The Characteristics of United States Hail Reports: 1955–2014

JOHN T. ALLEN

International Research Institute for Climate and Society, The Earth Institute of Columbia University, Palisades, New York

MICHAEL K. TIPPETT

Columbia University, New York, New York, and
Center of Excellence for Climate Change Research, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

The United States hail observation dataset maintained and updated annually by the Storm Prediction Center is one of the largest currently available worldwide and spans the period 1955–present. Despite its length, climatology of this dataset is nontrivial because of numerous characteristics that are nonmeteorological in origin. Here, the main features and limitations of the dataset are explored, including the implications of an increasing frequency in the time series, approaches to spatial smoothing of observations, and the sources that contribute to the hail dataset. Despite these problems, using limited temporal windows, spatial binning and judicious application of smoothing techniques reveals important characteristics of the hail dataset. The annual and diurnal cycles are found to be sensitive to the spatial shift northwards of observations and increasing report frequency in the Southeast. Hail days, in contrast to hail reports, show no national trend over the last 25 y. Regional and local influences on hail reporting are identified stemming from verification procedures and contributions from local officials. The change in the definition of severe hail size from 0.75 in (1.9 cm) to 1.00 in (2.5 cm) in 2010 has a particularly clear signature in the report statistics. The contribution of storm chasers and source of report factors beyond population to the hail dataset is also explored, and the difficulty in removing these changes discussed. The overall findings highlight the limitations and nonmeteorological features present in hail observations. Adding visual and descriptive metadata has the potential to improve the hail reporting process.

_____

1. Introduction

There are substantial biases in the reported frequency of severe thunderstorm phenomena such as hail, tornadoes and damaging winds (Kelly et al. 1985; Brooks et al. 2003a; Doswell et al. 2005; Doswell 2007; Verbout et al. 2006; Trapp et al. 2006; Johnson and Sudgen 2014; Allen et al. 2015a). The time-varying characteristics of these biases make analysis of the impacts of the climate system on these events especially challenging. Consequently, those unfamiliar with the characteristics of the

_____

Corresponding author address: John T. Allen, IRI, Columbia University LDEO, P.O. Box 1000, 61 Route 9W Palisades, New York, USA
E-mail: JohnTerrAllen@gmail.com

observations may perceive them to be more indicative of changes in the physical phenomena than is warranted. The characteristics of United States (U.S.) hail observations have been quantified relatively poorly for climatological applications in comparison to tornado observations. Investigations of the hail climatology by Stanley Changnon (e.g., Changnon 1977; Changnon 1999; Changnon and Changnon 2000; Changnon 2008; Changnon et al. 2009) and researchers from the National Severe Storms Forecast Center (NSSFC), Storm Prediction Center (SPC) and NSSL (e.g., Kelly et al. 1987; Schaefer et al. 2004; Doswell et al. 2005) have illustrated that many of the characteristics and variability of these reports for the U.S. are associated with societal and nonmeteorological changes.

However, not all changes in the hail climatology can be dismissed as non-meteorological in origin. Recently, environment proxies for hail occurrence have shown regional trends in favorable conditions for hail (Allen et al. 2015a). In contrast, these signals are masked in the observed hail dataset due to changes in reported frequency.

Several authors have discussed the limitations of hail reports and the influences of the forecast warning verification process (Wyatt and Witt 1997; Smith 1999; Schaefer et al 2004; Cintenio et al 2012; Doswell et al. 2005; Paulikas 2014; Allen et al. 2015a). Previously, hail has been stratified into "severe" hail ≥0.75 in (1.9 cm), and "significant" severe hail ≥2 in (5.1 cm) categories (Hales 1993), with the lower category used as the minimum threshold for warning verification. The recent elevation of the severe hail threshold in 2010 to ≥1 in (2.5 cm) has introduced further instability into the dataset related to the verification process, as we will illustrate later. To deal with this change in nomenclature, we will hereafter refer to hail 0.75 ≥ diameter < 1.0 in (1.9 cm) as subsevere, to hail 1.0 ≥ diameter < 2.0 in (2.5 cm) as severe hail, and significant hail when referring to hailstones ≥2 in (5.1 cm) in diameter, or otherwise specify the dimension or range of the longest axis of the hailstone(s).

Nonstationary features (i.e., aspects of the hail climatology that appreciably change over time) are also present on a regional basis. For example, Schaefer et al. (2004) identified a positive trend in hail events in excess of 4 in (10.2 cm) over the Southeast U.S. in the late 1990s, along with other nonstationary behavior in the time series in this region. Cintineo et al. (2012) hypothesized that differences between reported observations and their radar-derived severe hail climatology over the Southeast U.S. were related to a systematic over-reporting as part of the warning verification process. More recently, Allen et al. (2015a) highlighted that the formative environments favorable to severe hail did not indicate the equivalent frequency in this area, further suggesting that the report climatology in this region is difficult to interpret in terms of the environment. Whether this is a reflection of pulse (buoyancy-driven) thunderstorms producing hail events at or above severe thresholds requires further investigation, but this oddity appears not to reflect the expected severe thunderstorm signal. If we consider hail reports as depicted by the SPC, there are regional differences in reporting or collation of reports that vary along county warning area boundaries (P. Marsh, personal communication; Smith 1999). Such features are difficult to isolate as they can arise from local office policy influencing county warning areas (CWAs), contributions from local government officials or alternative sources of reports, and ongoing NWS policy (Weiss et al. 2002). Other nonstationarity may arise from factors external to the NWS. For example, Tuovinen et al. (2009) highlighted the recent impact of social media and mobile internet on the hail reports collected in Finland, and it is likely that a similar signal is present for the U.S. (Hyvärinen and Saltikoff 2010; Blair and Leighton 2012). Factors particular to the U.S. Great Plains also likely contribute to the observation process such as storm chasers and researchers observing and reporting hail from supercell storms during the spring months (Blair et al. 2014; Allen et al. 2015a). Furthermore, recent efforts to gather reports actively have introduced additional inhomogeneities to the hail dataset (Ortega et al. 2009).

Issues encountered in hail data depend on the time period over which the data is collated. An example of this behavior in other severe weather reports is the positive trend in tornado reports that occurs after 1970, contrasting a negative trend in strong tornadoes (Verbout et al. 2006). Severe thunderstorm phenomena tend to occur over small spatial areas. Clustering of observers toward areas of higher population can introduce a greater degree of regional variability, depending on where the conditions favorable to development occur in a given year. Analyses of trends in hail occurrence have shown these regionally varying characteristics to be associated with increased observer density (Schaefer et al. 2004; Brooks and Dotzek 2007; Doswell et al. 2005; Allen et al. 2015a). Areas with an increasing frequency of reports are present where population is large or growing, along road networks and where storm chasers often frequent. East of the Rocky Mountains, Brooks and Dotzek (2007) found strong variability in the number of reports of hail in excess of 7 cm, but no clear trend in the proxy severe-thunderstorm environment frequency from the past 50 y. More recently, Allen et al. (2015a) identified a tenfold increase in severe hail reports for the U.S. (1979–2012) that appears to have little meteorological (favorable environment-driven) reasoning, and results from changes to the total number of reports (Tippett et al. 2015).

The climatology of U.S. hail has reached 60 y in length (1955–2014), and given the substantial changes in the past two decades, it is prudent to outline known limitations. The purpose of this manuscript is to inform and illustrate to the research community the limitations of U.S. hail reports for applications to a variety of problems including climatology, satellite- and radar-derived product verification, insurance-portfolio-loss estimation and climate linkages. In doing so, we intend to inform future users of the underlying characteristics, rather than dissuade them from using this valuable dataset.

Regardless of its limitations, the U.S. hail dataset is one of the most complete currently available in the world (Tippett et al. 2015). Similar discussion of dataset limitations has been made for the tornado record (Brooks et al. 2003a; Verbout et al. 2006), and has not limited the application of the tornado record to analysis of climatological trends and questions of increasing variability (Brooks et al. 2014; Elsner 2014, Tippett 2014; Coleman and Dixon 2014). Comparatively little attention has been given to the characteristics of the hail dataset, as the nonmeteorological characteristics are more substantial than those of the tornado dataset, as we further illustrate here.

We do not claim to provide a comprehensive description of all nonmeteorological characteristics the hail dataset. We also do not intend to question the reports or source of any one CWA over any other, but only to illustrate that the challenges of the hail report data extend beyond population or easily modeled corrections that may be possible for tornadoes (Widen et al. 2013; Elsner et al. 2013). Characteristics on a local or county scale resulting from local office policy, observation sources and individual forecasters likely exist beyond what is detailed here, and these nuances may only be in the knowledge of local forecast offices or individuals.

The paper is structured as follows: in section 2, we detail the sources of the hail data described here. In section 3 we examine the annual and diurnal cycles of the climatology to assess the degree to which they are influenced by spatial and temporal changes. In section 4, the temporal characteristics of the dataset are investigated via time series along with the size distribution of reports. In section 5, we examine the spatial variations in hail reports and the contributions of local factors and changes over time to these

characteristics. The influence of the shift in minimum severe hail size in 2010 is then analyzed in section 6. In section 7, we explore the contribution of storm chasers to the regional frequency of hail reports over the Texas Panhandle, and illustrate why population density alone is not an appropriate function to smooth hail observational data. We also investigate how the source of hail observations has changed over the period 1998–2014, and the spatial variations in the fraction of reports by originating source of the report. Finally we outline the potential uses for the current hail dataset, steps that can be taken to improve the inputs to this record, and offer suggestions that we believe will enhance the ability of researchers to interpret changes in hail observations for a variety of applications.

2. Data and methods

The SPC Severe Weather Database (SWD, available at http://www.spc.noaa.gov/wcm/) is the primary source of U.S. severe weather occurrence information, and reports of hail are updated on a yearly basis with data provided by observers to local NWS offices (Schaefer and Edwards 1999). Features of the dataset can be categorized as stationary (little change through time), and nonstationary (changing through time). Report data are dependent on observer availability, and therefore have nonstationary features that are difficult to characterize. We only consider hail exceeding a diameter of 0.75 in (1.9 cm); only 13 reports are below this value. Corrections also were made for incorrect date information in four cases (mis-entered month) and data excluded if portions of the entry were missing. Swath information for hail was not used, as it was only available for 28.9% of observations, and in most cases repeated a single point source. Hence latitude and longitude data were taken from the beginning point for those data points that included an extended path of hail fall. The end result is a set of 266 282 hail observations, of which 201 296 (75.6%) occur in the last 20 y of the dataset between 1995–2014.

For temporal analysis of the seasonal cycle, we derive a histogram of the mean number of observations on each calendar day, and use a one-dimensional (1D) Gaussian kernel with $\sigma = 15$ days to smooth the frequency. This bandwidth is in line with smoothing applied in previous examinations of the annual tornado cycle (e.g., Doswell 2007). For the diurnal cycle, we instead apply a 1D Gaussian kernel

with σ = 2.0 h to the histogram given the smaller range of bins. Periodicity is also accounted for in each case by fitting the kernel across three duplicated instances of the binned data.

For spatial results, we first project the latitude-longitude of the observations to an axial equidistant areal projection over the CONUS. Reports of hail are then gridded analogously to Brooks et al. (2003a) and Doswell et al. (2005), to density using an 80 × 80 km grid. Following the gridding procedure, we then apply a

2-dimensional equidistant Gaussian kernel with σ = 1.5 × grid spacing (approximately equivalent to 120 km and SPC outlooks). Sensitivity to bandwidth was tested to ensure data were not oversmoothed. The procedural specification varies from the daily probabilistic likelihood of Brooks et al. (2003a) and Doswell et al. (2005); however, we feel that this fairly demonstrates the distribution of reports equivalent to prior results, and respects the discussion of kernel and bandwidth choice by Marsh and Brooks (2012).
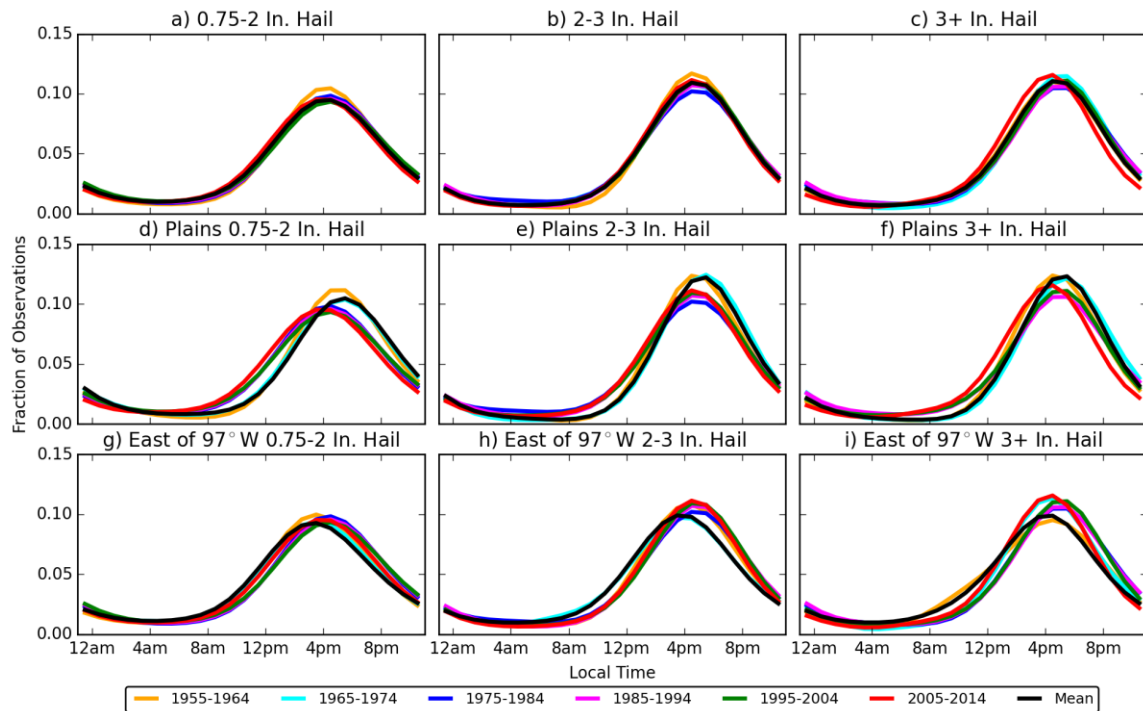

Figure 1: Diurnal cycle of hail reports in local time 1955–2014 stratified by hail size and decade, along with the overall fraction for selected regions. a),b),c) All US hail reports; d),e),f) hail reports restricted to a region in the northern Plains bound by 107°–100°W and latitude north of 36°N; and g),h),i) hail reports restricted to the area east of 97°W. a),d),g) ≥0.75 and <2.00 in (≥1.9 and <5.1 cm); b),e),h) ≥2.00 and <3.00 in (≥5.1 and <7.6 cm); and c),f),i) ≥3.00 in (≥7.6 cm). *Click image to enlarge.*

To address issues with clustering of reports on single days, we also calculate the number of days with hail (hail days) by filtering to a maximum of one per day within the predefined grid. In assessing the source of the reported hail, we also consider the full archival NCDC report data. This information is not included with the SPC data and only used here for exploration of the source information. The source field is available for the recent period (1998–2014). The total number of reports by source is collated and the gridding of these data is also to the 80 × 80

km gridded domain on an axial equidistant areal projection. No smoothing is applied to maintain the spatial fidelity of the fractional source information. Fractions are calculated based on the total number of reports from all sources for each year, and as a fraction of the national total of that report source. Further information and details regarding the sources of hail and other NCDC severe weather observations can be found at: ftp://ftp.ncdc.noaa.gov/pub/data/swdi/stormev ents/csvfiles/legacy/.

To derive maps of population density, county intercensal estimated population data on a yearly basis was retrieved from http://www.nber.org/data/census-intercensal-county-population.html, and higher resolution data sourced for 1995 and 2000 from the Center for International Earth Science Information Network (CIESIN) unadjusted high resolution gridded population density dataset (available at http://sedac.ciesin.columbia.edu/data/collection/usgrid/sets/browse). To produce the county-based density choropleths, we determine the area of respective counties in km^2 and use this to determine population density. As population is not uniformly stratified, for both the county and gridded data we apply a Jenks Natural Breaks method (Jenks 1967) typically used by geographers to identify the appropriate levels representative of population density at the chosen years of interest.

3. Diurnal and annual characteristics

a. Diurnal cycle

The largest difference in the relative diurnal distribution stratified by hail size occurs during the late afternoon hours, where subsevere and severe hail is typically recorded earlier than significant hail, and with a wider distribution throughout the day (Fig. 1a,b,c). We can explain this distributional shift for larger stones by considering that the large majority of significant hailstones are associated with supercells (Blair et al. 2011; Blair et al. 2014) and found in the Great Plains in higher proportions as a result (comparing Fig. 1b,e). This observed diurnal difference is plausible given the strong capping inversions that persist later in the diurnal cycle, when larger hailstone sizes are supported by the peak available CAPE. An alternate and complementary hypothesis is that storm-relative airflow patterns are the driving factor behind the largest hailstones (Nelson 1983; Ziegler et al. 1983; Nelson 1987) and thus more dependent on the development of supercell morphology that allows favorable growth trajectories. While this tendency for later afternoon occurrence of larger hail sizes is also present for the Southeast (Fig. 1g,h,i), the signal in the Plains shows a much greater fraction of events occurring in the early evening hours.

Another difference between the diurnal cycles of different hail sizes is found in the post daylight hours (i.e., after 2000 LT), where the low-level-jet strengthening coincides with the nocturnal decoupling of the boundary layer and elevated instability and favorable vertical wind shear persists into the morning hours (Mead and Thompson 2011). This activity often produces upscale growth (Kumjian et al. 2006) or elevated supercells that contribute to the small proportion of severe and significant hailstones at night.

Despite these results however, considerable variability remains between the decades of the climatology in diurnal cycle, particularly for subregions such as the Plains or over the East, with the shifts over time resulting in an apparently unchanged U.S. wide distribution for hail <3 in (<7.6 cm), and a slightly earlier peak for hail ≥3 in (≥7.6 cm). For the Plains, the shift suggests the diurnal peak is earlier in the day for all sizes in the most recent decades, with the overall number of reports from 0800–2000 LT increasing remarkably (7–11%). In contrast, for the east the distributional peak has shifted later, with additional reports recorded in the evening. These regional variations and nonstationarity suggest the need for careful examination of the diurnal cycle of hail from report data.

b. Annual cycle

The annual cycle of mean monthly hail occurrence has been discussed, both in terms of the monthly frequency and spatial pattern for the period 1979–2012 (Allen et al. 2015a). Here we consider the annual cycle as represented by the mean number of hail reports, days with hail for the period 1955–2014, along with subsets obtained using significant hail and spatial restrictions (Fig. 2). The smoothed Gaussian fit is chosen so as to explore the seasonal peak, though other kernel bandwidths or approaches may be appropriate for shorter temporal or regional features. Considering the CONUS-wide signal in subsevere and significant severe reports, the data clusters towards the beginning to middle of the year. The peak of the distribution occurs at the end of May and early June, with little decadal variation in the cycle over the length of the record (Fig. 2a,c).

The peak of subsevere hail days occurs from mid-May through the end of July, with a gradual decline (Fig. 2b). Unlike subsevere hail reports, there is considerable variability between the respective decades for subsevere hail days, with a broader seasonal peak in recent decades. Using significant hail days (a proxy for supercell thunderstorms), the peak shifts to the end of May and beginning of June. Spring favors a rapid

increase in the frequency of significant hail events compared to a slower tapering for the summer and autumnal months. The decadal variations are less evident in significant hail with a slight extension to the season in the summer months, suggesting that its use may be a better detection mechanism for variations in the annual cycle. Computing separate annual cycles for the northern Plains (Fig. 2e) and the east (Fig. 2f), the origins of this wider peak for subsevere hail are over the east rather than the central or northern Plains, as the later peak signal is evident for that region. The shifts in significant hail apparently arise from a progressively later shift in the annual peak of northern Plains hail (Fig. 2e).

These elements reveal three important features that must be considered when analyzing the annual cycle: the use of hail days is important in revealing the actual structure of hail frequency, regional choices can influence the picture obtained of the annual cycle, and the choice of size threshold can reveal very large differences in the resultant peak. These limitations exist in addition to the influence of changes to the annual cycle over the length of the dataset. Thus care must be taken to ensure that pooled data for the annual cycle is sufficient, and that variability is mitigated.
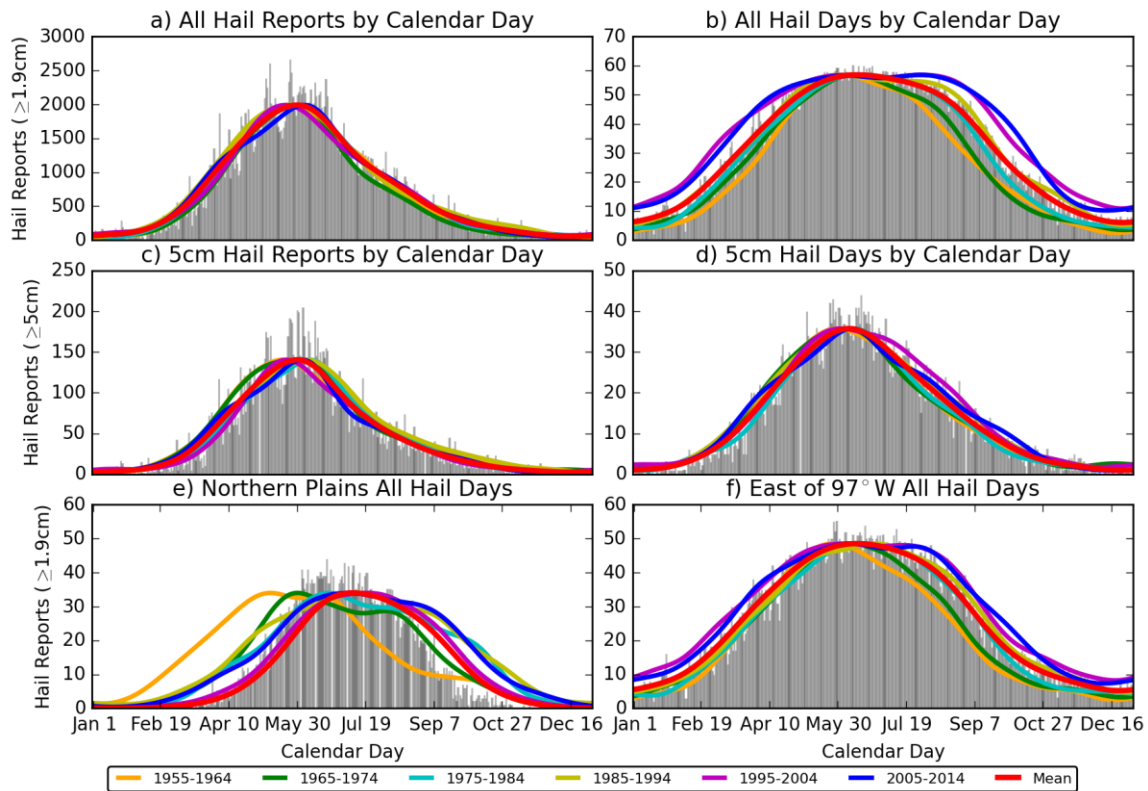

Figure 2: The annual cycle of hail reports and days for the U.S. 1955–2014: a) frequency of hail reports ≥0.75 in (1.9 cm) by calendar day, with the 30-day Gaussian smoothed annual cycle illustrated by the red line; b) as for (a) except frequency of hail days ≥0.75 in (1.9 cm) by calendar day; c) as for (a) except hail reports ≥ 2 in (5.1 cm); d) as for (b) except hail days ≥2 in (5.1 cm); e) as for (b) except hail days restricted to a region in the northern Plains bound by 107°–100°W and latitude north of 36°N; and f) as for (b) except hail days restricted to the area east of 97°W. *Click image to enlarge.*

4. Interannual variability

Over the full record, the reported frequency of hail displays nonstationary behavior across all sizes, with the largest changes occurring after the 1990s for hail sizes ≤2 in (≤5.1 cm) and

displaying trends where statistically significant increases are identified, using a two-sided p-value test whose null hypothesis is that the slope is zero. This positive trend in the number of reports is common to all aspects of severe-thunderstorm reports, though not always for the

more severe magnitudes (Doswell et al. 2005, Verbout et al. 2006). In the hail climatology, we note several breakpoints. For all hail reports, local increases appear to be associated with the establishment of the NSSFC in 1966 (examined
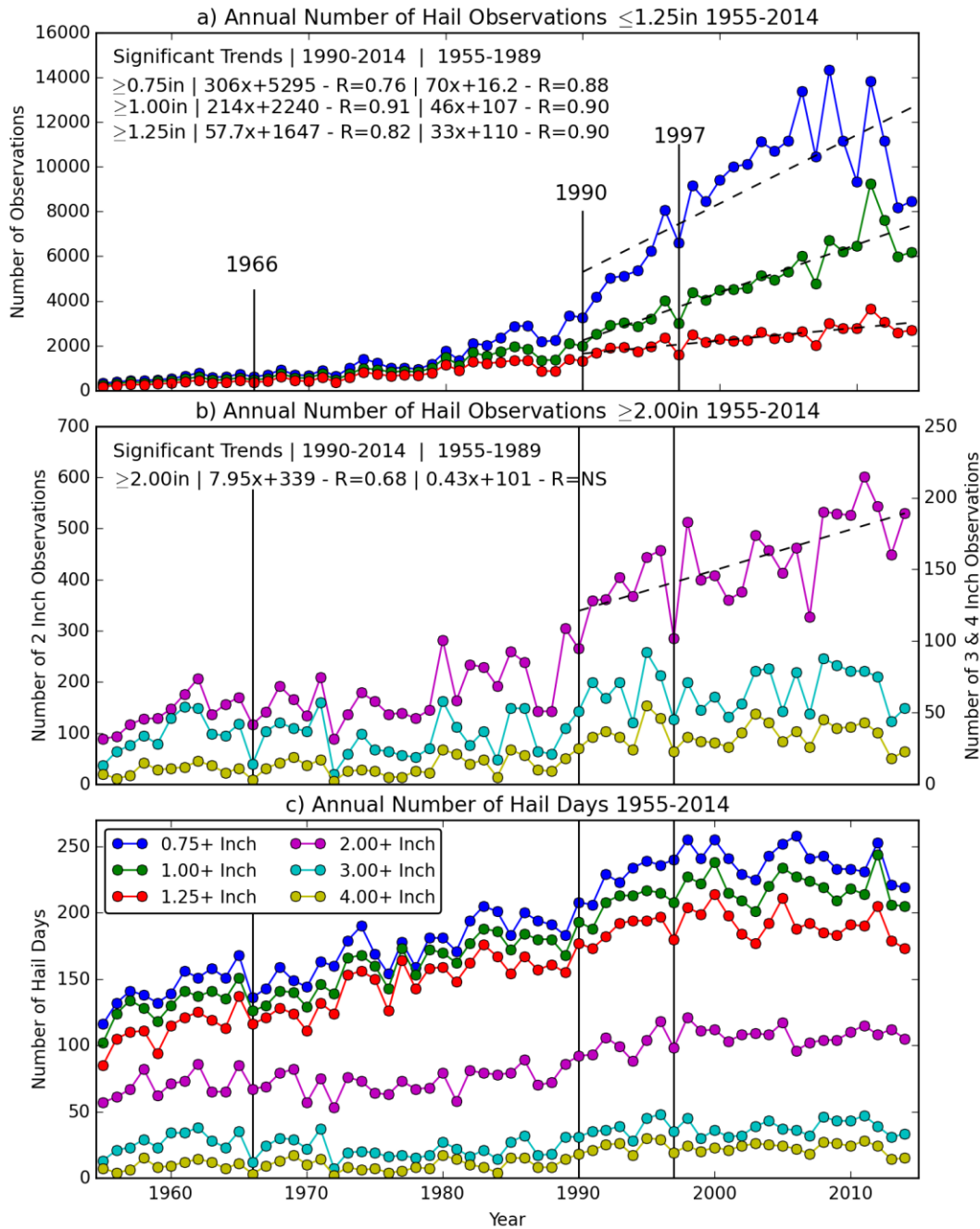

Figure 3: a) and b) Time series of hail reports (1955–2014) over the entire U.S. stratified by minimum size threshold to illustrate the interannual variability and increasing frequency of hail reports. c) Number of days with at least one hail report above the threshold size. Only statistically significant trends (two-sided p-value test p<0.01 with null hypothesis of zero slope) in hail reports and hail days over the period 1990–2014 are shown, with their respective coefficients and correlations, and corresponding trend values for the period 1955–1989 to illustrate break points. Vertical lines denote 1966, 1990 and 1997 mentioned in text. *Click image to enlarge.*
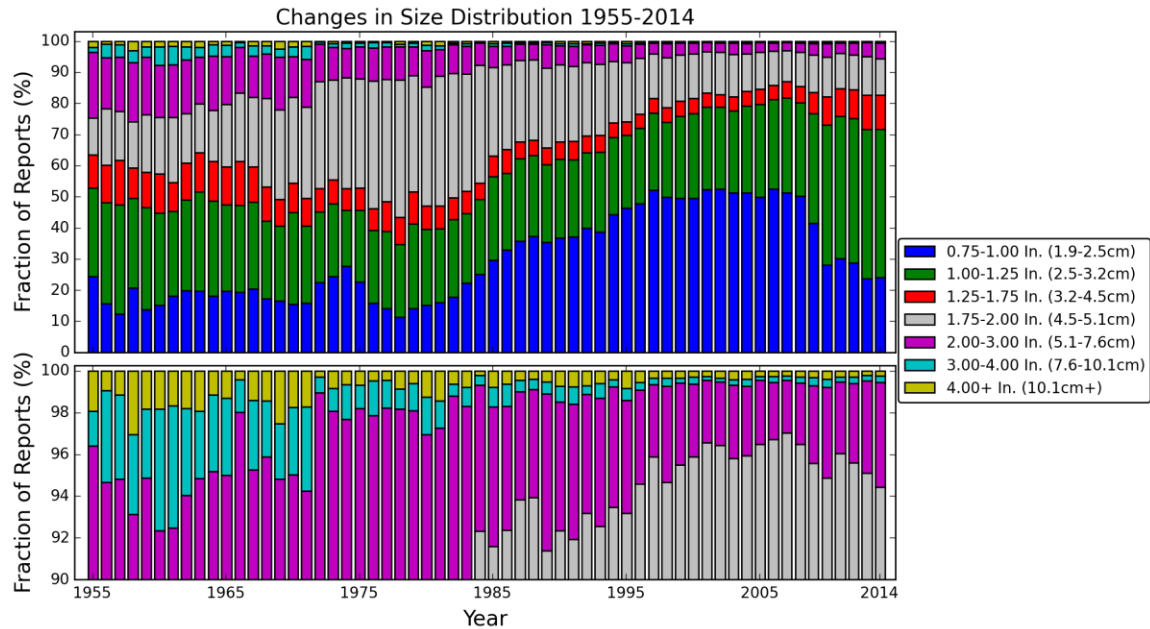
7

Figure 4: Fraction of total hail reports per year for 1955–2014 within specified size groupings indicated by the colorings shown in the legend. The lower panel is a zoomed view of the 90–100% fraction from the top panel to illustrate the contributions of the largest stones to the record. *Click image to enlarge.*

spatially later) and the formation of the SPC in 1995 (Fig. 3). The active maintenance of the reports database and the establishment of a network of spotters seem to contribute to the relatively steady increase until the early 1990s. The second, more major discontinuity begins in the early 1990s with the rollout of the WSR-88D radar network (Crum and Alberty 1993). This change is associated with an apparent rapid increase in trend for all hail sizes ≤2 in (≤5.1 cm) continuing throughout the rollout period. A likely explanation for this change is the contribution to the increased interest in severe storm warning forecast verifications by the NWS based on the improved radar coverage, similar to the change identified in F0 tornado detection in the SWD tornado dataset (Verbout et al. 2006).

The progressive rollout introduces regional variations in when this increasing trend begins, and thus few points in the U.S. can be considered to have a stationary frequency of reports prior to 1997. A possible explanation for the trend in significant hailstones leading up to 2000 is that more recent reports may reflect an increasing proportion of observers located within supercell swaths, such as storm chasers and research field projects (Blair et al. 2014), in addition to the Doppler radar rollout influence. The magnitude of these trends suggests that care should be taken

when restricting report data to near reporting thresholds, and perhaps avoided entirely. Using hail days at the respective thresholds (Fig. 3c) appears to alleviate any robust trends, and suggests a relatively reliable hail record is possible from the early 1990s onwards. Interestingly, the increases to 3-in (7.6-cm) and 4-in (10.1-cm) hail are not uniformly in line with the increases to severe and significant hail, and remain relatively unchanged since 1990 in terms of both hail reports and days. This difference may be related to the relatively restricted area over the Plains states where these hail sizes are commonly found, and potentially related to hail report collection by the NSSFC first in Kansas City, MO, and later by the SPC in Norman, OK.

Another characteristic of the time series of reports is the fraction of reported severe hail in a given size range (Fig. 4). As the record moved into the mid-1980s, the relative fraction of the largest hail sizes (≥1.75 in or 4.5 cm) began to decrease, while the fraction of both severe and 1.25–1.75 in (3.1–4.5 cm) hail remained consistent, reflecting the positive trend in reports. Given the change of severe diameter in 2010, the overall number of hail reports would be expected to decrease as subsevere reports potentially were discarded. However, a large jump in the relative fraction of severe hail
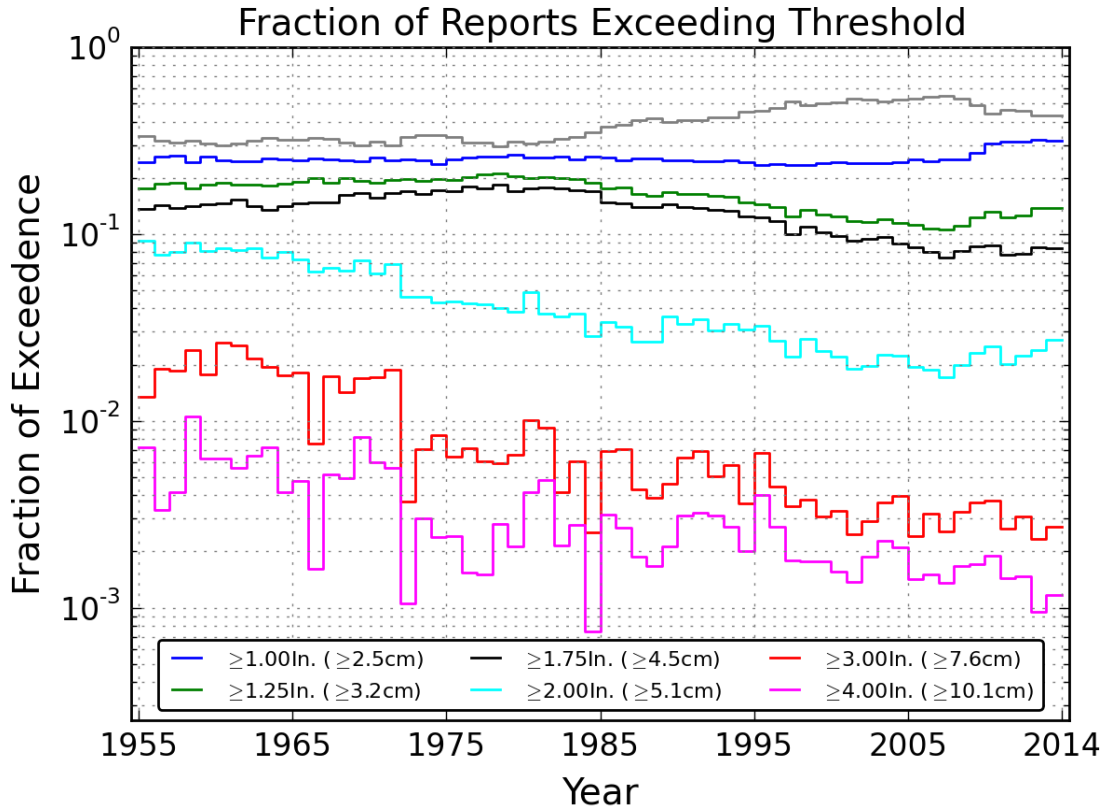
Fraction of Reports Exceeding Threshold

Figure 5: As for Fig. 4, except the cumulative fraction of hail reports stratified by minimum size, on a logarithmic y-axis to emphasize the shifts in the fraction of larger hailstones despite the trend in smaller stones. The grey line reflects the fraction of hailstones meeting the subsevere threshold. *Click image to enlarge.*

reports (≈20%) occurs, and there is also an increase in the fraction of 1.25–1.75 in (3.1–4.5 cm) reports. This jump coincident with the change of severe diameter is concerning, as it is possible that hail sizes are being inflated or quantized to meet elevated severe-hail warning criteria, and suggests a potential systematic bias that we explore further in section 6.

We also can consider the cumulative distribution of hail reports by size (Fig. 5). The number of hail reports greater than 3 in (≥7.6 cm) as a percentage of the total number of hail reports has steadily declined over the entire record. Reports of hail greater than 2 in (≥5.1 cm) show similar behavior but with an increase in relative frequency over the most recent decade, which is also present in reports from 1.25–1.75 in (3.1–4.5 cm).

Reports of hailstones ≥1.0 in occupy a nearly constant fraction of the total reports over most of the record. However, there is a notable jump in

2010 in the relative frequency of reports of hail greater than severe, which coincides with the change of severe threshold.

The relatively unchanging time series of significant hail suggests that taking into account natural variability, this subset may be potentially more consistent through the last two decades than using lower thresholds. Despite this, considering stratifying to avoid changes in the time series is not necessarily sufficient to resolve nonstationarity in the hail record.

5. Spatial changes

a. *Evolution of point reports through time*

We also can visualize changes through time by showing alterations to spatial patterns of point hail reports and smoothed isopleths of hail report frequency (Fig. 6). The data are broken into the six decades of the hail record with a color scale normalized to the final decade to illustrate how

these changes have taken place, while its companion figure illustrates how relative density has changed through time (Fig. 7). During the early years (1955–1965), reports are mostly found east of the Rocky Mountains, with small concentrations towards population centers and somewhat more uniform reports across Oklahoma. By 1965–1974, there is greater clustering towards population centers, particularly in the Plains states around Kansas

City, Oklahoma City and Wichita, potentially reflecting the influence of the establishment of the NSSFC in 1966. Increases continue steadily into the next decade, with larger population centers beginning to show increases in report frequency, such as Dallas-Fort Worth, San Antonio and Saint Louis areas, along with increasing numbers of reports from the northern Plains.
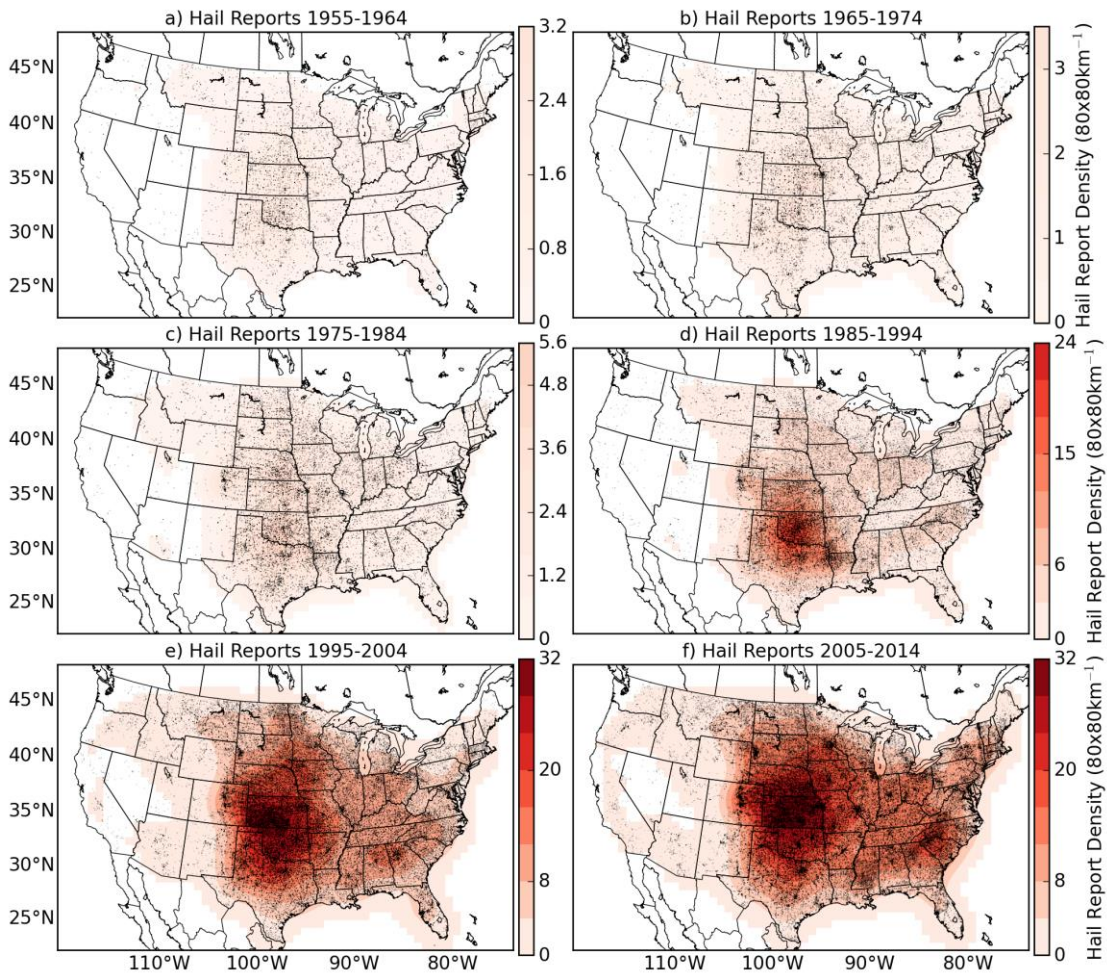

Figure 6: Mean annual Gaussian kernel-smoothed subsevere ($\geq$ 0.75 in or 1.9 cm) hail-report density for decade intervals for 1955–2014. Overlaid are point reports of hail diameter for the corresponding decades, illustrating the growth in report spatial frequency through time. Density contours are scaled by the peak density of the 2005–2014 period, such that the color scales are equivalent to the 0–32-report density per 80×80 km^{-1} range used in panel (f). In each case, the peak value of the color scale indicates the peak report density. Report latitude and longitude over the continental U.S. are projected onto an axial equidistant areal projection during the plotting procedure. *Click image to enlarge.*
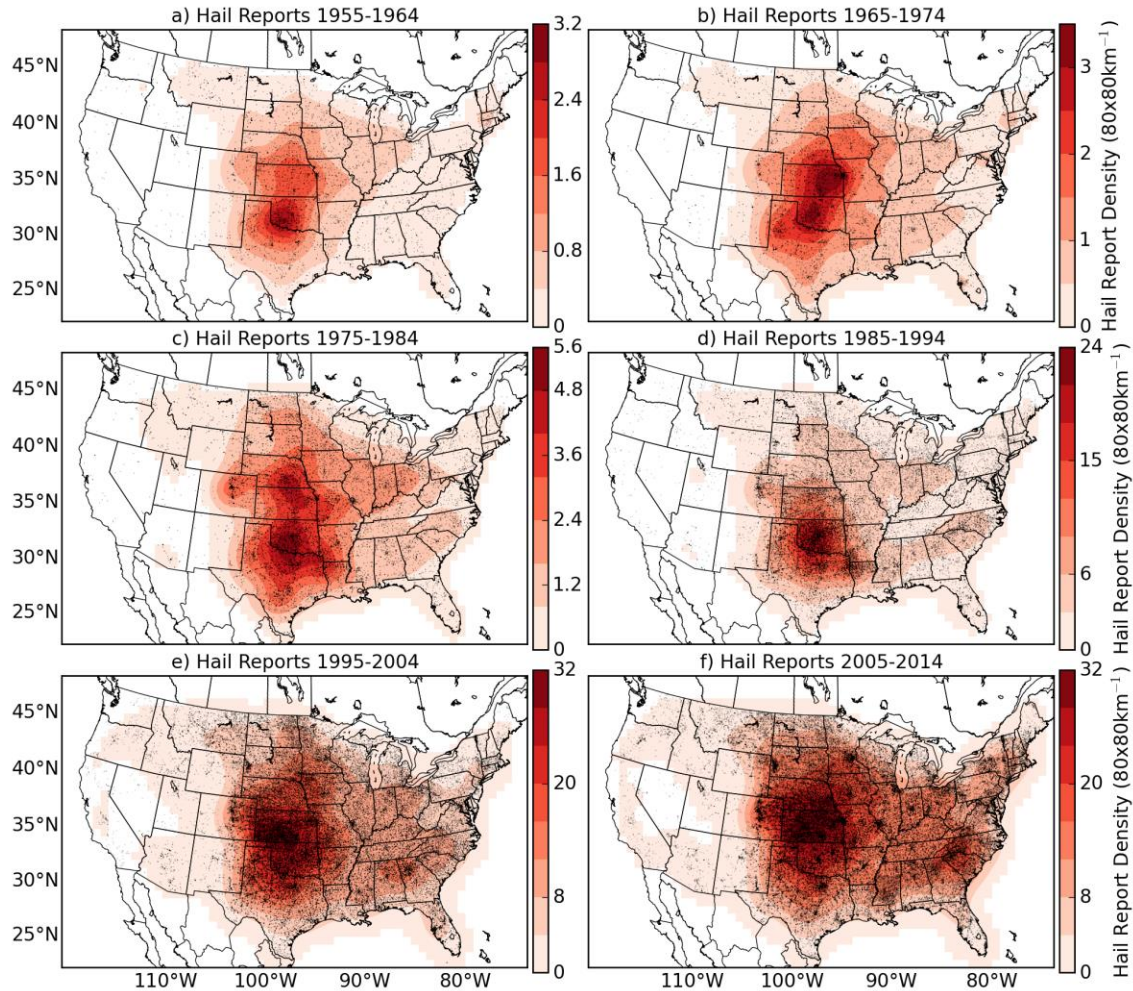
Figure 7: As for Fig. 6, except the hail-report-density color scale is relative to the peak density in the respective decades, as opposed to scaled by the final decade. *Click image to enlarge.*

During the period 1985–1994, a rapid growth in report frequency begins to saturate the outline of the state of Oklahoma, and to a lesser extent Kansas, while reports for the metropolitan areas of Denver, Oklahoma City and Dallas-Fort Worth continue to increase. A particular oddity can be identified in the forecast area surrounding the Shreveport, LA NWS office, which sees an increased frequency from the prior decade to a level equivalent to Oklahoma, followed by a rapid decline to much lower levels in the following two decades. The final two decades also see decreasing frequency over the state of Oklahoma, while densities over metropolitan regions continue to increase and spread from the Plains states. Over the final decade, we see that reports become more common over the northern Plains, the east and the northeast. There is also a northward shift in report frequency from

Oklahoma and northern Texas to Kansas and Nebraska.

An important difference illustrated by the most recent decade is that the hail report density is nonzero over most of the U.S., excluding the lightly populated regions of Nevada and Utah. Report density is clustered towards the region east of the Rockies, with bands stretching further east both north and south of the Appalachian mountain chain.

As the population within the mountains is relatively sparse, and road networks less dense, the density of hail reports over this region should be similar to the areas both north and south, as environmental conditions taken from the North American Regional Reanalysis (NARR) do not vary markedly (Allen et al. 2015a). However,

11

this difference is difficult to discern given NARR's noted limitations in resolving initiation and reliable convective environments around

topography. Despite the overall smaller number of reports, a similar pattern is noticeable for the east (*e.g.*, Atlanta, Indianapolis).
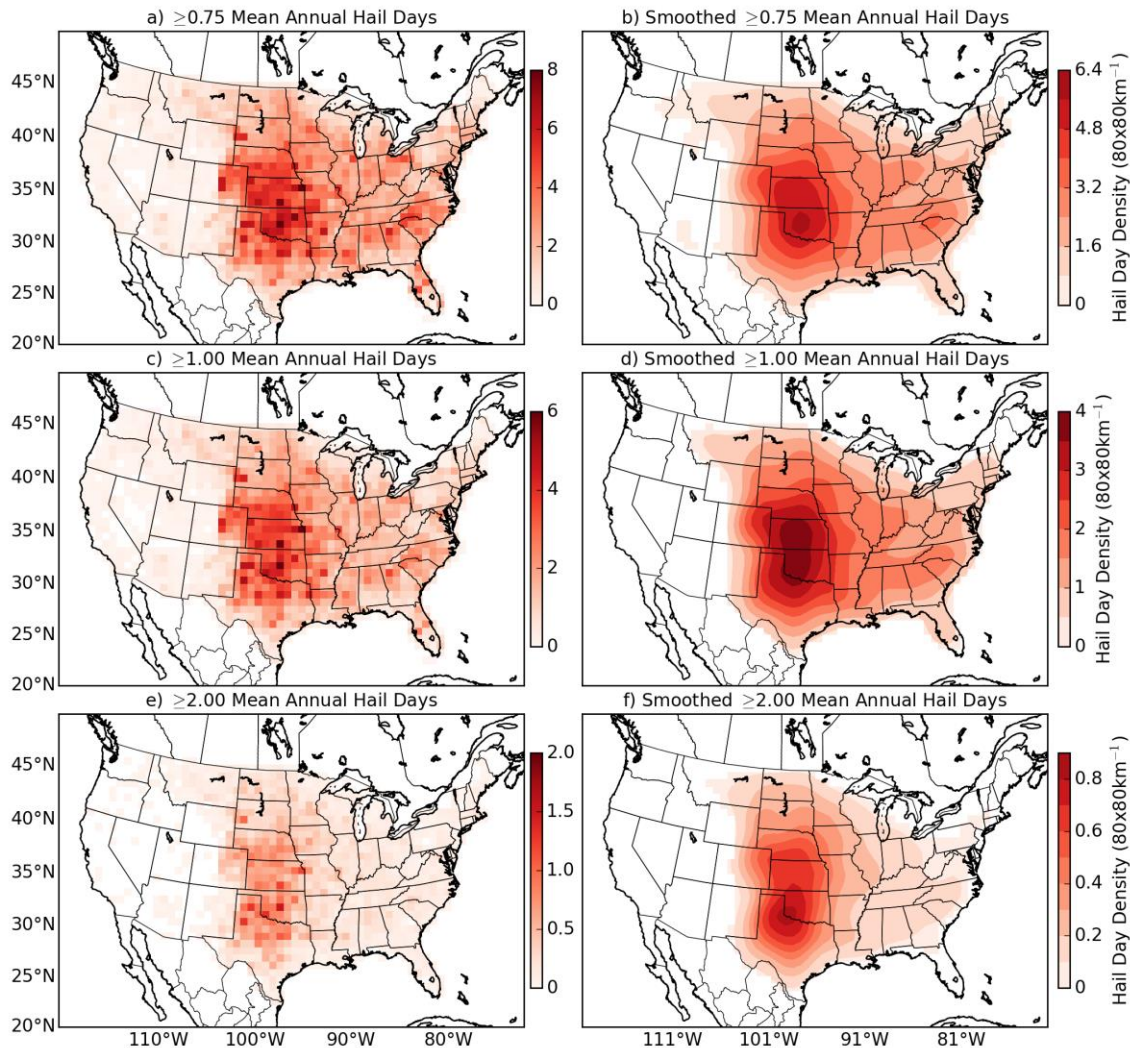

Figure 8: Mean annual frequency of gridded and Gaussian kernel-smoothed hail-day density for hail diameter equal to or greater than given thresholds for the length of the hail record (1955–2014): a) gridded hail-day density ≥0.75 in (1.9 cm); b) smoothed hail-day density ≥0.75 in (1.9 cm); c) as for (a) except hail days ≥1 in (2.5 cm); d) as for (b) except hail days ≥1 in (2.5 cm); e) as for (a) except hail days ≥2 in (5.1 cm); f) as for (b) except hail days ≥2 in (5.1 cm). *Click image to enlarge.*

If we consider the significant hail reports over the same period (not shown, but discussed further in section 5b), the frequency of point reports is more confined to the Plains, where reporting has changed less over time. There is also considerably less change in the spatial extent of where significant hail is reported than the subsevere reports category. The exception to this is a small increase in the last two decades over the Southeast, which is consistent with the

previously discussed findings of Schaefer et al. (2004). While metropolitan areas are slowly becoming more pronounced in the most recent decade for these reports, at least spatially, this dataset appears less prone to regional biases than the subsevere and severe size data. Nevertheless, the frequency of large hail reports likely is underrepresented due to the negative diameter bias in the SPC dataset, as illustrated by Blair et al. (2014).
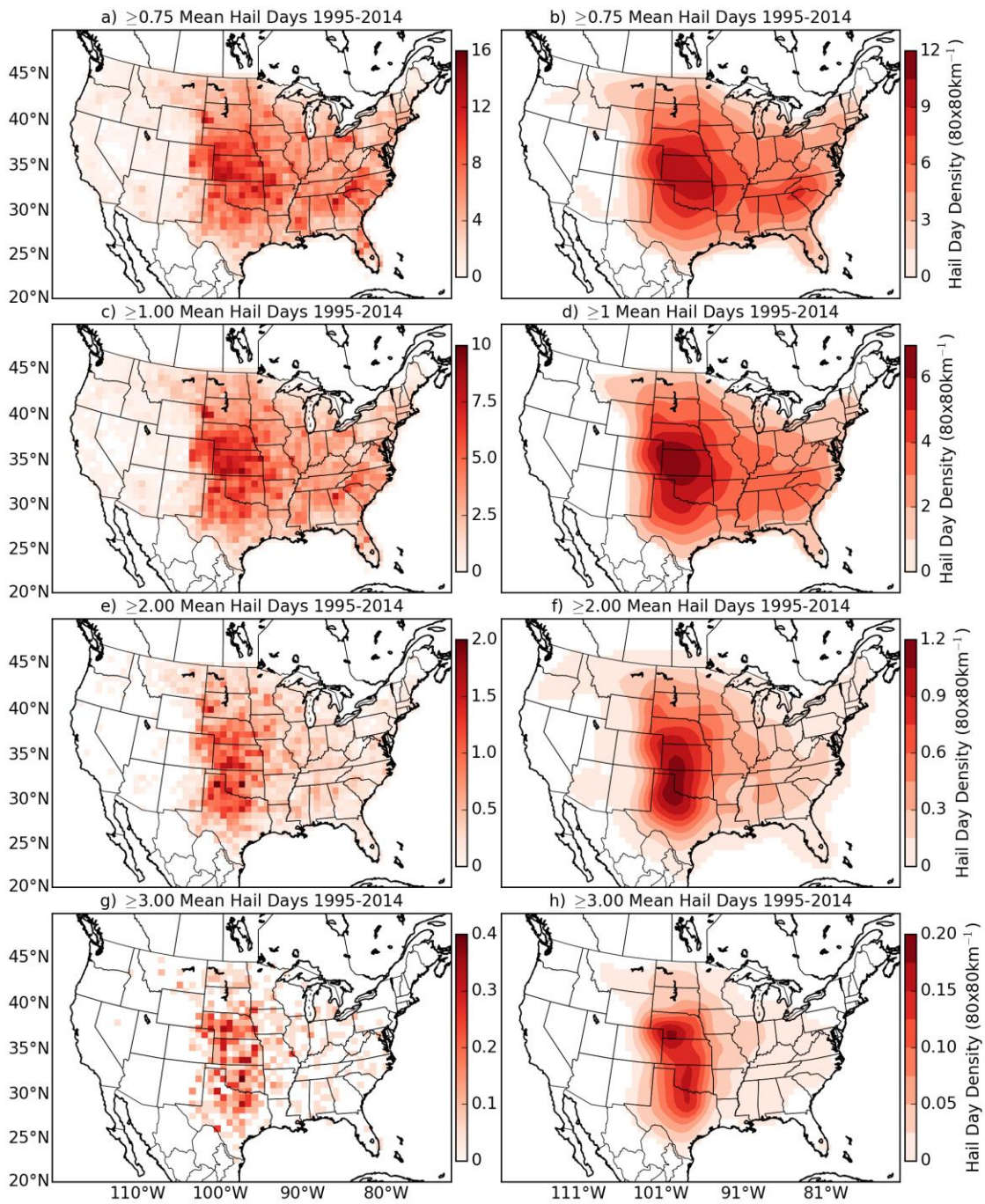
Figure 9: Mean annual frequency of gridded and Gaussian kernel smoothed hail day density for hail diameter equal to or greater than given thresholds for the recent two decades (1995–2014): a) gridded hail day density ≥0.75 in (1.9 cm); b) smoothed hail-day density ≥0.75 in (1.9 cm); c) as for (a) except hail days ≥1 in (2.5 cm); d) as for (b) except hail days ≥1 in (2.5 cm); e) as for (a) except hail days ≥2 in (5.1 cm); f) as for (b) except hail days ≥2 in (5.1 cm); g) as for (a) except hail days ≥3 in (7.6 cm); d) as for (b) except hail days ≥3 in (7.6 cm). *Click image to enlarge.*

b. Hail days and subjective choices

The problem with multiple reports of the same hailstorm over metropolitan areas as illustrated in Fig. 6 and discussed in the literature (Wyatt and Witt 1997; Blair et al. 2014: Paulikas 2014; Allen et al. 2015a) leads us to consider hail days as an alternative measure. Using this metric, a grid box for each location receives a binary value of zero or one for the presence of hail. The result is a depiction of the gridded and smoothed density of mean annual hail days (Fig. 8), which is somewhat less sensitive to the presence of population centers. The number of hail days is substantially smaller than the number of hail reports at any given location, and the peak value of 32 reports per year is reduced to 8 hail days per year for subsevere hail. The spatial distribution of hail days relative to that of hail reports shifts towards areas with a longer reporting record, e.g., the Plains states, particularly Oklahoma.

This reduction in peak frequency persists for both the severe and significant severe categories, with peak frequency of only six and two hail days per year, respectively. The smoothing approach also reveals limitations in working with the observed hail-day data. Individual grid boxes with high values are smoothed considerably when surrounded by relatively low frequencies, and thus displaying both the gridded and smoothed visualization may reveal useful information about the limitations of the dataset.

Smoothing these data, the peak frequency of subsevere hail-day density is found between Kansas and Oklahoma, extending northwestward into Colorado, and gradually declining northward into the Dakotas, and southward into northern Texas. Smaller regions of increased reports are found stretching between Indiana and Ohio, and from northern Alabama to South Carolina. The frequency of severe-hail reports in the Southeast decreases, suggesting the majority of days for at least some part of the time series is subsevere and the peak frequencies consolidate over the Plains states.

Subjective choices of time periods for studying hail occurrence are also a problematic contribution, more so than found for tornadoes (Fig. 9). Considering the last two decades, peak values of the gridded and contoured data for hail days are found to double per year relative to the entire record (unsurprising given 76% of the climatology is found in this period), and similar changes are possible on small regional scales for all but the last decade. Also, the spatial pattern for reports of severe threshold begins to look more like that for subsevere hail, a characteristic that may be related to the change to minimum diameter in 2010 (see section 6).

For significant hail, hail days are mostly confined east of the Rockies from the southern to northern Plains, with only modest numbers of reported events in the Southeast. This results in a smaller peak region that extends farther southward from southern Nebraska through western Oklahoma and into Texas, with the majority confined to the Plains states. However for 3-in (7.6-cm) hail, we see that a relatively small number of reports results in the peak frequency in northern Kansas and southern Nebraska, and extending southwards along the I-35 corridor in Oklahoma. This likely reflects a required greater density of population to identify 3-in (7.6-cm) hail, rather than suggesting that such events are rare elsewhere (Blair et al. 2014).

There are two important concepts to take from these results: 1) stratifying the dataset according to size reduces the available data, resulting in unusual or undesired small-scale spatial anomalies, and 2) choices made when gridding, smoothing, contouring and limiting temporal windows may lead to inconsistent and nonmeteorologically driven results in terms of variability.

6. Influence of the severe-hail-criteria change in 2010

The 2010 change in severe hail criteria has been illustrated above to have a marked influence on the distribution of hail reports by size. To further investigate how this change impacts the spatial record, we stratify into subsevere and severe hail, and consider isopleths for two equal 5-y periods either side of the change along with point reports and their change between the periods (Fig. 10). This period choice was made to reduce the influence of natural variability illustrated in Fig. 3 and avoid the influence of outlier years such as 2011.

For the period 2005–2009 (Fig. 10a,b) peak frequency for both subsevere and severe hail is relatively similar, though higher isopleths extend farther eastward for subsevere hail. This is particularly evident in the Southeast between Alabama and South Carolina, which has a higher frequency compared with the Plains. However, for the 5-y period following the change, we

identify a large shift in the frequency (Fig. 10c,e) from subsevere to severe hail days, reducing the number of subsevere hail days by more than half to two-thirds, while point reports become almost nonexistent for Oklahoma, Nebraska, Wyoming and Kentucky. This is contrasted by a large increase in reported severe hail that shifts the higher contours of the distribution farther northward and eastward compared to the preceding period (Fig. 10d,f). However, for the 5-y period following the which has a higher change, we identify a large shift in the frequency (Fig. 10c,e) from subsevere to severe hail days, reducing the number of subsevere hail days by more than half to two-thirds, while point reports

become almost nonexistent for Oklahoma, Nebraska, Wyoming and Kentucky. This is contrasted by a large increase in reported severe hail that shifts the higher contours of the distribution farther northward and eastward compared to the preceding period (Fig. 10d,f).

This result appears to demonstrate a wholesale shift in severe hail reporting. Two potential hypotheses to explain this shift are: the size of hail is increasing, which does not appear the case, except for near-severe hail sizes based on Fig. 4, or there is a systematic reporting change occurring that increases hail diameter by a range of 0.25 in to 0.5 in (0.63–1.27 cm).
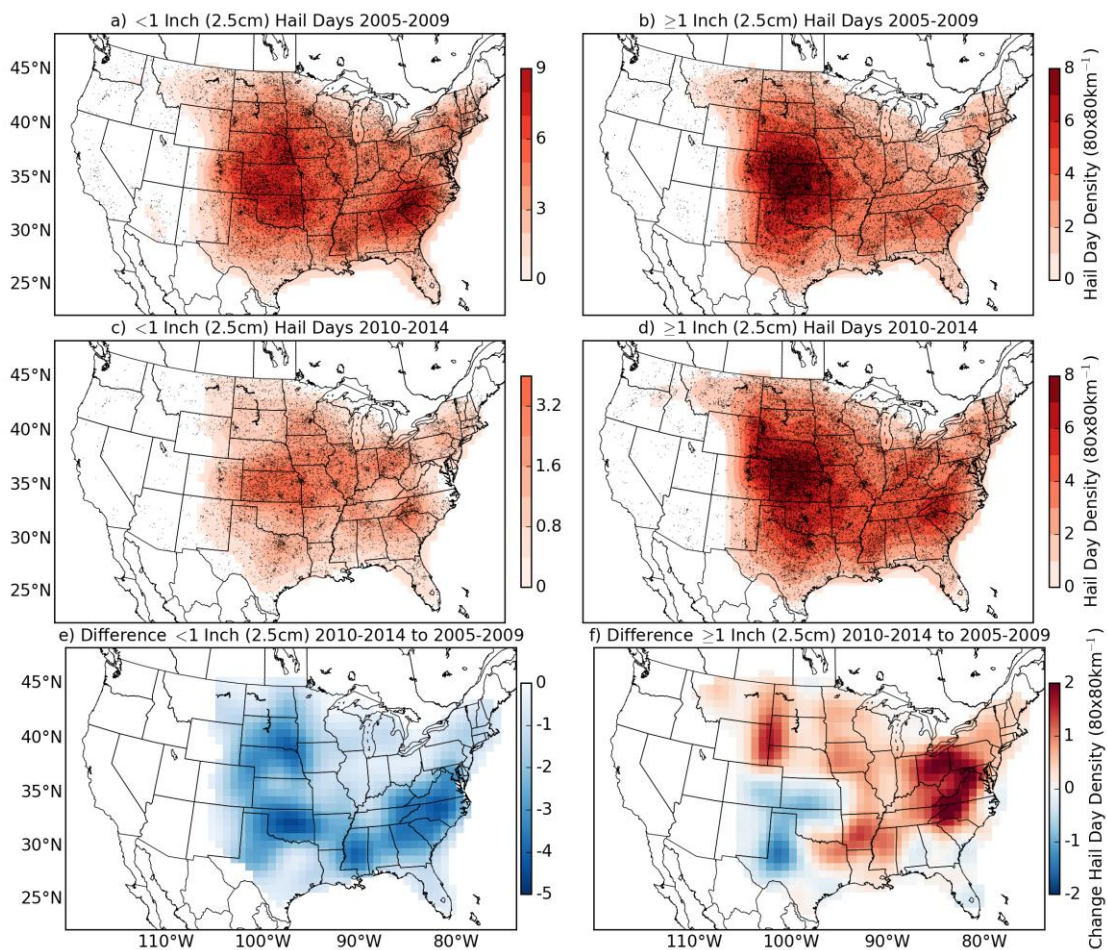

Figure 10: Differences in hail reporting either side of the severe hail size change at the beginning of 2010: a) Mean annual hail < 1 in (2.5 cm) day density per 80×80 km^{-1} box for the period 2005–2009, overlain by all hail reports <1 in (2.5 cm) for the same period; b) as for (a) except for hail days ≥1 in (2.5 cm) and reports meeting the same condition; c) as for (a), except mean annual hail days for 2010–2014, overlain by hail reports for the same period; d) as for (b) except mean annual hail days ≥1 in (2.5 cm) for 2010–2014; e) difference between the number of hail days <1 in (2.5 cm) for the 2010–2014 period and the 2005–2009 period; f) as for (e) except for hail days ≥1 in (2.5 cm). Peak values of the difference (e,f) are equal to a −33% to 25% change in the overall frequency (a–d). *Click image to enlarge.*
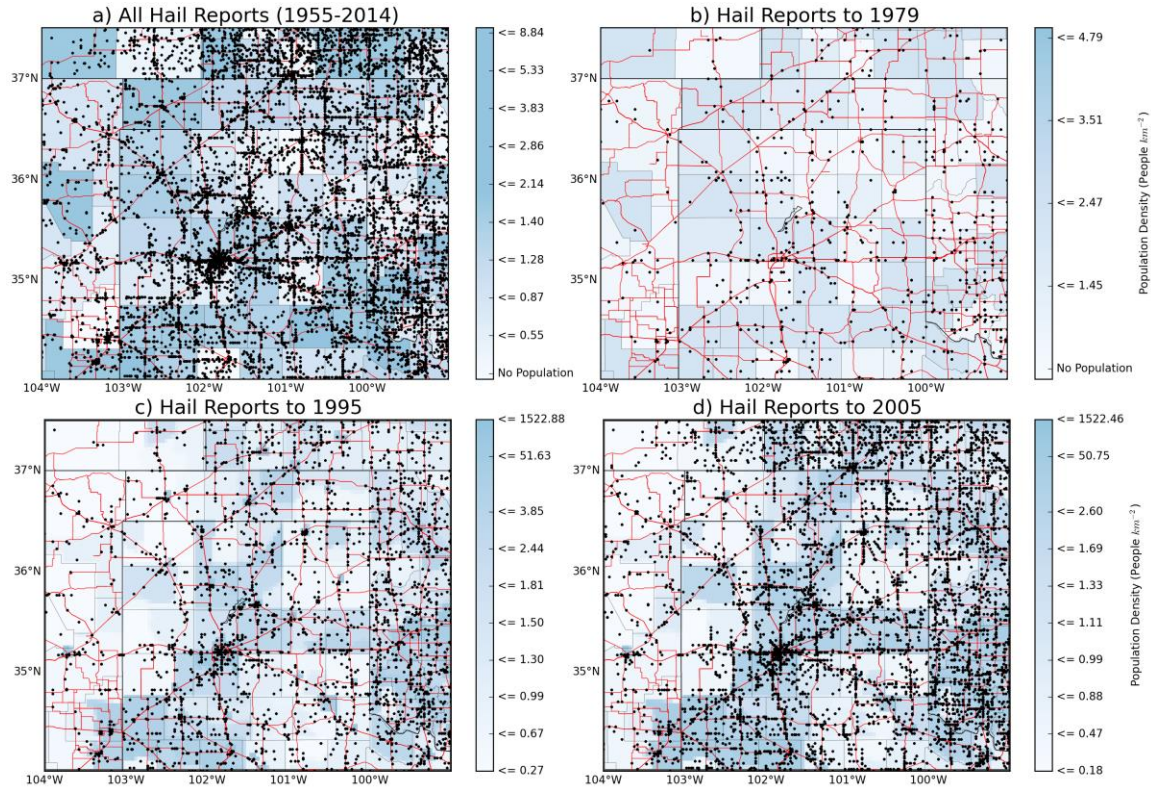
Figure 11: Point reports of hail >0.75 in (1.9 cm), over the Texas Panhandle and surroundings, with population choropleth of intercensal estimated population segregated by Jenks Natural Breaks: a) all reports 1955–2014, shown with mean population 1979–2012; b) hail reports 1955–1979 with 1979 population; c) as for (b) except hail reports 1955–1995 and 1995 population from the CIESIN gridded global population data; d) as for (c) except hail reports 1955–2005 and 2000 population from the CIESIN data. Primary interstates and highways are shown in red. *Click image to enlarge.*

Plausibly, this reporting change was induced by the new standard diameter for severe hail, potentially to ensure the verification of severe thunderstorm warnings.   While Blair et al. (2014) suspect that maximum hail size for storms is underestimated, this shift is not driven by natural variability and will act as a breakpoint in the hail dataset, similar to the impact of the shift to rating tornadoes using the enhanced Fujita scale in 2007 (Doswell et al. 2009; Edwards et al. 2013).

7. Impacts of population and report sources

Several studies have proposed that population can be used to adjust tornado report frequency (Elsner et al. 2013; Widen et al. 2013). However we illustrate here using a case study of the Texas panhandle that considering this factor alone may not be an effective approach to remove inhomogeneities in the reported hail data. The spatial clustering of reports in this area has

previously been shown to have little relationship to environmental characteristics, and instead is related to road networks as well as population centers (Allen et al. 2015a). Unlike tornadoes, hail is reported when an observer sees fallen hail during or following a storm and then provides a size estimation or formal measurement. This reporting procedure contrasts to tornadoes, where a Great Plains observer in many cases can see an event clearly up to ten to twenty miles away, and the event is typically post surveyed if it results in any assessable damage. Thus, to some extent, the clustering of hail reports towards road networks is somewhat plausible, as it reflects the likelihood of hail interacting with the local population who could make reports. However, if the road network was the sole reason for these reports, this set of observations should be relatively stationary over time. This is not the case for the Texas Panhandle, which experiences a considerable increase in reports, with additional clustering towards both population

centers and road networks (Fig. 11). Further complicating this puzzle, we see that in eastern New Mexico and the less populated western Texas Panhandle there is only a relatively small comparative increase in hail reports, which reflects a reduced road network and lower population. Overlaying this on population density and the road network, we note that few counties in either region have positive trends in population in the recent two decades and the majority exhibit decreases, yet the proportion of reports along road networks has increased relative to the clustering around population centers. We can extrapolate this information to two potential sources of reports:

- Increased telecommunications, mobile internet and ease of reporting via social media, websites and mobile phones, leading to road users increasingly reporting hail in line with the suggestions of Tuovinen et al. (2009).

- An increased presence of storm chasers in the Great Plains appearing to coincide with the growth of this interest group leading into the late 1990s, and rapid growth in the 2000s and onwards.

As the resident population has not increased appreciably, local road network usage likely has not changed appreciably outside of interstate routes or major highways. Thus, simply modeling changes to frequency of hail reports based on the local population will fail to capture the nonstationary nature of hail reports in these areas, and more complex modeling approaches are necessary. To further extract potential contributions to such changes, we next consider the observer sources of hail reports.

Reports of severe thunderstorms come from observers who may or may not have meteorological training (Doswell et al. 2005). As illustrated above, factors beyond population are influencing the trends in the hail climatology and these vary from region to region. Thus, one way we can examine the reason for this characteristic is to analyze the original source of these hail observations. For each raw hail report from the period 1998–2014, the report source field was recorded (Table 1). Sources are not consistently named over time, and we group them into 17 primary categories.

Table 1: Sources of hail-report data 1998–2014, with the condensed category list shown on the left, and the categories that comprise them shown on the right.

Source Category	Included Categories
NWS EMPLOYEE	NWS Employee (Off Duty)
SOCIAL MEDIA	Social Media
UNKNOWN	Unknown, Empty
GOVERNMENT OFFICIAL	State Official, County Official, Govt Official
LAW ENFORCEMENT	Law Enforcement
SHAVE/COCORAHS	SHAVE project, COCORAHS
STORM CHASER	Storm Chaser
BROADCAST MEDIA	Broadcast Media, TV Meteorologist
EMERGENCY MANAGER	911 Call Center
NWS STORM SURVEY	Official NWS Observations
TRAINED SPOTTER	Meteorologist (Non-NWS), Mariner, Airplane Pilot
FIRE/RESCUE	Fire Department, Rescue Squad
FEDERAL AGENCiES	Park/Forest Service, Post Office, Department of Highways
AMATEUR RADIO	Amateur Radio
PUBLIC	General Public, Insurance Company, Public Utilities
NEWSPAPER	Newspaper
WEATHER STATIONS	AWOS, ASOS, Mesonet, RAWS, NPOP, COOP Observer

To illustrate how observation sources contribute to the frequency of hail reporting, the time series of source by report fraction was analyzed (Fig. 12). Law-enforcement hail reports were the third largest contribution (20% per year) until 2003, when they began to decline steadily, and now contribute <5% of reports per

year. Since 2003, a greater number of reports have originated from broadcast media, fire and rescue, NWS employees and weather stations. Another interesting change was the fraction of reports sourced from storm chasers in 1999, and subsequent low frequency until 2009 when the frequency rebounded to levels similar to those in

1999. Other changes include the growth in reports from government officials after 2005, the total decline of NWS storm survey efforts for hail after 2006, and increasing contributions from weather stations, observational field programs (e.g., the Severe Hazards Analysis and Verification Experiment; SHAVE; Ortega et al. 2009), crowd-sourced hail observations (e.g., the Community Collaborative Rain, Hail and Snow Network; CoCoRAHS 2015), and the rapid increase of social media as a category since its inclusion in 2012.
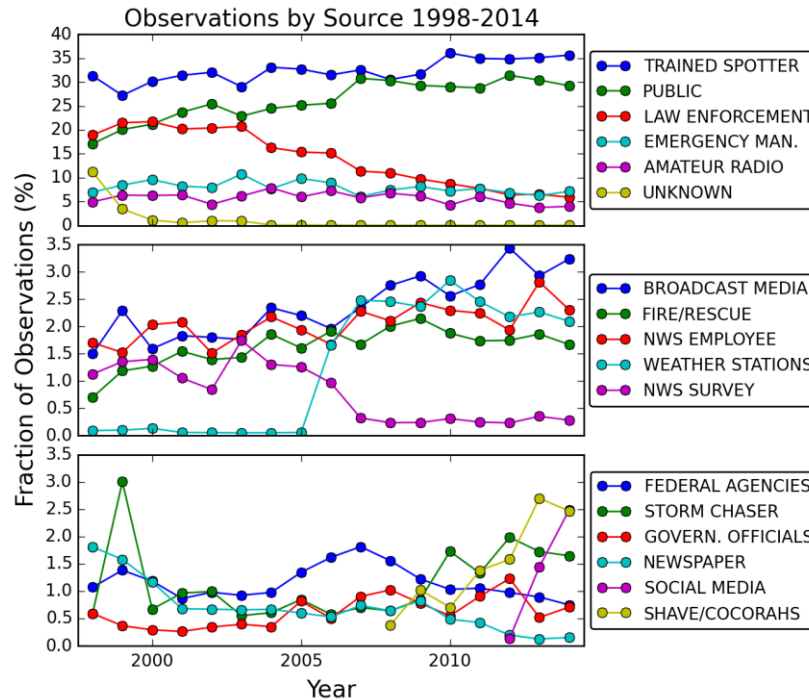

Figure 12: Time series of the fraction of hail reports ≥0.75 in (1.9 cm) by the respective sources described in Table 1 for the period 1998–2014. *Click image to enlarge.*

Table 2: Compiled sources of hail reports for the period 1955–2014, listing the number of hail reports ≥ 0.75 in (1.9 cm), ≥ 2 in (5.1 cm) and ≥ 3 in (7.6 cm), along with the fraction these contribute to all reports of the respective size thresholds.

| Source	All Hail Reports		2 Inch+ (5cm) Hail		3 Inch (7.5cm) Hail	
	Obs.	%	Obs.	%	Obs.	%
NWS EMPLOYEE	4399	2.06	284	2.81	45	3.02
SOCIAL MEDIA	403	0.19	40	0.40	5	0.34
UNKNOWN	1488	0.70	75	0.74	9	0.60
GOVERNMENT OFFICIAL	1377	0.64	44	0.44	11	0.74
LAW ENFORCEMENT	29445	13.76	1269	12.56	202	13.57
SHAVE/COCORAHS	1235	0.58	100	0.99	12	0.81
STORM CHASER	2355	1.10	331	3.28	62	4.16
BROADCAST MEDIA	5128	2.40	422	4.18	74	4.97
EMERGENCY MANAGER	17205	8.04	884	8.75	135	9.07
NWS STORM SURVEY	1626	0.76	160	1.58	38	2.55
TRAINED SPOTTER	69313	32.39	2890	28.61	392	26.33
FIRE DEPARTMENT/RESCUE	3591	1.68	209	2.07	39	2.62
FEDERAL AGENCiES	2532	1.18	90	0.89	16	1.07
AMATEUR RADIO	12561	5.87	582	5.76	74	4.97
PUBLIC	57043	26.66	2328	23.04	299	20.08
NEWSPAPER	1406	0.66	264	2.61	58	3.90
WEATHER STATIONS	2862	1.34	130	1.29	18	1.21
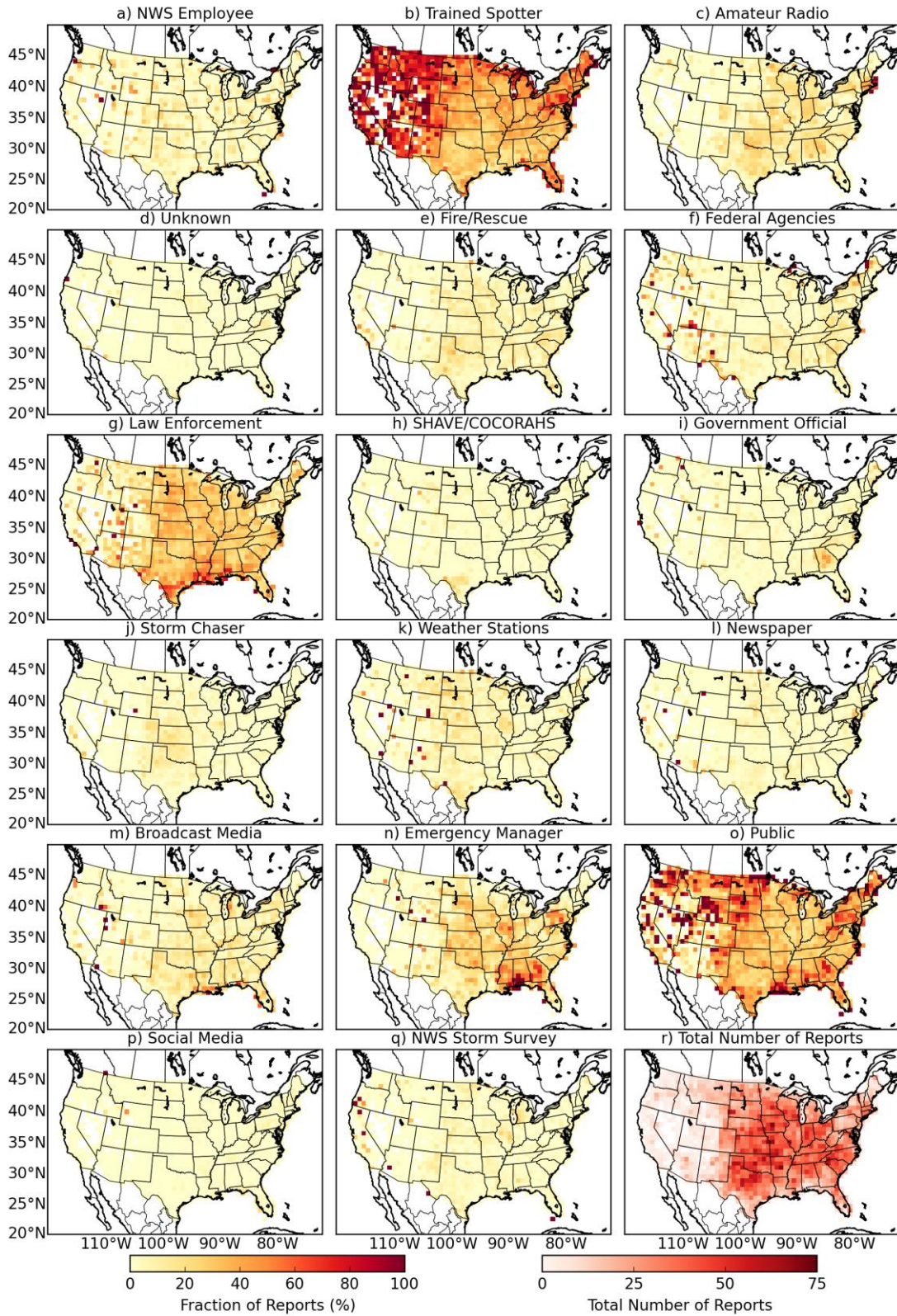
Figure 13: a–q) Fraction of the total gridded ≥0.75 in (1.9 cm) hail reports originating from the respective observation sources (left scale) for the period 1998–2014; r) total number of reports with source information available 1998–2014 (right scale). *Click image to enlarge.*

Another perspective is given by the relative fraction of reports of a given size that sources add to the dataset (Table 2). In terms of contributions stratified by size, only small variations were found, mainly reflecting a decreasing fraction of public and trained spotter reports for significant hail, and increasing fractions of storm chaser reports (quadrupling to 4.2% of the total for ≥3 in (7.6 cm) hail), media sources, NWS employees, and other emergency management sources. Considering the time series for significant hail, few changes are noticeable, except for the jump from the lowest set to the 3rd highest source of reports by storm chasers since 2009, and the increasing contributions of social media and field programs.

We also can consider the relative fraction of reports by the respective sources on the same gridded projection as the SPC filtered hail observation data, to appraise spatial biases in source (Fig. 13). The relative fraction of reports from sources such as the public or trained spotters provide the largest proportion of reports over all grid locations based on the total frequency, which is true for many locations. However, some of the smaller categories are not spatially uniform. For example, near each regional forecast office, there are a higher proportion of reports originating from NWS employees or storm surveys in the earlier record (Fig. 14 a,q), consistent with the issues raised by Doswell et al. (2005). Given few NWS employees are located away from the forecast office's grid box, the low values are not unexpected, but it does reveal that a large fraction of reports in the vicinity of areas of population (40+%) originate from this source, rather than solely due to increased population.

Storm-chaser reports also illustrate a spatial bias towards the Plains states, but only contribute a relatively small fraction (≈20%). A larger number of reports likely are associated with storm chasers in the Plains, where population is limited, and are either misclassified as public or entered instead as trained spotter. Also, a considerable fraction of reports in the Atlanta region of Georgia originate from county and state officials (≈50%, Fig. 13i), a characteristic not found anywhere else in the country, suggesting perhaps a local agreement for reporting hail. A similar pattern of federal-agency reporting is identified for much of the Southeast, Oklahoma and into the northeast.

Law enforcement, one of the largest categories, contributes throughout the country and typically provides 20–40% of reports; however, a substantial difference is highlighted for the Gulf Coast area and southern Texas, where fractions are 60–80% of the relatively small number of reports (Fig. 13g). A similar effect can also be identified in other parts of the Gulf Coast for the emergency-manager category (Fig. 13n), and the public source in southern Louisiana. Social media thus far only contributes a small fraction for the areas surrounding forecast offices in the eastern portion of the country, and contributes little to hail reports in the Plains states (Fig. 14p). This implies a stronger argument for increased telecommunications, storm chasers and trained spotters contributing to the growth in reports for the Texas Panhandle.

While field programs are a growing contribution, their relative influence is mainly confined to southern Texas and a limited number of reports scattered through the northern Plains (Fig. 14h). Though numerous regional intricacies could be illustrated by these results, here we highlight how regional differences in collection of hail reports may artificially inflate or decrease the number of hail events in collated time series or spatial maps.

8. Discussion

The lack of an appreciable increase in the number of hail days in the past decade, in contrast to the growing number of reports, suggests to account for duplicate reporting when considering the hail data. Not all changes in frequency of reports are solely due to the reporting problem, and some of the differences may relate to report sources. These results imply that researchers from a wide variety of backgrounds using hail data (including insurers, climatologists, radar- and satellite-based climatology and detection verifications, etc.) need mind a vast array of potential confounding factors that may influence their results.

We have outlined a number of characteristic factors that the community should understand, however this list cannot be considered exhaustive. We can be more confident in pooled statistics or patterns derived from hail observations than we can be of localized regions; examples of this include the annual and diurnal cycles, national or regional time series, and by using hail days rather than hail reports.
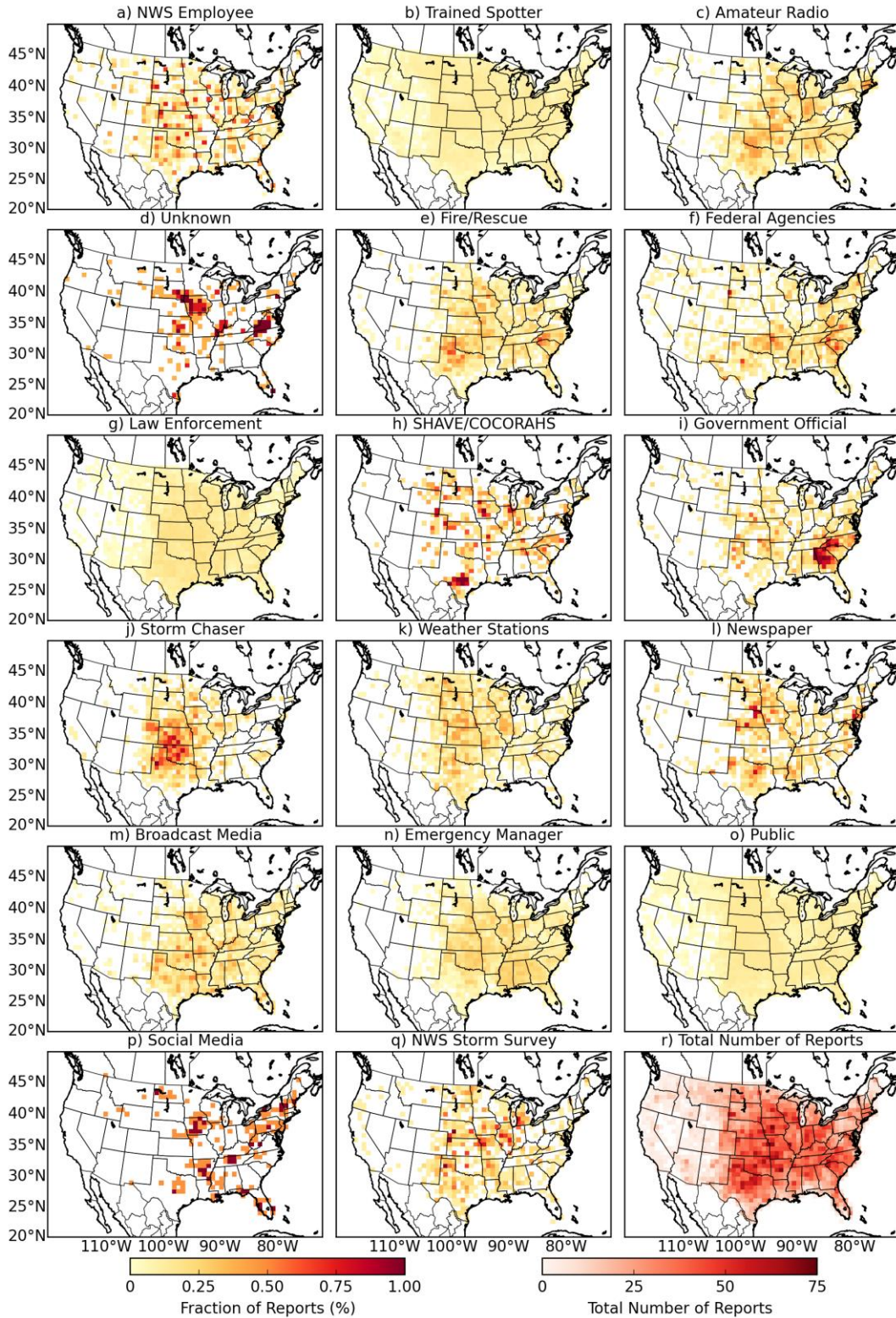
Figure 14: a–q) As for 13, except instead showing the fractional contribution of each grid cell (left scale) to the total number of each observation source (i.e., as a percentage of the total number of reports of a given source summed across all grid points, and thus the fraction at all grid points sums to 100% of a given source) for the period 1998–2014. *Click image to enlarge.*

The large temporal changes in the period after 1994 suggest that in order to obtain a fair representation of mean hail climatology, it is necessary to restrict the hail data chosen to the last one to two decades. Even then however, caution must be taken in handling the heavy size quantization, and 2000 may be a better starting point for climatologies of observations for comparisons to radar or satellite derived products (e.g. Cintineo et al. 2012; Cecil and Blankenship 2012). It also remains to be seen how much of a problem the change to the severe hail diameter will cause in the long term, as for the past 5 y the influence of this change appears substantial (Figs. 4, 5, 10).

Applying simple corrections to rectify spatial biases in hail also appears problematic. Unlike tornado reports, which are often surveyed by the National Weather Service and can be observed easily from locations within 5 mi (8 km) of the event, hail reports require an observer to be present in or shortly after the storm to make a valid observation. As we have illustrated here, this leads to a spatial bias towards road networks in the hail dataset, which is becoming increasingly problematic as the number of storm chasers in the Plains increases. Similar results can be presented for other locations, such as western Kansas, eastern Colorado and western Nebraska; however, the greater density in the local road network may weaken the relationship to major highways.

Identifying the influence of storm chasers on the hail report dataset is also not simple, as in NWS records these observers are classified into a mixture of categories, including the public, trained-spotter and storm-chaser subsets. This suggests that a positive move for the dataset would be to ensure the separation of this category for future analysis. The justification for this change would be separating by level of expertise: a storm chaser can encompass an experienced individual or researcher familiar with accurate or measured reporting, or a member of the public observing a storm for the first time. In contrast, trained spotters have (at the very least) completed rudimentary training about how to make a severe weather report. As we move forward, the fraction of observation sources likely will continue to move towards mobile reporting, online or social media derived information, rather than traditional sources such as broadcast or print media, and these changes are evident in the recent data for the source of reports for 1998–2014.

Regional oddities remain in the relative fraction of sources that contribute reports, and these appear to vary greatly between regional forecast offices. Whether this reflects a local policy characteristic to value one source of reports over another, or an absence of the more traditional observer sources remains unknown. Exploring these CWA-induced characteristics in more detail would also be a valuable course of future work (similar to Weiss et al. 2002), which would allow improved understanding for local verification of simulations or understanding changes in reports over shorter temporal windows.

Given that the frequency of reports has been relatively stable outside natural variability in the past 15 y, we can be confident that changes to verification practices should not have a marked effect on the number of hail days. However, the shift in the minimum hail size in 2010 from 0.75 in (1.9 cm) to 1.00 in (2.5 cm) as severe, appears to have clearly influenced the fraction of reports that are now recorded as severe as well as the spatial distribution, and potentially influences hail reports between severe and 2 in (5.1 cm) in diameter. This would imply that hail sizes were being inflated to the new severe threshold in order to meet warning verification criteria, rather than reflecting physical measurements. Avoiding similar changes to the arbitrary thresholds we use to define hail size may assist in avoiding future negative influences on the dataset quality. For end users, the best way to avoid these biases may be to use thresholds that are distinct from the official values.

Another approach to mitigate potential trends in hail size, or its misrepresentation, would be to enforce transition from solely textual hail reports to a visual confirmation system, with a common reference object or ruler alongside to allow appropriate interpretation of size as suggested by Blair and Leighton (2012). This would also assist in alleviating the present difficulty for researchers interested in hail size caused by the quantization of hail reports to reference objects rather than the use of measured hail diagnostics.

Further, the value of hail diameter as the sole measure of hail magnitude is an ongoing problem in the science, as diameter is not the sole contribution to a hailstone's density or mass. That two stones on their longest axis can be of equal diameter and yet on their secondary or tertiary axis be of varying dimension has broad

implications for the potential fall velocity of the hailstone (proportional to the mass of a spheroid object), and commensurate damage to property. A similar problem also exists for the density of hailstones, which are rarely uniform between or within storms. Recent field efforts have moved toward understanding density and multi-dimensional measurement, but these observations are still not extensive, and do not yet contribute to the NCDC hail dataset (Brown et al. 2014; Giammanco et al. 2014). Including dimensional information or even the weight of a given stone would allow improved recognition of conditions favorable to large hail sizes and assist in providing a more accurate warning, informing models or postprocessed radar information which can project potential hail size. These discussions also have been made recently using the results of field campaigns to identify the underreporting of hail size (e.g., Ortega et al. 2009; Blair et al. 2014; Heymsfield et al. 2014) and illustrate that there is a need to improve how we collate hail data in the national climatic datasets, similar to the noted need to improve the metadata associated with tornadoes (Edwards et al. 2013). Another complimentary perspective would be reaching out to the storm chaser and trained spotter communities to ensure they are familiar with appropriate measuring techniques for hail dimensions. Attaching a greater importance to using simple digital-age reporting tools likely will be the best way to move forward past the current paradigm.

In view of the limitations of the observed hail dataset, we advocate caution in examining whether the results obtained via analysis reflect real climate signals, or are a result of temporal inhomogeneities. Simple tests involving removal of outliers, and subsampling of climatological periods will likely reveal these limitations, as suggested by Doswell (2007). Authors also should understand that observations may not reveal a climatologically significant signal, but this does not imply the *absence* of a climatic influence on hail. As we, and others have demonstrated elsewhere, climate signals in hail occurrence may be masked by interannual variability and the large nonmeteorological changes in observations, but the same is much less true for environments (e.g. Brooks et al. 2003b; Barrett and Henley 2015; Allen et al. 2015a; Allen et al. 2015b). Thus in the context of future climate-relationship studies, we recommend (in addition to observations and rigorous scrutiny of the reality of temporal and spatial statistics) that

environmental proxies be used to provide additional evidence that any such connection truly is related to the underlying meteorology.

REFERENCES

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015a: An empirical model relating U.S. monthly hail occurrence to large-scale meteorological environment. *J. Adv. Model. Earth Syst.*, **7**, 226–243. doi: 10.1002/2014MS000397.

——, ——, and ——, 2015b: Influence of the El Nino/Southern Oscillation on tornado and hail frequency in the United States. *Nature Geosci.*, **8**, 278–283, doi: 10.1038/ngeo2385.

Barrett, B. S., and B. N. Henley, 2015: Intraseasonal variability of hail in the contiguous United States: Relationship to the Madden–Julian Oscillation. *Mon. Wea. Rev.*, **143**, 1086–1103. doi: http://dx.doi.org/10.1175/MWR-D-14-00257.1.

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

——, D. R. Deroche, J. M. Boustead, J. W. Leighton, B. L. Barjenbruch, and W. P. Gargan, 2011: A radar-based assessment of the detectability of giant hail. *Electronic J. Severe Storms Meteor.*, **6** (7), 1–30.

——, D. Cavanaugh, J. Laflin, J. Leighton, K. Sanders, and K. Ortega, 2014: High-resolution hail observations: Implications for NWS warning operations and climatological data. Preprints, *27th Conf. on Severe Local Storms,* Madison, WI, Amer. Meteor. Soc., 18.1.

Brooks, H. E., 2013: Severe thunderstorms and climate change. *Atmos. Res.*, **123**, 129–138.

——, and N. Dotzek, 2007: The spatial distribution of severe convective storms and an analysis of their secular changes. *Climate Extremes and Society*. H. F. Diaz and R. Murnane, Eds., Cambridge University Press, 35–53.

——, C. A. Doswell III, and M. P. Kay, 2003a: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640.

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

——, G. W. Carbin, and P. T. Marsh, 2014: Increased variability of tornado occurrence in the United States. *Science*, **346**, 349–352.

Brown, T. M., I. M. Giammanco, and M. R. Kumjian, 2014: IBHS Hail Field Research Program 2012–2014. Preprints, *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 16B.1.

Cecil, D. J. and C. B. Blankenship, 2012: Toward a global climatology of severe hailstorms as estimated by satellite passive microwave imagers. *J. Climate*, **25**, 687–703.

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

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

——, 2008: Temporal and spatial distributions of damaging hail in the continental United States. *Phys. Geography*, **29**, 341–350.

——, and D. Changnon, 2000: Long-term fluctuations in hail incidence in U.S. *J. Climate*, **13**, 658–664.

——, ——, and S. D. Hilberg, 2009: *Hailstorms Across the Nation: An Atlas about Hail and its Damages.* Illinois State Water Survey, 95pp.

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.

CoCoRAHS, cited 2015: CoCORAHS: The Community Collaborative Rain, Hail and Snow Network. [Available online at http://www.cocorahs.org/Media/docs/CoCoRaHSBrochure15FINAL.pdf.]

Coleman, T. A. and P. G. Dixon, 2014: An objective analysis of tornado risk in the United States. *Wea. Forecasting*, **29**, 366–376.

Crum, T. D., and R. L Alberty, 1993: The WSR-88D and the WSR-88D Operational Support Facility. *Bull. Amer. Meteor. Soc.*, **74**, 1669–1687.

Doswell, C. A. III , 2007: Small sample size and data quality issues illustrated using tornado occurrence data. *Electronic J. Severe Storms Meteor.*, **2** (5), 1–16.

——, 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.

——, ——, and N. Dotzek, 2009: On the implementation of the enhanced Fujita scale in the USA. *Atmos. Res.*, **93**, 554–563, doi:10.1016/j.atmosres.2008.11.003.

Edwards, R., J. G. LaDue, J. T. Ferree, K. Scharfenberg, C. Maier, and W. L. Coulbourne, 2013: Tornado intensity estimation: Past, present, and future. *Bull. Amer. Meteor. Soc.*, **94**, 641–653.

Elsner, J. B., L. E. Michaels, K. N. Scheitlin, and I. J. Elsner, 2013: The decreasing population bias in tornado reports across the central Plains. *Wea. Climate Soc.*, **5**, 221–232.

——, S. C. Elsner, and T. H. Jagger, 2014: The increasing efficiency of tornado days in the United States. *Clim. Dynamics,* **45**, 651–659, doi: 10.1007/s00382-014-2277-3.

Giammanco, I. M., T. M. Brown, and M. R. Kumjian, 2014: Observations of hailstone sizes and shapes from the IBHS Hail Measurement Program: 2012–2014. Preprints, *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc, 16B.1.

Hales, J. E. Jr., 1993: Biases in the severe thunderstorm data base: Ramifications and solutions. Preprints, *13th Conf. on Weather Forecasting and Analysis*, Vienna, VA, Amer. Meteor. Soc, 504–507.

Heymsfield, A. J.*,* I. M. Giammanco*,* and R. Wright*,* 2014*:* Terminal velocities and kinetic energies of natural hailstones. *Geophys. Res. Lett.,* **41**, 8666–8672, doi: 10.1002/2014GL062324.

Hyvärinen, O. and E. Saltikoff, 2010: Social media as a source of meteorological observations. *Mon. Wea. Rev*., **138**, 3175–3184.

Jenks, G. F., 1967: The data model concept in statistical mapping. *Int. Yearbook Cartogr.*, **7**, 186–190.

Johnson, A. and K. Sugden, 2014: Evaluation of sounding-derived thermodynamic and wind-related parameters associated with large hail events. *Electronic J. Severe Storms Meteor.*, **9** (5), 1–42.

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.

Kumjian, M. R., J. Evans and J. Guyer, 2006: The relationship of the Great Plains low level jet to nocturnal MCS development. Preprints, *23rd Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., P1.11.

Marsh, P. T., and H. E. Brooks, 2012: Comments on "Tornado risk analysis: Is Dixie Alley an extension of Tornado Alley?" *Bull. Amer. Meteor. Soc.*, **93**, 405–407.

Mead, C. M., and R. L. Thompson, 2011: Environmental characteristics associated with nocturnal significant-tornado events in the central and southern Great Plains. *Electronic J. Severe Storms Meteor.*, **6** (6), 1–35.

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

——, 1987: The hybrid multicellular–supercellular storm—an efficient hail producer. Part II: General characteristics and implications for hail growth. *J. Atmos. Sci.*, **44**, 2060–2073.

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. Amer. Meteor. Soc.*, **90**, 1519–1530.

Paulikas, M. J., 2014: Examining population bias relative to severe thunderstorm hazard reporting trends in the Atlanta, GA metropolitan region. *Meteor. Appl.*, **21**, 494–503, doi: 10.1002/met.1394.

Schaefer, J. T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. Preprints, *11th Conf. on Applied Climatology,* Dallas, TX, Amer. Meteor. Soc., 603–606.

——, J. J. Levit, S. J. Weiss, and D.W. McCarthy, 2004: The frequency of large hail over the contiguous United States. Preprints, *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc*,* 3.3.

Smith, P. L., 1999: Effects of imperfect storm reporting on the verification of weather warnings. *Bull. Amer. Meteor. Soc.*, **80**, 1099–1105.

Tippett, M. K., 2014: Changing volatility of U.S. annual tornado reports, *Geophys. Res. Lett.*, **41**, 6956–6961, doi: 10.1002/2014GL061347.

——, J. T. Allen, V. A. Gensini, and H. E. Brooks, 2015: Climate and hazardous convective weather. *Curr. Clim. Change Rep.*, **1**, 60–73*,* doi: 10.1007/s40641-015-0006-6.

Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinski, and R. Wolf, 2006: Buyer beware: Some words of caution on the use of severe wind reports in postevent assessment and research. *Wea. Forecasting*, **21**, 408–415.

Tuovinen, J.-P., A.-J. Punkka, J. Rauhala, H. Hohti, and D. M. Schultz, 2009: Climatology of severe hail in Finland: 1930–2006. *Mon. Wea. Rev.*, **137**, 2238–2249.

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1954–2003. *Wea. Forecasting*, **21**, 86–93.

Weiss, S. J., J. A. Hart, and P. R. Janish, 2002: An examination of severe thunderstorm wind report climatology: 1970–1999. Preprints, *21st Conf. Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 446–449.

Widen, H. M., J. B. Elsner, R. B. Cruz, G. Xing, E. Fraza, L. Migliorelli, S. Strazzo, C. Amrine, B. Mulholland, M. Patterson, and L. Michaels, 2013: Adjusted tornado probabilities. *Electronic J. Severe Storms Meteor.*, **8** (7), 1–12.

Wyatt, A. and A. Witt, 1997: The effect of population density on ground-truth verification of reports used to score a hail detection algorithm. Preprints, *28th Conf. on Radar Meteorology*, Austin, TX, Amer. Meteor. Soc., 368–369.

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

REVIEWER COMMENTS

[Authors' responses in *blue italics.*]

REVIEWER A (Walker S. Ashley):

Initial Review:

Recommendation: Accept with minor revisions.

General Comments: The authors provide a thorough overview of the U.S. hail database, including discussions of: data sources, changes in reporting strategies, and biases; spatial and temporal distributions and changes in climatology; and broad implications for (mis)interpreting hail observation signals. Overall, I have no—what I would consider—major suggested edits or requests. With a few minor changes and/or additions, the manuscript will be a nice addition to the body of literature that has examined severe storm climatology, as well as issues affiliated with the observed severe storm record.

We thank the reviewer for the positive feedback, and have made changes as suggested unless otherwise stated.

Substantive Comments: The issue regarding CWAs should be illustrated and discussed in more depth. The authors find and briefly discuss a couple examples of CWA-specific biases in the manuscript (Shreveport, Atlanta/Peachtree City), but further analysis and illustration of this issue is warranted. I recommend following the analyses provided by Weiss, S. J., J. A. Hart, and P. R. Janish, 2002: An examination of severe thunderstorm wind report climatology: 1970–1999. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 446–449. Using a CWA boundary file on specific report types for various temporal windows could uncover a number of biases. Initial exploration could be performed via *SeverePlot*. As a side, I recommend the authors reach out to the WCMs at the offices where specific oddities were found (SHV, FFC) to discuss what may be the influencing factors. This would remove any sort of speculation and would be very useful information for the readership.

We have included the suggested references to this material along with further clarification and discussion to illustrate our point. However, in response to the reviewer's suggestion, we are reluctant to pursue this analysis in any further detail here as it is beyond the explanation needed to understand the long-term climatological elements of the dataset, which is the focus of this paper. We do however point out that Figs. 6 and 7 illustrate this point on a decadal basis with the reports through time on a grid basis (we have now made these enlargeable to make the point reports more easy to discern)—to understand this for every CWA would take a considerable period and is beyond the scope of the work we present here. We clarified this in the original manuscript on page 3: "We also do not intend to question the reports or source of any one CWA over any other, only illustrate that the challenges of the hail report data extend beyond population or easily modeled corrections… It is also likely that characteristics on a local or county scale resulting from individual forecasters exist beyond what is detailed here, and these nuances may only be in the knowledge of local forecast offices or individuals."

From a philosophical perspective, we were also concerned that this would raise political issues that complicate this work unnecessarily, and would be better addressed in a separate paper exploring these aspects specifically in greater detail (for which we would welcome the reviewer's input).

How was the limited hail-swath information (which clearly has a CWA bias) employed?

This category was not addressed here as we felt it was limited in usefulness to researchers in the available set (only available consistently after 2008, but while 28.9% of data points had this information, the majority were repeated points, not a swath). Instead a report containing this information was simply taken from the starting latitude longitude point of the swath. We clarify this now in the data section for reproducibility. This particular characteristic probably warrants further investigation in its own right in

future exploration as the climatology grows, perhaps in the context of radar-derived swathes that would be useful for verification.

[Minor comments omitted...]

Second Review:

Recommendation: Accept with minor revisions.

General Comment: I am satisfied with the response from the authors. I think the manuscript will be in publishable form after a few minor issues are addressed.

[Minor comments omitted...]

REVIEWER B (Matthew R. Kumjian):

Initial Review:

Recommendation: Accept with minor revisions.

Summary: This very well-written manuscript is sorely needed in our community, and as such, makes a significant contribution to the literature. This should be an oft-cited paper in coming years. I found it extremely informative and enjoyable to read. Overall, I have only minor comments reflecting some clarifications, corrections, and small issues that should be taken into consideration. These are detailed below.

We appreciate the reviewer's feedback, and have made every attempt to address the comments and critiques.

[Editor's Note: A few comments initially labeled "minor" are included below, since they resulted in small but substantive revisions or clarifications.]

[M]any studies from the 1980s have demonstrated that CAPE is only a necessary (not sufficient) condition for large hail. In fact, several studies (e.g., Nelson 1983 *JAS*; Ziegler et al. 1983 *JAS*; Nelson 1987 *JAS*) suggest that storm-relative airflow patterns (not instability) are critical for large-hail production. Thus, part of this diurnal cycle may also be related to supercell morphology, allowing the storm to mature to the stage where favorable precipitation-growth trajectories are possible.

This is a really important point that wasn't appropriately stressed, we have clarified and included references as suggested by the reviewer.

I don't disagree about population biases probably playing a major role in this [orographic] discrepancy. However, does the NARR used in Allen et al. (2015a) have sufficient resolution to capture possible orographic effects on convective initiation and/or maintenance? I'm especially skeptical that the NARR's representation of topography for the lower boundary condition is sufficient to reveal any possible effects.

This is an excellent point, as we noted in the cited paper NARR's abilities around topography are already suspect, so this suggestion deserves at least clarification.

I think these results have broader applicability besides just climate researchers! Radar- and satellite-based studies, insurance companies, etc. come to mind, to name a few.

This is a great point we overlooked in our interest in our given applications. There is definitely broad applicability to the cautions presented here, and we have clarified as such.
[Minor comments omitted…]

REVIEWER C (Dennis Cavanaugh):

Initial Review:

Recommendation: Accept with minor revisions.

General Comments: Overall I think the authors have done an excellent job in identifying trends and biases in the evolution of the continental United States hail database. The analysis of these data appear to be well organized and logical, and the figures provided are extremely useful in helping the reader to visualize the trends and biases identified by the authors in this research. I think this paper identifies many important changes in this historical hail database that need to be accounted for in future studies that attempt to apply statistical analysis techniques to the data to relate hail occurrence to changes in the climate system. The statistical analysis techniques used to identify trends and biases in the data appear to be chosen and executed properly for the authors' stated purpose for performing this research.   My only substantive comment follows, but overall any recommended changes to the paper are what I would consider minor in nature.

We appreciate the reviewer's feedback on the manuscript and his opinion as to the value of this work. We believe the comments provided add useful clarification and improve the manuscript and hence have addressed them as requested unless otherwise stated.

Substantive Comment: My only general comment to the authors is that there are times in this paper where the conclusions drawn from the analysis are stated as fact or proven when the conclusions appear to be inferential in nature. My problem with this isn't that the inferences or conclusions drawn are wrong, rather, they simply appear to be overstated in terms of confidence. There are times when causality appears to be stated as proven by results when there is really not mathematical proof supporting this conclusion. For what it's worth, in most instances, the author's inferences are probably on the right track, but I think some wording changes to cite the true confidence in these inferences is needed to appropriately represent what can be inferred from the results of their analysis. These concerns are individually stated as minor comments below.

This is a fair observation, and we have made all attempts to clarify the uncertainty in each of the stated cases unless otherwise discussed, and further reviewed the manuscript for any other cases we could identify. To some degree, the inferential conclusions are needed to highlight some of the elements as the sources would be difficult to isolate and mostly exist as here-say or knowledge in personal communication.

[Editor's Note: A few comments initially labeled "minor" are included below, since they resulted in small but substantive revisions or clarifications.]

Regarding your Gaussian smoothing choices: From Fig. 2, your choice of σ looks appropriate in that it appears to capture the nature of the background histogram you provide for the types trends that you identified in your discussion. I noticed some interesting local extrema in the spring months in Fig. 2(a,c) and wondered if you had tried any lower σ values to incorporate any of those local peaks and valleys that don't show up on your hail days graphs. I guess my comment here in general is: Did you take a look at your data utilizing any other smoothing specifications or did you follow guidance primarily from previous research? Regardless, from the figures and information you've provided, it seems like your smoothing accurately captures the larger scale frequency trends you were modeling from your histogram. I simply wondered if there was any value in some of the smaller wavelength trends that are noted in the background histogram.

Undoubtedly there are gains to be had by considering different kernel bandwidths or alternate ways to capture the wavelength (e.g. Lu et al 2015), depending on the intended application. We explored a range of sigma values for this application, and settled on the chosen approach, which turned out to be similar to that used by Doswell (2007) for the pooled data we used. On smaller regional scales or when exploring other elements of the temporal, and for more localized data, there is definitely value at considering different kernel sigma values (small temporal or spatial scale variability is often masked by such a large

bandwidth), or potentially alternatives to the Gaussian approach (for example, a narrower seasonal peak might be better represented by a Epanechenov or other kernel). We intend to explore this suggestion in future research, and have clarified this within the manuscript.

[Figure 3] and graphs are very interesting and illuminating, I think. Did you consider comparing a regression equation for the period of record from 1955–1989 to the regression equations you provided? It might be interesting for the reader to see that the slope of the regression equation associated with >1.25" hail is very similar "before and after" 1990. The comparison of slope for the other two equations would probably stand out quite a bit.

This is an excellent suggestion that we overlooked, and have updated the figure to illustrate this useful point. We have included the respective regressions, all of which show significance other than that for 2 inches. It is interesting to note that ≥1.25" hail still sees an increase in slope across the 1990 discontinuity, though nowhere near as pronounced as for the smaller sizes. We have also updated the text to reflect this information.

Mentioning significant break points in the hail climatology is fine, but it would probably help to note what criteria the authors utilized to identify these break points. If the criteria is specified and does in fact correlate to 1966, 1995 and 1996, it seems highly speculative to attribute those break points with the establishment of the NSSFC, the SPC, and the rollout of the Doppler radar network. The authors could very well be right that these are the driving factors behind these break points, but there was no evidence provided that supported this. There does not seem, to me at least, to be any evidence that supports this conclusion. I don't have any problem with the authors speculating that these break points are related to these events, but some data to support this conclusion is recommended. Language that is utilized later in this section: "A possible explanation for trend in significant hailstones leading up to 2000 is that more recent reports may reflect an increasing proportion of observers…" This type of language seems more appropriate in that it conveys uncertainty while positing a reasonable explanation for the trend in data observed.

This is a good suggestion, it is true that there is insufficient certain evidence in the current paper to point to these root causes. We have clarified this language, added annotations in Fig. 4 to highlight the discussed break points. We would argue that the influence of the Doppler radar network is not speculative but commonly held and published—while the SPC establishment may have a co-related contribution to the Doppler rollout, the quadrupling of reports in the period coinciding with the rollout while not influencing the number of hail days supports this contention. As this has been identified in other literature (e.g. Verbout et al. 2006) and elsewhere, we are confident of the networks contribution to hail report growth (Schaefer et al. 2004).

Figure 11: I assume that your population bins were assigned as a tool in ArcGIS and then the data displayed as specified in the 4 panels of your figure. Granted, I don't have any experience with the mathematics involved in the assignment of data bins from the Jenks Natural Breaks algorithm, but the population density maps seem a bit suspect in cases. Assuming Amarillo, TX is depicted in the high density of hail reports seen in the middle of the image, why is there a pixel of very low population density intersecting the city of Amarillo? Knowing how quickly population density changes out in west Texas, it seems strange that a grid box containing some of Amarillo would represent a local minimum in the population-density map. I realize the point of the figure is to show that the rate of population density increase/change does not explain the increase in hail reports. I completely agree with your analysis, but the background population density map doesn't seem to make sense in this case.

As you suggest, this was an artifact of the population data used. The process was conducted in Python using shape files and intercensal county population data, with the population scale altered using Jenks Natural Breaks (which means that you don't get a color scale biased toward the largest values, which is a typical problem for chloropleth population maps). This point was also raised by reviewer B, and we have taken steps to address it by replacing county population with a higher resolution product for the period when it is available (c,d). The challenge was not making the figure too cluttered by showing higher-

resolution population and report changes, when the focus is on the reports, and we believe the new rendering with roads achieves this goal.

[Minor comments omitted...]

Second Review:

Recommendation: Accept with minor revisions.

General Comments: The authors satisfactorily addressed all of my comments and suggestions from the first round of reviews. In my opinion I think that their work is in a form that is ready to be published. My only remaining comments and suggestions are minor in nature and are along the lines of grammatical or aesthetic content. Once again, I think the authors have done an excellent job in identifying important biases and trends in the United States hail climatology that should be considered for anyone engaged in using these data for future research.

[Editor's note (manuscript editor): There are a couple instances in the paper (noted by two reviewers in P.2 and P.3) where "individual forecasters" are suggested to play a role in the irregularities of the hail database, noted by discontinuities along CWA boundaries.   I agree with the reviewers as I'm unsure how forecasters in their individual role would have any bearing on altering incoming hail reports or collecting hail reports for warning verification purposes.

During a severe weather event, it is fairly uniform from office to office that at least one forecaster/meteorologist is responsible for collecting ground-truth reports to support operations. Where this can vary from one office to another, and where you may see long-term trends or discrepancies between offices, are differences in: 1) local office policy (importance and aggressiveness of warning verification and reports); 2) utilizing proven or innovate collection methods versus a standard spotter database;, and 3) available workload and staffing strategies. Perhaps these concepts were what you were trying to convey, and I would encourage you to consider some of these contributing factors for the differences in the hail database related to NWS operations, or expand how you envision the role forecasters play in shaping the hail database.]