year return period at Atlanta, with values of up to 76 mm (3 in.) expected at intervals as short as 20 years. Finally, the model is evaluated over the mid-Atlantic and Northeast regions, using the grid closest to Philadelphia, Pennsylvania. Hail sizes in this region are again comparatively smaller, with most hail observed close to the minimal severe thresholds, and the largest sizes on record between golfball and baseball (45-70 mm or 1.75-2.75 in.). The wider spread in this region reflects the variations induced by a larger number of reports above 1 in. (25 mm), with the remaining sample at the minimum severe level, which leads to point-to-point variations in the estimation due to a non-continuous observational distribution despite dithering. Nonetheless, 76 mm (3 in.) hail is certainly possible in Philadelphia and surrounds, with this size stone occurring between the 10- and 200-year return levels, suggesting at least some degree of regularity. The spatial variability of return sizes at given probabilities should not be interpreted on a point basis in the unsmoothed form, as point-to-point sample variation can lead to larger variations in estimated return level, especially at the longest returns. A 100-year return period implies that there is a 0.01 probability at any point of a given maximum hail size within a 1° x 1° grid box, however at higher resolution and consequently smaller grid area (which is not examined in this study), the 100-year value could be equal to this or smaller. While there is a considerable spread in the confidence intervals, there is a strong degree of consistency of hail at least in excess of 51 mm (2 in.) at the 10-year return level for all locations, reflecting a hazard to property and vehicles.

Generalizing this analysis to the entire fitted domain, we evaluate the fitted point distribution to determine the hail size (in inches) at the respective return periods (Fig. 8). This reveals that for most locations east of the Rockies, that hail sizes at the 2-year return level are over 25 mm (1 in.), and a large majority of grid points in the Great Plains exceed 50 mm (2 in., Fig. 8b), with values reaching as high as 76-101mm (3-4 in.) at the 5-year interval. Increasing the return interval to the 10-year level, sizes generally range between 76 and 127mm (3 and 5 in.), with the higher values mostly confined to grid points in the Great Plains. Given the length of the record (35 years), the 20 and 50-year return values most closely resemble the maximum hail size observations, with higher values for many points. This is as would be expected for a fitted distribution, as there is considerable point uncertainty in event occurrence, especially when combined with the existence of a number of rarer large observations (Fig. 8a,d,e). The values at these levels range between 75 and 152 mm (3-6 in.), suggesting that the model is representative of the data to which it is fitted and includes extensions into the Southeast and Midwest. At longer return periods (100-year return), hail sizes of 150mm (6 in.) are identified for most of the fitted hail domain outside of the northeast

U.S., with the highest values particularly concentrated through the Great Plains and toward the Canadian border. Extending this to an extremely long return period with low-confidence, at the 200-year level (Fig. 8), large portions of the Great Plains including Oklahoma, Kansas, and Texas would suggest return diameters of 152-203mm (6-8 in.) or more, which are consistent with the largest values in the existing hail record. Analyzing throughout the return periods, there is low probability but high magnitude potential over the northern Great Plains and Midwest, reflecting rarer excursions of environmental parameters favorable to the development of storms producing this diameter hail. Over much of the domain east of the Rockies, including through the Southeast, east of the Applachian Mountains, and the eastern population centers, 200-year return levels are well in excess of 101mm (4 in.), suggesting the potential for catastrophic hail storms in areas which comparatively rarely experience these events. As with all extreme value estimates of return levels, the largest potential errors exist in the outer tails, especially when sample size is limited. Nonetheless, the fact that 101mm (4 in.) or greater measurements are not unusual anywhere within the domain (consistent with the estimated 20-50 year or greater return period), suggests that longer return levels are not unreasonable, but must be viewed with greater uncertainty.

The gridbox-to-gridbox variations suggest a more pronounced influence of spatial observational quality on the return period estimates rather than reflecting robust differences in hail size at varying return levels. To offset this, we apply a 2-D Gaussian Kernel smoother to the return period data with a σ=1.00 (1 degree smoother) kernel bandwidth to produce a more spatially consistent hazard profile (Fig. 9). The smoothed spatial return map for the maximum hail size for 1979-2013 suggests peak values of approximately 127-152mm (5-6 in.), with the highest likelihood for these hail sizes over the central to northern Great Plains and extending into both the upper Midwest and Southeast. Values as high as 100mm (4 in.) extend through New Mexico into Arizona and to the Canadian border and into Montana. At the 2-year return level, much of the Great Plains exhibit values up to 50mm (2 in.), with a steep gradient towards the east of the Rockies and fairly uniform coverage that extends from Montana to Southern Maine, south to central Florida and stretching west into the desert southwest (Fig. 9b). At the 10-year level, much of the region east of the CONUS has return hail sizes of 51-76 mm (2-3 in.), and 101 mm (4 in.) over the central Great Plains. This extension of significant hail (50 mm or 2 in.) is found into southern New York, Pennsylvania and New Jersey. Return values increase substantially over the Great Plains at the 20- and 50- year return levels, with a slower increase over much of the remainder of the eastern CONUS, with the smoothed 50-year return level qualitatively similar to the maximum observed hail size, and the largest difference being the reduction in northern extent (reflecting the tendency of the smoothing kernel to flatten absolute point maxima). There is also some smearing by the smoothing procedure of four grid boxes in Arizona with sufficient very large hail measurements to justify fitting the model, where hail up to 101mm (4 in.) has been observed in the recent past. Even following the smoothing procedure, hail sizes over 100-125mm (4-5 in.) are likely over much of the eastern CONUS at return levels of over 100 years, with 150mm (6 in.) appearing a likely value for the more convectively prone regions of the Great Plains, Midwest and Southeast, rising to 175-200mm (7-8 in.) in the central Great Plains (Fig. 9g,h).

Testing the modeled hail sizes further, the return periods for hail of 25mm (1 in.) in diameter are evaluated by comparing them with the annual occurrence rate of a proxy for hail derived from environmental parameters (Allen et al. 2015). This proxy produces a spatially unbiased hail climatology using a combination of monthly environmental parameters favorable to hail development (CAPE, 0-3km Storm Relative Helicity, convective precipitation, mean 0-90mb above ground level specific humidity) in a Poisson regression to simulate the monthly frequency of hail $\geq$25 mm (1 in.). The point comparison suggests that this hail size has a return period of one year over most of the U.S., particularly over the Great Plains, Midwest and Southeast (Figure 10a). For comparison, the Gumbel distribution used here cannot provide a return period of less than 1 year as it is defined by $(1/p)$, and thus the environmentally derived rate should be considered equivalent for all values between 0.10 and 1.00 (i.e. where one or more $\geq$25mm (1 in.) hail storms occur per year). This comparison reveals that the two maps are spatially consistent. The inverse probability derived from the environmental proxy suggests a less than one event per year return rate over much of the Great Plains and remainder of the eastern CONUS (Figure 10b), which would imply the Gumbel model underestimates the return rate of the hail hazard for $\geq$25mm (1 in.) hail and smaller sizes.

Following this positive test, the evaluation threshold is raised for temporal return period through the respective sizes of interest (38, 45, 51, 76 mm or 1.5, 1.75, 2, 3 in.). Even for hail sizes as large as 75mm (3 in., Fig. 11), the minimum return period is between 2 and 5 years for much of the Great Plains, while hail stone diameters of up to 50mm (2 in.) have return periods of 1-3 years over the Great Plains, Southeast and Midwest. These hail sizes appear to be more infrequent along the Appalachian Mountains and into the Northeast, particularly for hail sizes in excess of 50mm (2 in.). For the lower thresholds, much of the domain experiences hail of up to golf ball diameter (45mm or 1.75 in.) at a likelihood of an event every 1-2 years throughout the Great Plains. These results suggest that hailstones capable of producing considerable damage to structures, vehicles and property ($\geq$45mm or 1.75 in., Brown et al. (2015)) are relatively commonplace on a yearly basis for the Great Plains and Southeast, reflecting a likely hazard irrespective of the available observations. While this does not necessarily imply certainty at any location such as a sub-grid scale city given grid boxes of $\sim$ 100x100 km, it does suggest that these large hail events have a higher rate of occurrence than may have been anticipated based on existing observationally derived climatologies.

To evaluate these fitted distributions for performance in representing quantiles, next continental scatter diagrams of observed (percentiles from all hail observations) and modeled (percentiles derived from annual maxima) hail size were explored (Figure 12). As the observational data are limited in quantity, we restrict this to the 80th, 90th, 95th, and 98th percentiles at each grid point (5,10, 20, 50 year return levels) for the undithered and dithered observations to both the point and smoothed model return periods. Against both the quantized and dithered observed quantiles, the point model performed well at the 80th, 90th and 95th percentiles with high degrees of correlation (Figure 12). At the 98th percentile, the model also appeared to perform relatively well, however, this is harder to assess as the observations are less representative due to the limited sample, which can explain the slight upward bias in the modeled quantiles. Comparison to the dithered data results in a considerably higher degree of fit, suggesting that it provides a more representative depiction of hail size quantiles over the domain. This supports the conclusion that performance across the U.S. is very good out to the 20-year return level, and perhaps slightly overestimating the return size at the 50-year return period if the 35 years of observations are representative of the true distribution. Considering instead the smooth return levels sampled on a point basis, there is a greater degree of spread in the compared points where hail size quantiles are both under and over-estimated relative to observed quantiles owing to the smoothing of the sample (Figure 12c,f,i,l). Nonetheless, the degree of correlation is significantly high between the observed and smoothed data, with relatively small point variations particularly at the 80th, 90th, and 95th percentiles.

Finally the frequency with which observed hail sizes do not exceed the model percentiles (e.g. ideally 95% of observed hail sizes are below the modeled 95th percentile) are summarized as percentiles of non-exceedance. Each grid point was compared to the fitted model through the range of quantiles over the continental U.S. and regionally to establish any localized biases at a given return level. Over the entire domain (Figure 13a), the grid box model shows a relatively good fit through the middle quantiles (80th-99th) with upward divergence at the 50th and 99th percentiles, the lowest and highest values shown. This suggested that the grid box model may be overestimating the size of hail for a given return period, which is consistent with the comparison of the Gumbel mean and the mean annual maximum values. The result is somewhat unexpected given that this part of the distribution has the most available data for evaluation (though data limitations influence a number of the fitted grid points), but may also reflect the influence of a lower bound on hail size incurred by the severe thresholds that preclude recording of smaller diameters (Allen and Tippett 2015). Another potential explanation is that the tendency of the Gumbel distribution to weigh toward the center of the data leads to the fitted curve being skewed at the extreme tail and the lower return frequencies. The values at the higher quantiles also display this divergence related to the limitations in the maximal observed sizes of the distribution. On a regional basis (Figure 13b), the model appears to perform well over each of the respective NOAA climate regions (Allen et al. 2015), with similar positive biases over regions with fewer observations over the record length compared to the central Great Plains (e.g., the Northeast).

4. Discussion

A climatology of large hail occurrence and maximum size potential has been derived from Storm Data using observations from 1979-2013, providing insight to hazard modeling for large hail. The spatial hazard maps of hail size return intervals generated using this approach illustrate that a simple EVD can produce a first of its kind spatial model for observed hail size return intervals for the central and eastern CONUS. However, it has also been demonstrated that it is necessary to carefully explore the limitations of the observed hail record and statistical techniques to accurately ascertain the hazard, or the reasons for point to point variability in the results obtained using this comparatively crude approach. Nonetheless, the performance of the EVD model based on the evaluation conducted here would indicate that it provides a useful analysis of the hazard posed by large hail in the U.S., and higher than expected potential for large portions of the country compared to observational climatology, with the east of the country exposed to hail up to 75mm (3 in.) diameters on a 20- to 50-year interval, and over the Great Plains on a 10- year recurrence period, and exposed to damaging hail (45 mm/1.75 in. or greater) every one to two years.

Perhaps the most stark limitation of this technique and assessment of U.S. hail size data is the significant quantization present in the size of hail reports resulting from a limited diversity of reference objects available for observers and the very basis of the reporting system (Blair and Leighton 2012; Allen and Tippett 2015; Blair et al. 2017). This challenge can be mitigated by dithering to some extent, but this data processing step introduces additional potential errors (albeit small) in the estimation of fitted distribution parameters. On the other hand, the step allows for a fairer evaluation of the modeled quantiles compared to the quantized observations, and likely reflects the errors that are naturally introduced by observers (Blair and Leighton 2012; Blair et al. 2017). Theoretically and physically, there must be an upper bound to the largest possible hail size at any one location (Knight and Knight 2001), as updraft speed cannot increase without bound for realistic environments. However, for most locations the sample is incomplete or not reflective of the narrow swathes of the largest stones for each storm (Blair et al. 2017), and thus may be under-representative. Additionally the spatial distribution of observations reflects the characteristics of population, not only the actual distribution of hail size observations. This results in errors in the fitted scale and location parameters, particularly where fewer observations or longer return sizes are found. While this error can be mitigated by smoothing procedures or possibly sampling over a wider region to fill out the distribution, there is the potential that this step over or understates the hail size potential. This suggests a need to divorce the observations from the hail size model, perhaps by using environmental distributions (e.g., Brooks et al. 2003; Gilleland et al. 2013), particularly where data are limited or do not exist.

The limitations of the observational data also lead to issues in the parameter estimation, as there are insufficient samples in many locations to explore the characteristic of the tail of the distribution. It is likely that if additional data were available to constrain parameter estimations, a more general GEV model might be possible, which may include a tailing behavior toward the Weibull distribution like many processes of increasing rarity (Fraile et al. 1992; Dotzek et al. 2009). The point Gumbel model is one possible solution with currently available data. It is also plausible that this result may be sensitive to the spatial resolution of the grid chosen, which merits future investigation. A further complication is that it is not clear how these grid box results translate to the true probability of experiencing hail at the sub-grid scale. However the nature of the Gumbel as a collector distribution and lack of tailing characteristics in the extremes to provide an upper bound can mean over- or underestimation of hail sizes depending on the available fit using the existing data. This suggests that as future data becomes available, it may be possible to improve on the modeled result here and possibly that the point distribution will converge to a Weibull distribution, or be better modeled using a Generalized Pareto Distribution. However, at the current juncture neither of these approaches produced stable results due to large grid box-to-grid box variations in the estimated shape parameters. Known long return period observations (35 years) appear to be consistent with the modeled distribution, suggesting that the outer tail is being reasonably well captured by the Gumbel model. Despite the limitations of the Gumbel approach, for the 2-100 year return periods, the regional and national performance metrics provide confidence that this model for U.S. hail size performs well in assessing the threat posed to the U.S. by large hail events.

Acknowledgments. The authors acknowledge support from FM Global in conducting this research and preparing this manuscript. JTA conducted the data analysis, production of figures and

writing of the paper. All authors contributed to the research design and writing and editing of the final paper.

References

Allen, J. T., and M. K. Tippett, 2015: The characteristics of United States Hail Reports: 1955-2014. *Electron. J. Severe Storms Meteorol.*, **10**, 1–31.

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015: An empirical model relating U.S. monthly hail occurrence to large-scale meteorological environment. *Journal of Advances in Modeling Earth Systems*, **7 (1)**, 226–243, doi:10.1002/2014MS000397.

Bardsley, W., 1990: On the maximum observed hailstone size. *J. Appl. Meteor.*, **29 (11)**, 1185–1187.

Blair, S. F., and J. W. Leighton, 2012: Creating high-resolution hail datasets using social media and post-storm ground surveys. *Electron. J. Oper. Meteor*, **13 (3)**, 32–45.

Blair, S. F., and Coauthors, 2017: High-resolution hail observations: Implications for NWS warning operations. *Wea. Forecasting., In Press*.

Brooks, H. E., 2013: Severe thunderstorms and climate change. *Atmos. Res.*, **123**, 129–138, doi:10.1016/j.atmosres.2012.04.002.

Brooks, H. E., J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67-68**, 73–94.

Brown, T. M., W. H. Pogorzelski, and I. M. Giammanco, 2015: Evaluating hail damage using property insurance claims data. *Weather, Climate, and Society*, **7 (3)**, 197–210.

Changnon, S. A., 1999: Data and approaches for determining hail risk in the contiguous United States. *J. Appl. Meteor.*, **38**, 1730–1739.

Changnon, S. A., 2008: Temporal and spatial distributions of damaging hail in the continental United States. *Physical Geography*, **29 (4)**, 341–350.

Changnon, S. A., and D. Changnon, 2000: Long-term fluctuations in hail incidences in the United States. *J. Climate*, **13**, 658–664.

Changnon Jr, S. A., 1966: Note on recording hail incidences. *J. Appl. Meteor.*, **5 (6)**, 899–901.

Changnon Jr, S. A., 1977: The scales of hail. *J. Appl. Meteor.*, **16 (6)**, 626–648.

Cheng, L., M. English, and R. Wong, 1985: Hailstone size distributions and their relationship to storm thermodynamics. *J. Climate Appl. Meteor.*, **24 (10)**, 1059–1067.

Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective high-resolution hail climatology of the contiguous United States. *Wea. Forecasting*, **27**, 1235–1248, doi:10.1175/WAF-D-11-00151.1.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Cox, M., and P. R. Armstrong, 1981: A statistical model for the incidence of large hailstones on solar collectors. *Solar energy*, **26 (2)**, 97–111.

Doswell, C. A., III, 2001: Severe convective storms—an overview. *Meteorological Monographs*, **28**, 1–26, doi:10.1175/0065-9401-28.50.1.

Doswell, C. A., III, H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595, doi:10.1175/WAF866.1.

Dotzek, N., J. Grieser, and H. E. Brooks, 2003: Statistical modeling of tornado intensity distributions. *Atmos. Res.*, **67**, 163–187.

Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. *Atmos. Res.*, **93**, 575–586, doi:http://dx.doi.org/10.1016/j.atmosres.2008.10.020.

Fisher, R. A., and L. H. C. Tippett, 1928: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180–290.

Fraile, R., C. Berthet, J. Dessens, and J. L. Sánchez, 2003: Return periods of severe hailfalls computed from hailpad data. *Atmospheric Research*, **67**, 189–202.

Fraile, R., A. Castro, and J. Sánchez, 1992: Analysis of hailstone size distributions from a hailpad network. *Atmospheric research*, **28 (3-4)**, 311–326.

Frechet, M., 1927: Sur la loi de probabilité‘ de l'cart maximum. *Ann. de la Soc. Polonaise de Math.*, **6**, 93–117.

Gilleland, E., B. G. Brown, and C. M. Ammann, 2013: Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics*, **24 (6)**, 418–432.

Gumbel, E. J., 1958: *Statistics of Extremes*. Columbia University Press, New York, USA.

Gunturi, P., and M. K. Tippett, 2017: Managing severe thunderstorm risk: Impact of ENSO on U.S. tornado and hail frequencies. Tech. rep., WillisRe.

Heymsfield, A. J., I. M. Giammanco, and R. Wright, 2014: Terminal velocities and kinetic energies of natural hailstones. *Geophys. Res. Lett.*, **41 (23)**, 8666–8672.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood, 1985: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, pp. 251–261.

Jenkinson, A. F., 1955: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.*, **81**, 158–171.

Kelly, D. L., J. T. Schaefer, and C. A. Doswell III, 1985: Climatology of nontornadic severe thunderstorm events in the United States. *Mon. Wea. Rev.*, **113**, 1997–2014.

Knight, C. A., and N. C. Knight, 2001: Hailstorms. *Severe Convective Storms, Meteorological Monographs*, **28**, 223–248.

Martins, E. S., and J. R. Stedinger, 2000: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, **36 (3)**, 737–744, doi: 10.1029/1999WR900330.

McMaster, H., 1999: The potential impact of global warming on hail losses to winter cereal crops in New South Wales. *Climatic Change*, **43**, 455–476.

Morgan Jr, G. M., and N. G. Towery, 1975: Small-scale variability of hail and its significance for hail prevention experiments. *J. Appl. Meteor.*, **14 (5)**, 763–770.

Munich RE, 2015: *Severe convective storms and hail - Icy cricket balls from above, Available at: http://www.munichre.com/australia/australia-natural-hazards/australia-storm/hailstorm/index.html, Accessed on 25th November 2015*. Munich RE.

Nelson, S. P., 1983: The influence of storm flow structure on hail growth. *J. Atmos. Sci.*, **40 (8)**, 1965–1983.

Nelson, S. P., 1987: The hybrid multicellular-supercellular storm-an efficient hail producer. part ii. general characteristics and implications for hail growth. *J. Atmos. Sci.*, **44 (15)**, 2060–2073.

Nelson, S. P., and S. K. Young, 1979: Characteristics of Oklahoma hailfalls and hailstorms. *Journal of Applied Meteorology*, **18 (3)**, 339–347.

Ortega, L., T. M. Smith, K. L. Manross, K. A. Scharfenberg, A. Witt, A. G. Kolodziej, and J. J. Gourley, 2009: The severe hazards analysis and verification experiment. *Bull. Am. Meteor. Soc.*, **90**, 1519–1530.

Palutikof, J., B. Brabson, D. Lister, and S. Adcock, 1999: A review of methods to calculate extreme wind speeds. *Meteor. Appl.*, **6 (2)**, 119–132.

Pavia, E. G., and J. J. O'Brien, 1986: Weibull statistics of wind speed over the ocean. *J. Appl. Meteor. Climatol.*, **25 (10)**, 1324–1332.

Sánchez, J., R. Fraile, J. De La Madrid, M. De La Fuente, P. Rodríguez, and A. Castro, 1996: Crop damage: The hail size factor. *J. Appl. Meteor.*, **35 (9)**, 1535–1541.

Sander, J., J. Eichner, E. Faust, and M. Steuer, 2013: Rising variability in thunderstorm-related US losses as a reflection of changes in large-scale thunderstorm forcing. *Wea. Clim. Soc.*, **5 (4)**, 317–331.

Schaefer, J. T., and R. Edwards, 1999: The SPC Tornado/Severe Thunderstorm Database. *Preprints, 11th Conf. Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 215–220.

Schaefer, J. T., J. J. Levit, S. J. Weiss, and D. W. McCarthy, 2004: The frequency of large hail over the contiguous United States. *Preprints, 14th Conf. Applied Climatology, Seattle, WA, Amer. Meteor. Soc*, 7pp.

Smith, P. L., and A. Waldvogel, 1989: On determinations of maximum hailstone sizes from hallpad observations. *J. Appl. Meteor.*, **28 (1)**, 71–76.

Strong, G., and E. Lozowski, 1977: An Alberta study to objectively measure hailfall intensity. *Atmosphere*, **15 (1)**, 33–53.

Swiss RE, 2017: *Natural catastrophes and man-made disasters in 2016: A year of widespread damages.* Swiss RE, URL http://www.swissre.com/library/editors-pick/sigma_2_2017_en.html.

Wilks, D. S., 2006: Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteorol. Appl.*, **13**, 243–256.

Ziegler, C. L., P. S. Ray, and N. C. Knight, 1983: Hail growth in an Oklahoma multicell storm. *J. Atmos. Sci.*, **40 (7)**, 1768–1791.

LIST OF FIGURES

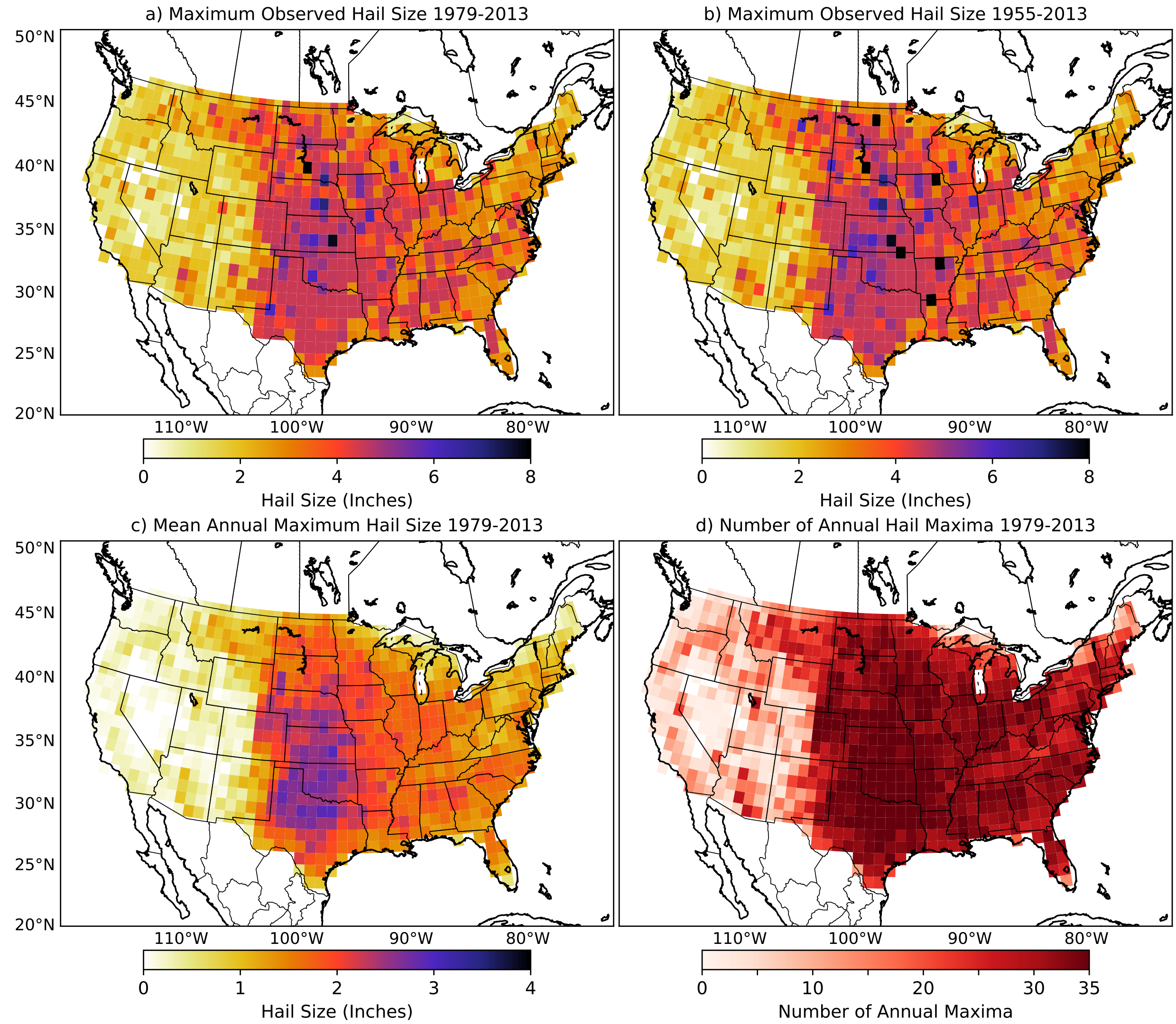
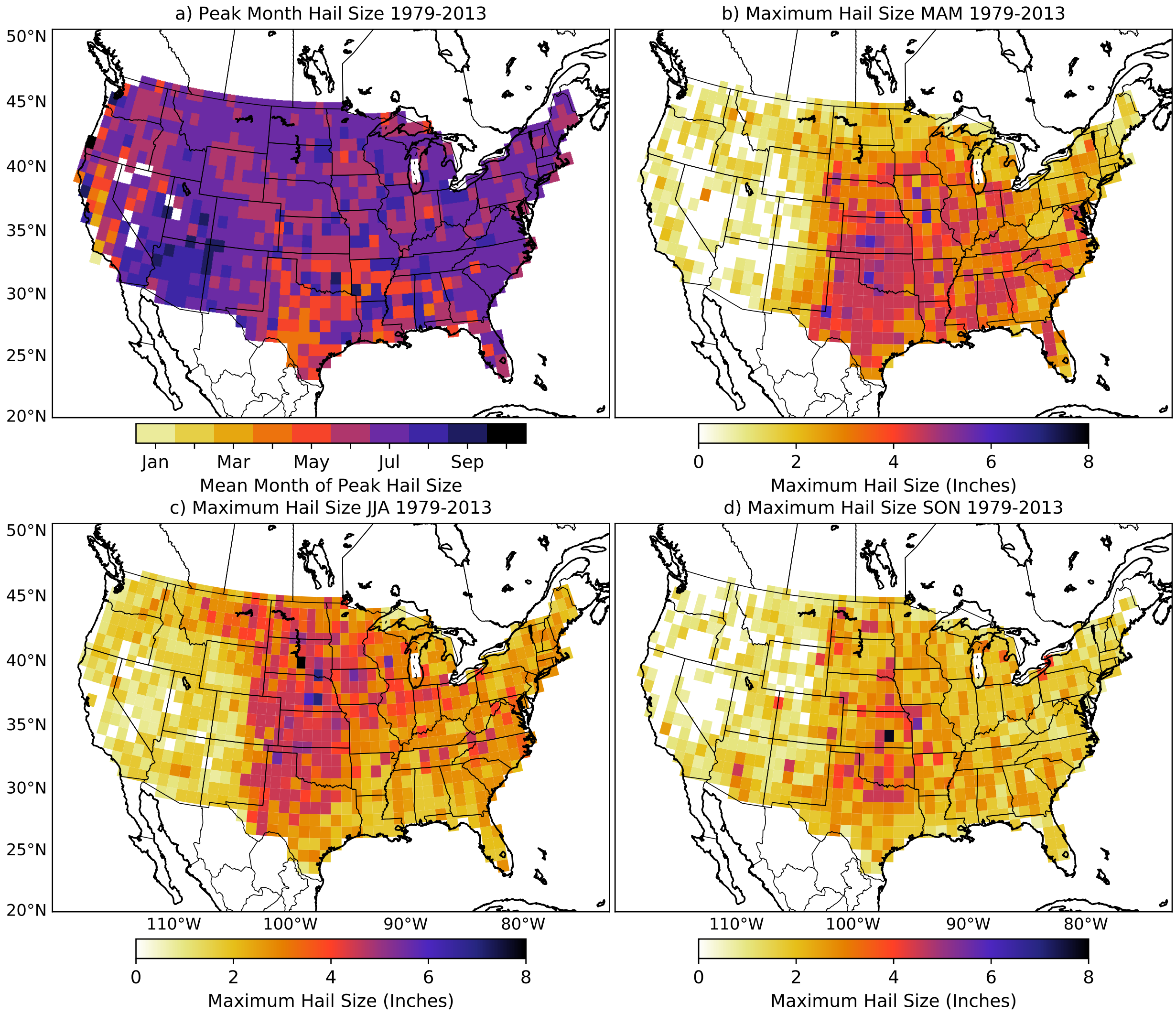


FIG. 1. Maximum observed U.S. hail size from the NCEI dataset for the periods a) 1979-2013 and b) 1955-2013. c) Mean annual maximum hail size 1979-2013. d) Number of annual maxima in each grid box over the period 1979-2013.



728 FIG. 2. Seasonality of maximum hail size from 1979-2013 in terms of the a) Peak month of hail size (month with the largest hail size) based on the gridbox mean of non-zero months for the period 1979-2013, b) March, April, May (MAM) maximum hail size, c) June, July, August (JJA) maximum hail size, and d) September, October November (SON) maximum hail size.

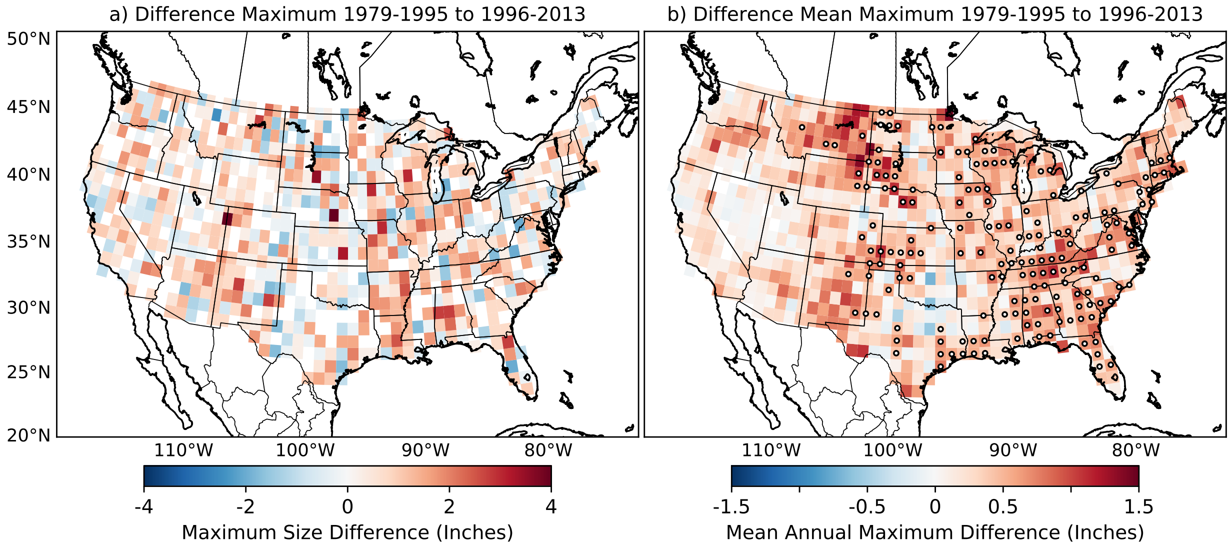
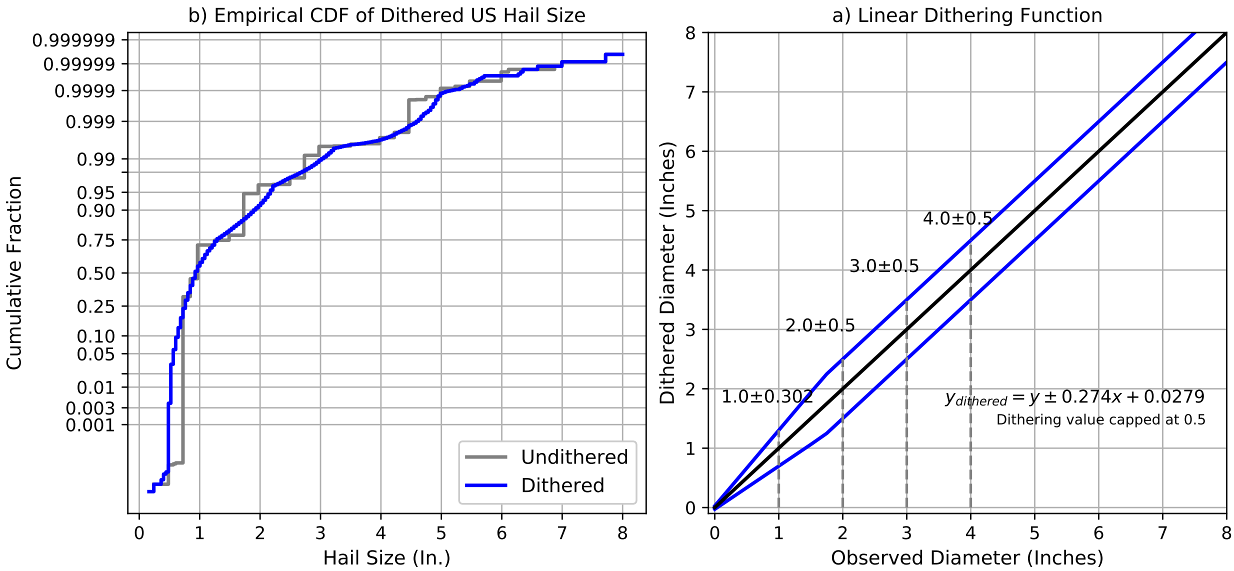


FIG. 3. Changes between the period 1979-1995 and 1996-2013 in terms of a) The largest recorded annual maximum hail size and b) Mean annual maximum hail size over the U.S. Stippling shows where a Wilcoxon signed-rank test of medians has a p-value of less than 0.05, and reflects a rejection of the hypothesis of no change in median between the periods.



736 FIG. 4. a) Illustration of the impact of capped linear dithering on hail size and b) the resulting empirical CDF
737 of U.S. hail size observations following the dithering procedure.

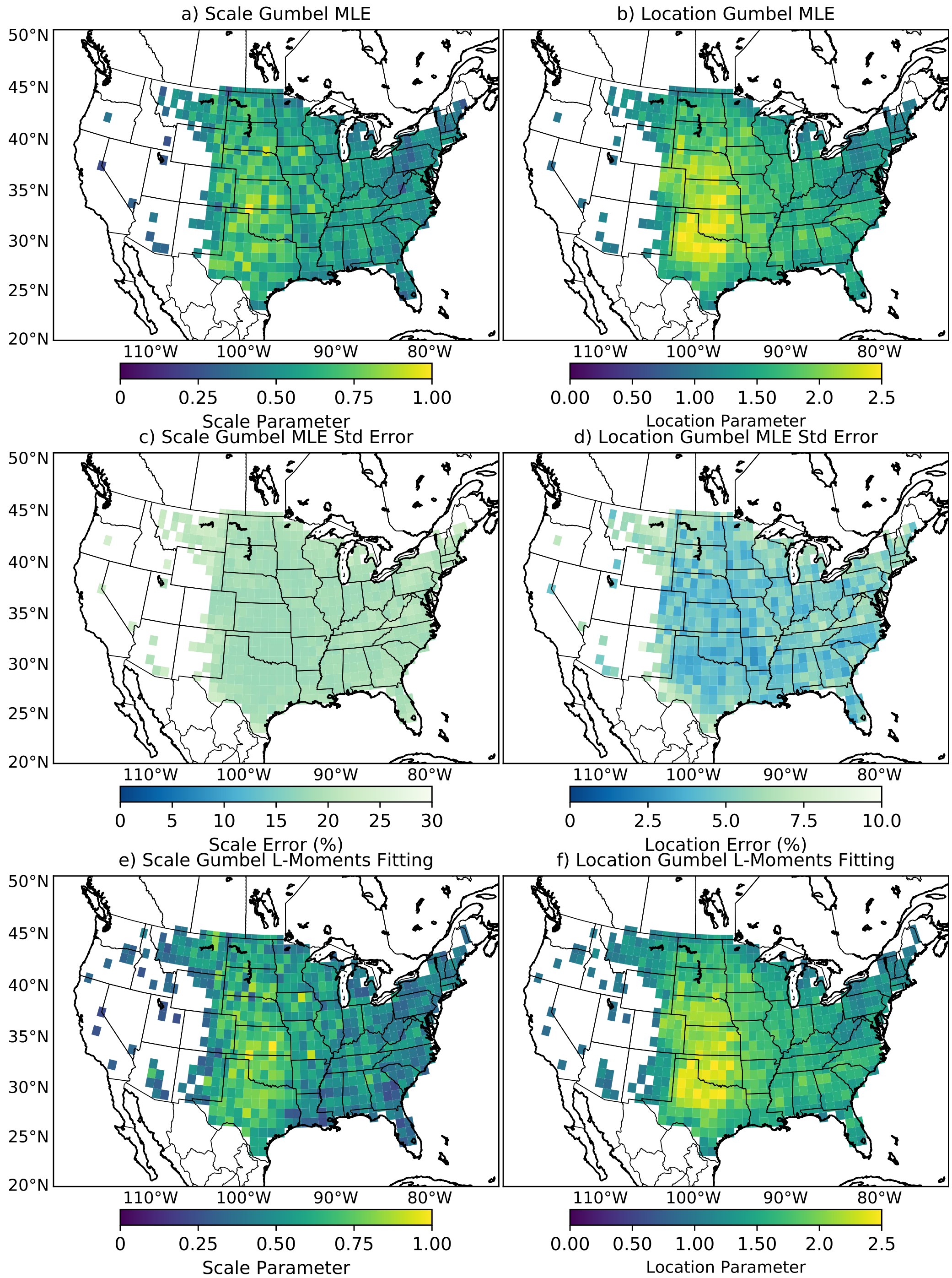


FIG. 5. Gumbel distribution parameter estimates and their standard errors for the point fit of dithered annual maxima observations with more than 30 years for the period 1979-2013. a) Scale parameter using MLE fitting, b) Location parameter using MLE fitting, c) Percentage standard error in scale parameter estimates from MLE, d) Percentage standard error in location parameter estimates from MLE, e) Scale parameter using L-Moments fitting, f) Location parameter using L-Moments fitting.

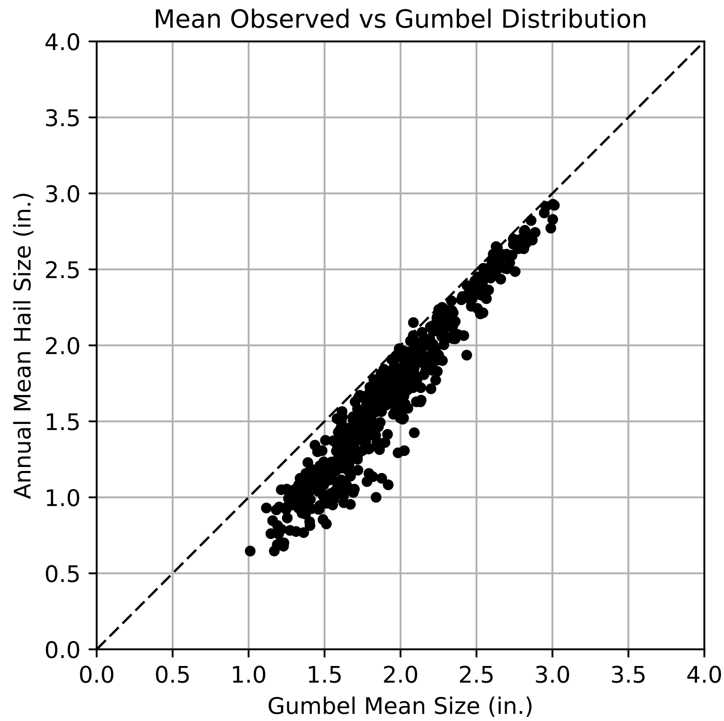


FIG. 6. Comparison of the mean annual maximum hail size at each grid box used for the model fitting and the mean of the derived Gumbel distribution at each point, determined by $(Gumbel_{mean} = 0.5772scale + location)$

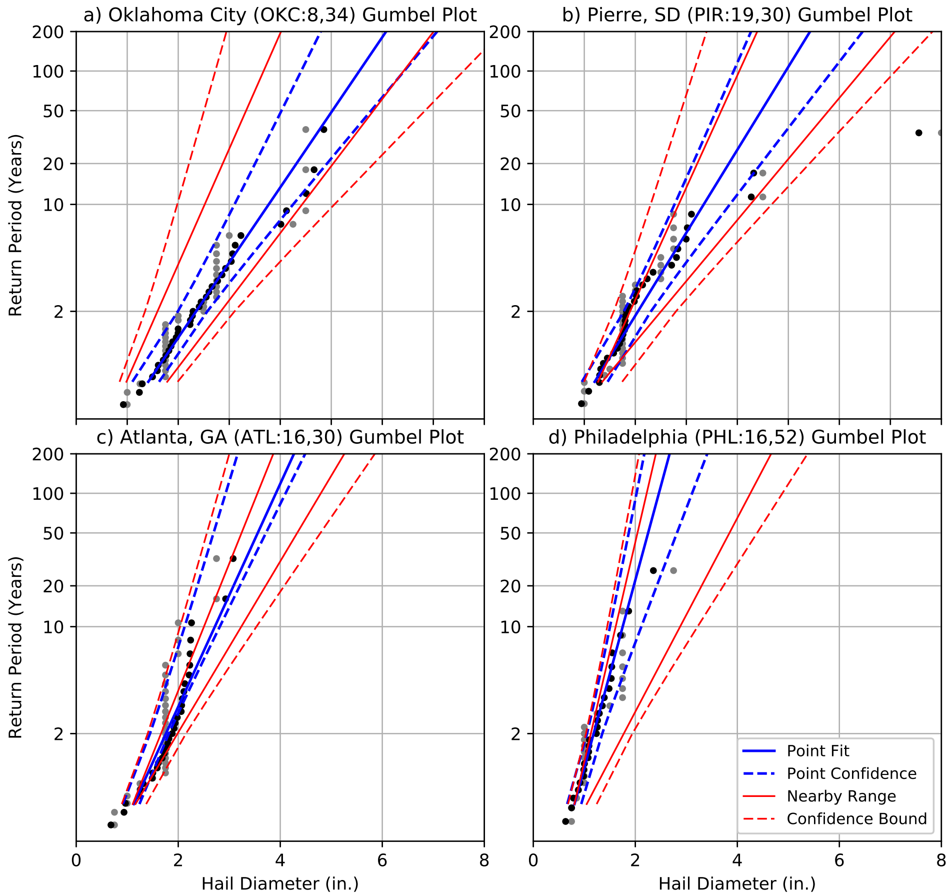
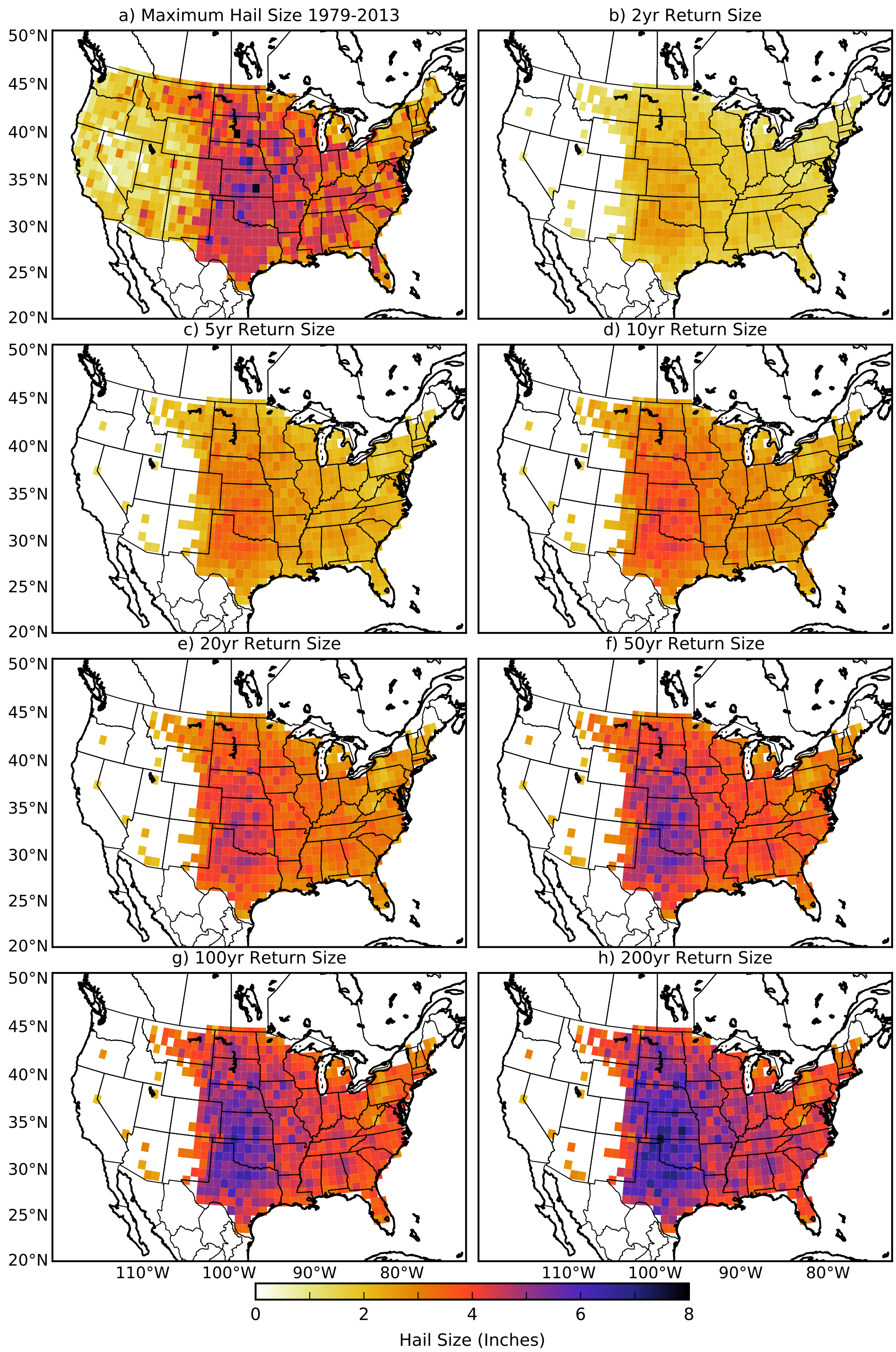
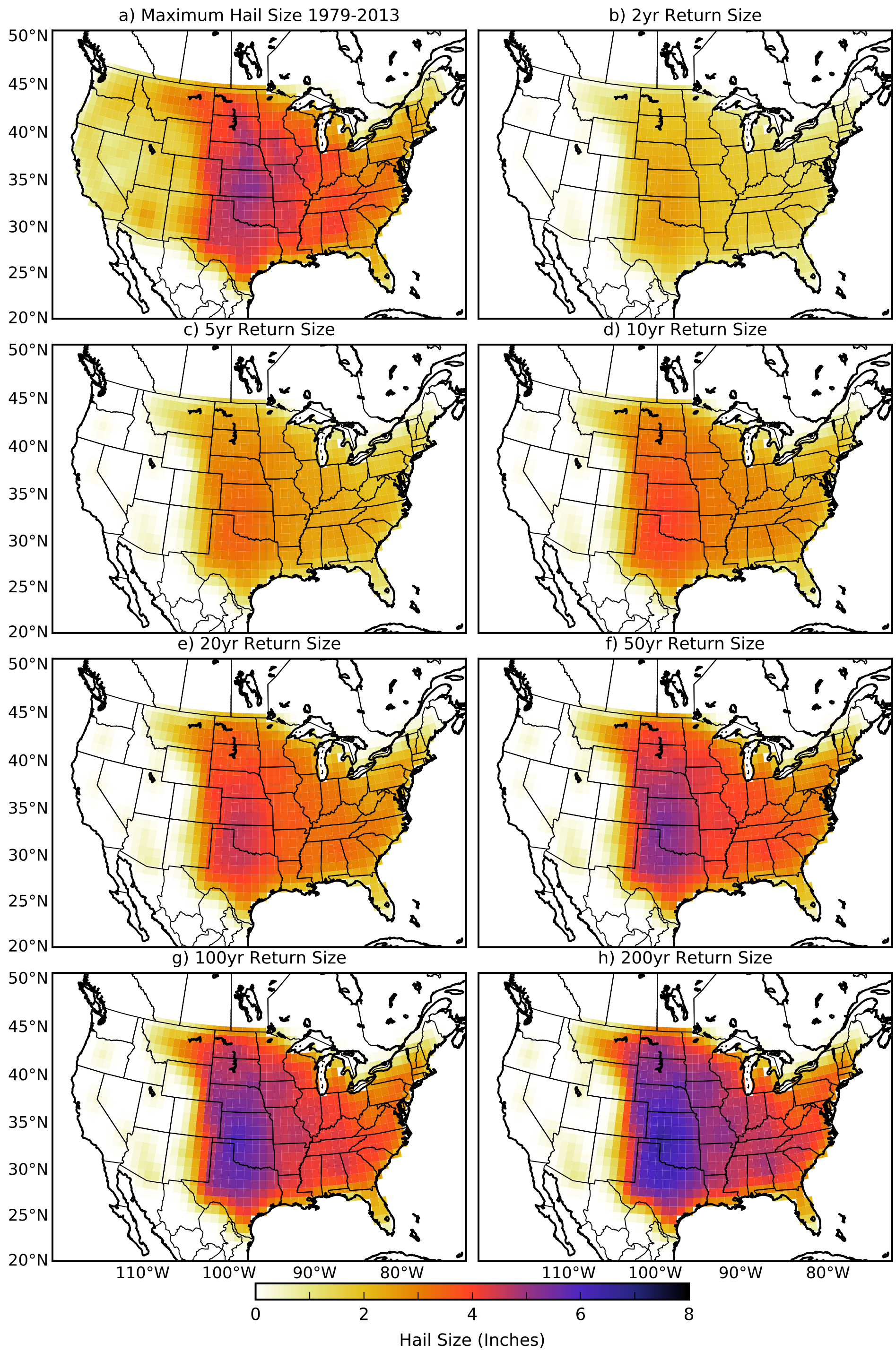


FIG. 7. Evaluation of the maximum expected size of hail at grid points and the nearby region for givens return periods in years as illustrated on a Gumbel plot. Dots represent the raw observations of the point (grey) samples and dithered observations (black). Continuous lines indicate the return curve for the point fitted model on the dithered data (blue), and the range of model fits for the surrounding ±3 grid boxes (48 grid boxes total, in red). Confidence intervals for the point (blue), and the surrounding grid (red) are indicated by the dashed lines. Nearest gridpoints are shown to a) Oklahoma City, Oklahoma, b) Pierre, South Dakota, c) Atlanta, Georgia, and d) Philadelphia, Pennsylvania.



752 FIG. 8. Fitted point dithered Gumbel estimated return hail sizes for the respective quantiles. a) Maximum observed hail size for each grid point 1979-2013. b-h) Modeled return hail sizes at the 2,5,10,20,50,100 and 200 year intervals, for points with at least 30 annual maxima on the 1° x 1° grid.



755 FIG. 9. As for Figure 6, except hail return sizes as determined derived from Gaussian kernel-smoothing of the
756 raw Gumbel return values using a kernel with a 1.00 sigma (1 degree) bandwidth.

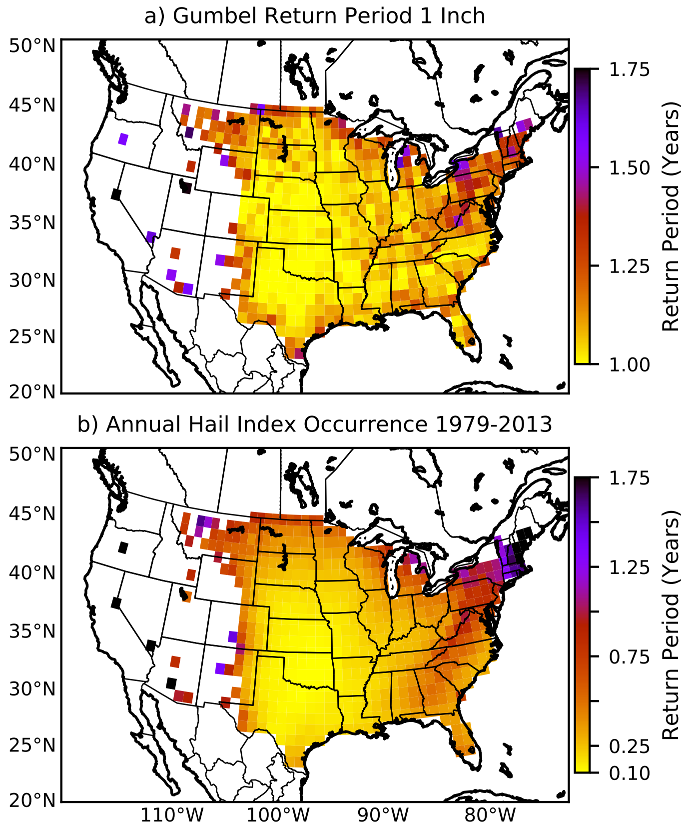
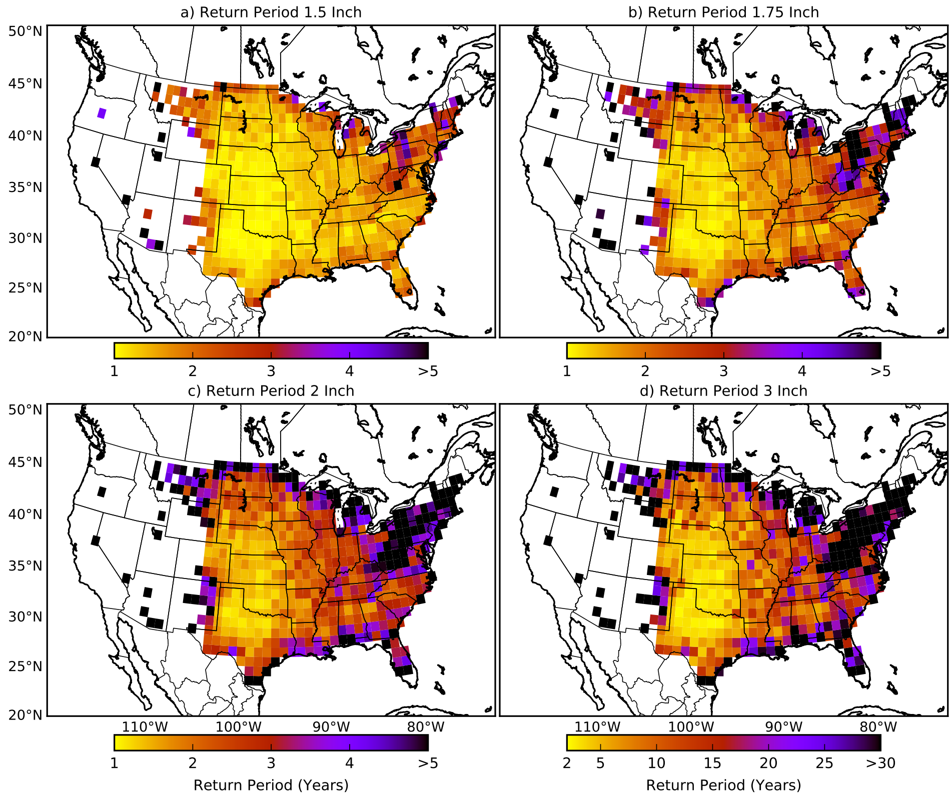


FIG. 10. Comparison of return period over the climatology 1979-2013 from the a) 1° x 1° gridbox model for 25 mm (1 in.) hail with b) Inverse probability calculated using one on the mean annual hail occurrence above 25 mm (1 in.) determined using the hail index derived from North American Regional Reanalysis monthly environmental data (Allen et al. 2015). Note that by construction the minimum value of a Gumbel return period here is 1 year (1/p), whereas the occurrence model is capable of producing intervals of less than 1 year and thus the color scales for the two panels differ.



763 FIG. 11. Fitted Gumbel Return Periods on the 1° x 1° grid for the chosen size thresholds (1.5,1.75,2 and 3 in. respectively). Note that there is a different scale for return period values for panels a),b),c) as compared to panel d) to reflect a longer range of returns for the larger hail sizes.

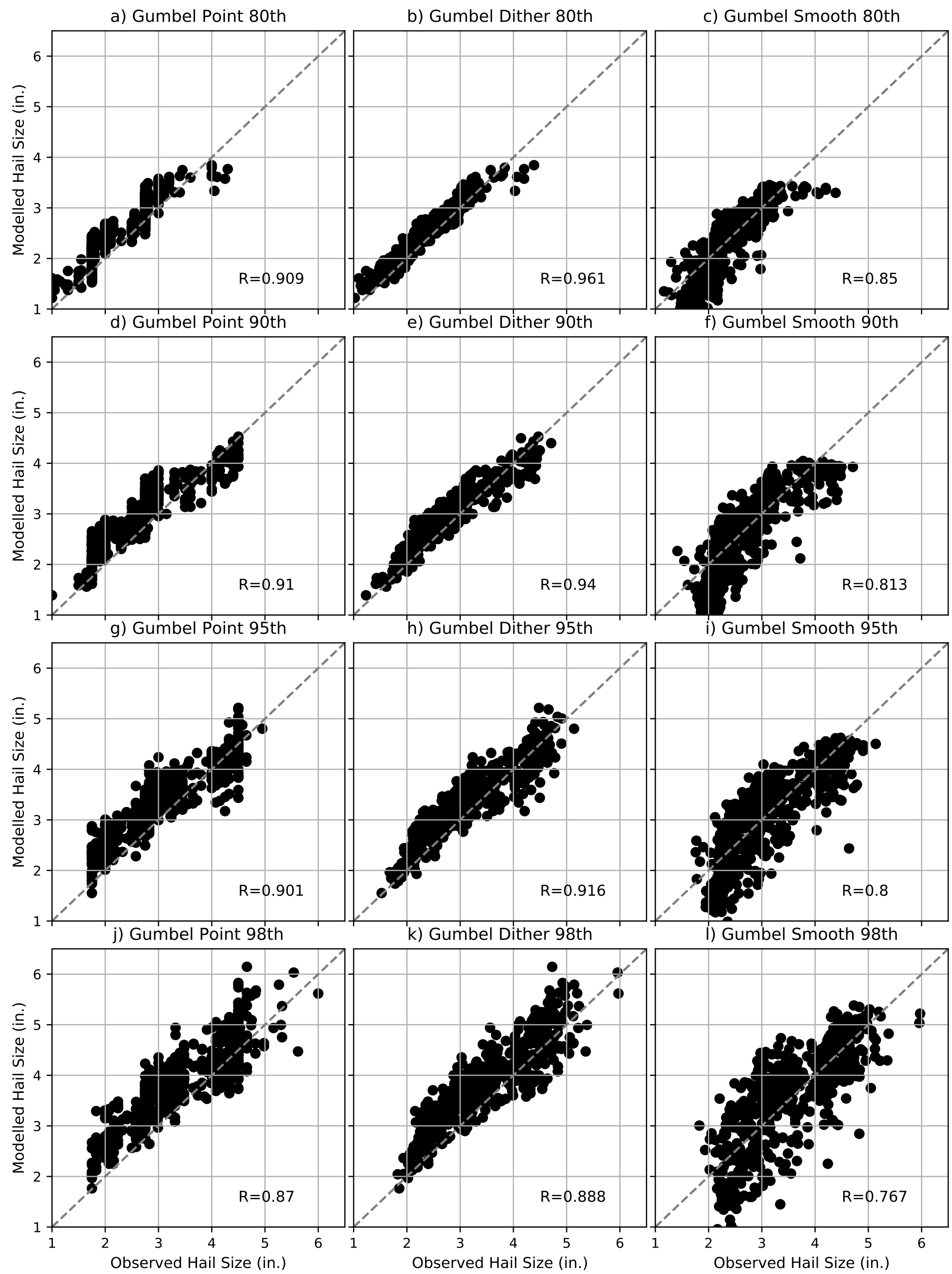


FIG. 12. Comparison of point modeled hail size at the 80th, 90th 95th and 98th percentiles and observed hail size for undithered observations (a,d,g,j), dithered observations (b,e,h,k) and dithered with the smoothed modeled hail size (c,f,i,l), for locations with at least 30 annual maxima observations. Significant Pearson correlations are shown for each comparison.

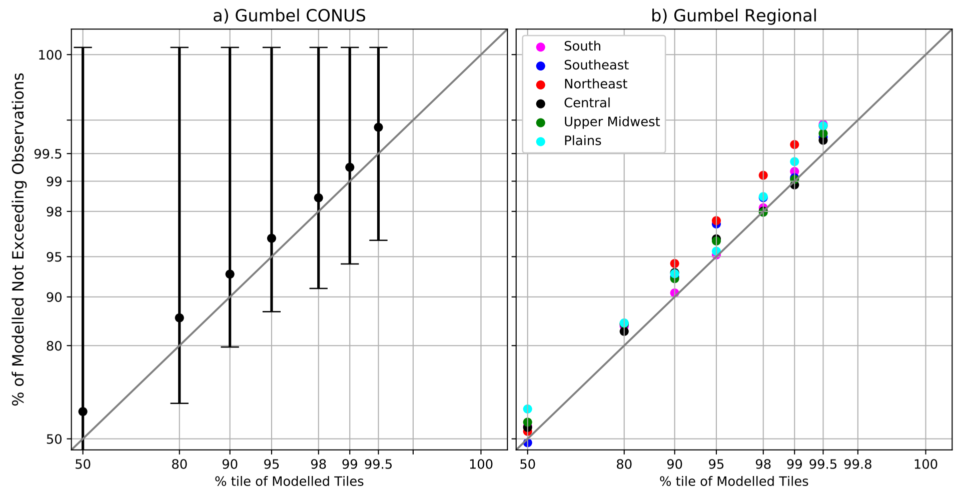


FIG. 13. Percentile of non-exceedence plot for point and areal Gumbel modeled tiles compared to the number of tiles not exceeding the observed values at that grid point. Comparison is made at each of the respective quantiles, with the box and whiskers in a), while b) makes the same comparison over limited NOAA climate regions as defined in Allen et al. (2015b).