An Hourly Climatology of Operational MRMS MESH-Diagnosed Severe and Significant Hail with Comparisons to *Storm Data* Hail Reports

NATHAN A. WENDT[a] AND ISRAEL L. JIRAK[a]

[a] *NOAA/National Weather Service/Storm Prediction Center, Norman, Oklahoma*

(Manuscript received 8 September 2020, in final form 1 February 2021)

ABSTRACT: The Multi-Radar Multi-Sensor (MRMS) system generates an operational suite of derived products in the National Weather Service useful for real-time monitoring of severe convective weather. One such product generated by MRMS is the maximum estimated size of hail (MESH) that estimates hail size based on the radar reflectivity properties of a storm above the environmental 0°C level. The MRMS MESH product is commonly used across the National Weather Service (NWS), including the Storm Prediction Center (SPC), to diagnose the expected hail size in thunderstorms. Previous work has explored the relationship between the MRMS MESH product and severe hail (≥25.4 mm or 1 in.) reported at the ground. This work provides an hourly climatology of severe MRMS MESH across the contiguous United States from 2012 to 2019, including an analysis of how the MESH climatology differs from the severe hail reports climatology. Results suggest that the MESH can provide beneficial hail risk information in areas where population density is low. Evidence also shows that the MESH can provide potentially beneficial information about severe hail occurrence during the night in locations that are climatologically favored for upscale convective growth and elevated convection. These findings have important implications for the use of MESH as a verification dataset for SPC probabilistic hail forecasts as well as severe weather watch decisions in areas of higher hail risk but low population density.

KEYWORDS: Hail; Severe storms; Climatology; Radars/Radar observations; Forecast verification/skill; Operational forecasting

1. Introduction

Hail contributes to a substantial portion of the total insurance damages for crops and other property in a given year. During the past decade, thunderstorm-related damages have exceeded $30 billion (U.S. dollars) globally with hail estimated to account for $8–$14 billion of those total losses each year (Podlaha et al. 2020). Because of the potential societal impact of severe hail (≥25.4 mm or 1 in.), the ability to make skillful forecasts for the timing and location of severe hail is of great importance. The Storm Prediction Center (SPC) issues operational probabilistic forecasts for severe hail during the Day 1 (i.e., current day) and Day 2 periods and currently relies on local storm reports to verify these forecasts. To provide meaningful, skillful, and reliable probabilistic forecasts for any severe hazard on a subdaily time scale, a forecaster needs to be calibrated by baseline climatological values of risk on that same subdaily time scale. Some work has been done by Krocak (2017) and Krocak and Brooks (2018) that offers guidance on hourly climatological risk for hail and tornadoes, respectively. Those studies utilized severe hail and tornado reports from the National Climatic Data Center (NCDC) *Storm Data* publication. For hail, which is the focus of the present study, the *Storm Data* report database contains known issues that are well-described by several previous works (e.g., Hales 1993; Wyatt and Witt 1997; Davis and LaDue 2004; Jewell and Brimelow 2009; Allen and Tippett 2015; Blair et al. 2017). Of particular relevance to this study is the presence of estimated hail size by the public. Because hail is often estimated by comparison to some reference object of known size, the distribution of hail reports becomes quantized into sizes specific to the common reference objects (Doswell et al. 2005). The true hail size can be overestimated or underestimated with this approach (Allen et al. 2017). The lack of spatiotemporal coverage of human–issued reports is also a concern. In areas of low population density, hail may not be reported at all and, if it is, it may not represent the largest hail that occurred (Blair et al. 2011, 2017). All of these issues can pose problems when using the data for climatological and verification purposes.

One tool that forecasters at SPC, and the NWS more broadly, use in a situational awareness and nowcasting sense is the maximum estimated size of hail (MESH; Witt et al. 1998) product from the National Severe Storms Laboratory (NSSL) Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) product suite. The MESH is a radar-based hail-size estimation algorithm (described in section 2) that uses radar reflectivity within preferred temperature layers for hail growth to estimate hail size. While the original Witt et al. (1998) MESH algorithm is currently used in the operational MESH product, new formulations have been developed to try to improve hail size estimates (see Murillo and Homeyer 2019). The MESH product is useful to forecasters owing to the overall consistency, known biases, and spatiotemporal coverage in areas that receive very few severe hail reports. Leveraging these desirable qualities of the MESH, this study seeks to extend the work of Krocak and Brooks (2018), using the MESH as an estimate of severe hail occurrence to create an hourly climatology. With an hourly

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WAF-D-20-0158.s1.

Corresponding author: Nathan A. Wendt, nathan.wendt@noaa.gov

DOI: 10.1175/WAF-D-20-0158.1

climatology of MESH-diagnosed severe hail, questions can be answered on where and how the MESH-based severe hail climatology differs from the *Storm Data*–based severe hail climatology on a subdaily time scale. Of particular interest is how they differ during the night (i.e., between sunset and sunrise) when receiving a report of severe hail is hypothesized to be less likely, as suggested by Ashley et al. (2008) for nocturnal tornadoes and Bunkers et al. (2020) for hail and wind.

A major goal of this work is to provide the science that will ultimately support the SPC in fulfilling its mission. One aspect of that goal is to evaluate the utility of the MESH for use in verifying the probabilistic hail forecasts produced by the SPC. As the MESH is less prone to nonmeteorological artifacts that are possible with traditional storm reports (e.g., human availability at various times of day, population density affecting report frequency, etc.), it may be able to add value in some scenarios. Understanding where MESH estimates severe hail occurrence in a climatological context will further help to assess MESH being used as a forecast verification tool. A better understanding of the differences in the climatological risk indicated by MESH and *Storm Data* can also provide a forecaster useful information for watch and warning decisions in areas that are less likely to receive severe weather reports. As part of this process, this work will also investigate the new MESH formulations developed by Murillo and Homeyer (2019) and compare their performance with that of the original Witt et al. (1998) MESH algorithm. Furthermore, as alluded to earlier, there is a drive to provide probabilistic forecasts on shorter time scales as part of the Forecasting a Continuum of Environmental Threats (FACETs) initiative (Rothfusz et al. 2018). For the SPC, specifically, this means providing temporal forecasts and information for severe hail between the outlook and watch as well as between the watch and warning.

This paper is organized as follows: section 2 describes the data and quality control used. Section 3 describes the methods used to derive the hourly climatologies of MESH and *Storm Data* hail reports, section 4 presents the results of the analysis, and section 5 discusses the implications of the findings and potential operational applications.

2. Data

MESH is a radar-derived, gridded hail-size estimate based on an exponential fit to the severe hail index (SHI; Witt et al. 1998). The SHI is a weighted vertical integration of flux values of hail kinetic energy (Waldvogel et al. 1978; Federer et al. 1986). Values of hail kinetic energy are weighted by both reflectivity and temperature. Flux values are given increasing weight with reflectivity greater than 40 dBZ and are fully weighted at and above 50 dBZ. Thermally, flux values are given increasing value above the 0°C level and are fully weighted at or above the −20°C level. The integration continues through the depth of the storm scanned by radar so long as the minimum weighting thresholds are met. Caveats to the accuracy of the MESH are 1) it will provide more accurate estimates when multiple radars are sampling a storm (Ortega et al. 2005, 2006), 2) the algorithm was designed to be an overforecast such that 75% of all observed hail sizes will fall below the radar estimate (Witt et al. 1998), and 3) the algorithm was formulated using a very small sample of hail events from a limited geographic domain (only Oklahoma and Florida), none of which were smaller than 0.75 in. (≈19 mm) (Witt et al. 1998). The data are on a grid, with approximately 1-km spacing, that spans the entirety of the contiguous United States and is updated every 2 min. Details and additional references on the process of taking single-radar MESH data and placing them on a 3D Cartesian grid can be found in Smith et al. (2016). These grid-based data are in contrast to the cell-based data. Cell-based data track individual storms cells and create a vertical reflectivity profile that is then processed by the MESH algorithm. Cell-based data can come from both multiradar (Stumpf et al. 2002) and, more often, single radar (Witt et al. 1998) sources. As mentioned, multiradar grid-based MESH statistically provides better estimates, which is why they are used in this study. The SPC receives MRMS MESH data through operational data feeds from the National Centers for Environmental Prediction (NCEP). The MESH data within this study comes from MESH data archived from that feed. This analysis compares climatological characteristics between the MESH and severe hail reports from the *Storm Data* publication over the same 2012–2019 period. This period was selected as both archived MESH data and *Storm Data* reports were available. Any reference to reports is synonymous with hail reports from the *Storm Data* publication.

3. Methods

a. Data quality control

To ensure that MESH data were not spurious radar artifacts, a quality control procedure similar to Wendt et al. (2016) and Melick et al. (2014) was used. Hourly MESH data were used in this step to ensure that MESH output was associated with thunderstorms through an hourly lightning quality control step. Quality-controlled National Lightning Detection Network (NLDN) cloud-to-ground lightning data were used to determine if MESH pixels are associated with a thunderstorm. NLDN data were aggregated into hourly files and time matched with MESH data from the same hour. Those MESH pixels that fall within 40 km of a detected flash during the same hour are included in this study. MESH values above 127 mm (5 in.) were removed as there is evidence to support those values being very rare and likely spurious (J. L. Cintineo et al. 2016, personal communication; Blair et al. 2011).

b. Hourly climatology creation

Creation of hourly climatological estimates of MESH-diagnosed severe hail and *Storm Data* severe hail reports largely followed the methods of Krocak and Brooks (2018). As this study is concerned with hourly severe hail occurrence, the 60-min maximum MESH data at the top of each hour were used. Use of this data allows for examination of whether a grid cell met or exceeded severe MESH thresholds at least once during the hour. Hourly maximum grids were similarly computed for *Storm Data* severe hail reports. To identify severe hail in the MESH data, a 29 mm (1.14 in.) threshold was chosen based on previous work by Cintineo et al. (2012) and

Murillo and Homeyer (2019) showing this value to be the most skillful in discriminating severe from subsevere hail. This is nearly identical to the threshold of 30 mm (1.19 in.) identified by Ortega (2018). Note that both thresholds are higher than the 25.4 mm (1 in.) threshold used by NWS forecast offices and the SPC to define severe hail. Severe MESH and *Storm Data* severe hail reports were then regridded to an 80-km Lambert Conformal Conic (LCC) grid, which is similar to the grid that SPC uses to verify its outlooks (i.e., probability of severe hail within 25 mi of a point). The interpolation used was a maximum nearest-neighbor method where, for each grid point and each hour, the maximum hail size is mapped to the nearest 80-km LCC point. While detrending of the hail report database was done in Krocak (2017), the short time series of data (i.e., 8 years) used in this study did not show a significant trend. In terms of significantly severe hail (≥50.8 mm or 2 in.), both Cintineo et al. (2012) and Ortega (2018) did not find MESH thresholds that showed predictive skill. Despite that caveat, it is still instructive to investigate where and when MESH suggests significant hail occurs and how that compares to *Storm Data* reports. As such, this work does include an hourly analysis of significantly severe hail as diagnosed by MESH. The only deviation from the previously described methods is to use a threshold of 50.8 mm (2 in.) in the MESH grids for significantly severe hail.

Using the hourly 80-km LCC grid data, yearly grids were created by concatenating all hourly data together into one array with an extra day inserted for non–leap years. The grid points that were equal to or above the defined severe thresholds were then made into a binary grid where values of one represent a severe report, with zeroes assigned at all other grid points.

Though the 80-km LCC grid is already coarse, additional spatial smoothing of the binary grids was done to account for spatial uncertainty inherent in SPC outlooks (Brooks et al. 2003; Hitchens and Brooks 2014; Krocak and Brooks 2018). The equation for this step is as follows:

$$P = \sum_{n=1}^{N} \frac{1}{2\pi\sigma^2} e^{-d^2/2\sigma^2}, \tag{1}$$

where P is the kernel density estimate (KDE) of severe hail probability, N is the total number of grid boxes with severe hail events, d is the distance from grid point to the severe hail location, and σ is the smoothing parameter for the Gaussian filter. A σ value of 120 km was used as it produces fields that are qualitatively similar to SPC outlooks. Next, each hourly grid was then smoothed in time using

$$P = \sum_{n=1}^{N} \frac{1}{2\pi\sigma^2} e^{-t^2/2\sigma^2}, \tag{2}$$

which is exactly as in Eq. (1), but now distance is measured in time t. Smoothing in time was done in two different ways: 1) using σ = 15 days and 2) σ = 2 h. These temporal smoothing parameters are consistent with Krocak and Brooks (2018). The 15-day smoothing parameter would smooth across the same hour from each day (e.g., all 1800 UTC) to preserve the seasonal cycle. The 2-h smoothing parameter smooths adjacent hours to preserve the diurnal cycle. After all the smoothing, the yearly grids with all hours for the year were averaged to produce a KDE of severe hail hours per year. While the amount of processing of the data may appear excessive, each step is important to producing an analysis that is appropriate for use in providing probabilistic guidance to forecasters as well as verifying probabilistic severe hail outlooks. To see what this process produces, animations of the hourly estimated hail probabilities of severe and significant hail using the MESH data are available in the online supplemental material.

c. *Daytime and nighttime severe hail risk*

To analyze the severe hail events during the day and night separately, an hourly binary grid representing daytime (sunrise to sunset) and nighttime (sunset to sunrise) was created for the contiguous United States. Sunrise and sunset calculations were handled by the Python astronomical calculations package, "Skyfield" (v1.16; Rhodes 2019). Using this grid as a mask, calculations on hail occurrence during daytime and nighttime hours were performed.

4. Results

a. *Comparing MESH to hail reports*

Climatologies for both MESH and *Storm Data* severe hail were computed for 2012–2019, the results of which are shown in Figs. 1a and 1b. The figures show estimates of hail hours per year (i.e., the number of hours during a given year an 80-km grid cell experiences severe hail). There are two main differences between the MESH and *Storm Data* severe hail estimates. The first is the overall difference in magnitudes of hail frequency. Even in areas of the plains where the population density is relatively high, the estimated severe hail hours from the MESH can be 2–4 times greater than those estimated from *Storm Data*. In areas where population density is low, the differences can be even higher. MESH significant hail hours are similarly 2–4 times larger than the estimates from significant hail reports (Fig. 2c). The second is the MESH analysis shows a western extension of estimated severe hail hours into areas of low population density farther south–near the Big Bend region of Texas—and farther west—into more of the High Plains, Raton Basin, and along and south of the Mogollon Rim in Arizona. The differences are further highlighted when taking a difference field between MESH and *Storm Data* (Fig. 1c). For this analysis, the MESH estimates of severe hail hours per year are higher at all CONUS locations compared to estimates from reports. The same comparison was done for significant hail and is shown in Figs. 2a and 2b. The largest differences are within western portions of the plains into the High Plains and, especially, in the vicinity of the Black Hills of South Dakota and within the broader High Plains region. More subtle differences exist within Mississippi, Alabama, Georgia, and central Florida. Differences in the southeastern United States are likely diurnally driven. In those areas, large buoyancy during the day can support updrafts capable of producing hail aloft that will then be detected by the MESH algorithm. The lack of instability in the midtroposphere as well as a low-level environment conducive to melting hail likely explains why values are lower in the reports database (Cintineo et al. 2012; Murillo et al. 2020).

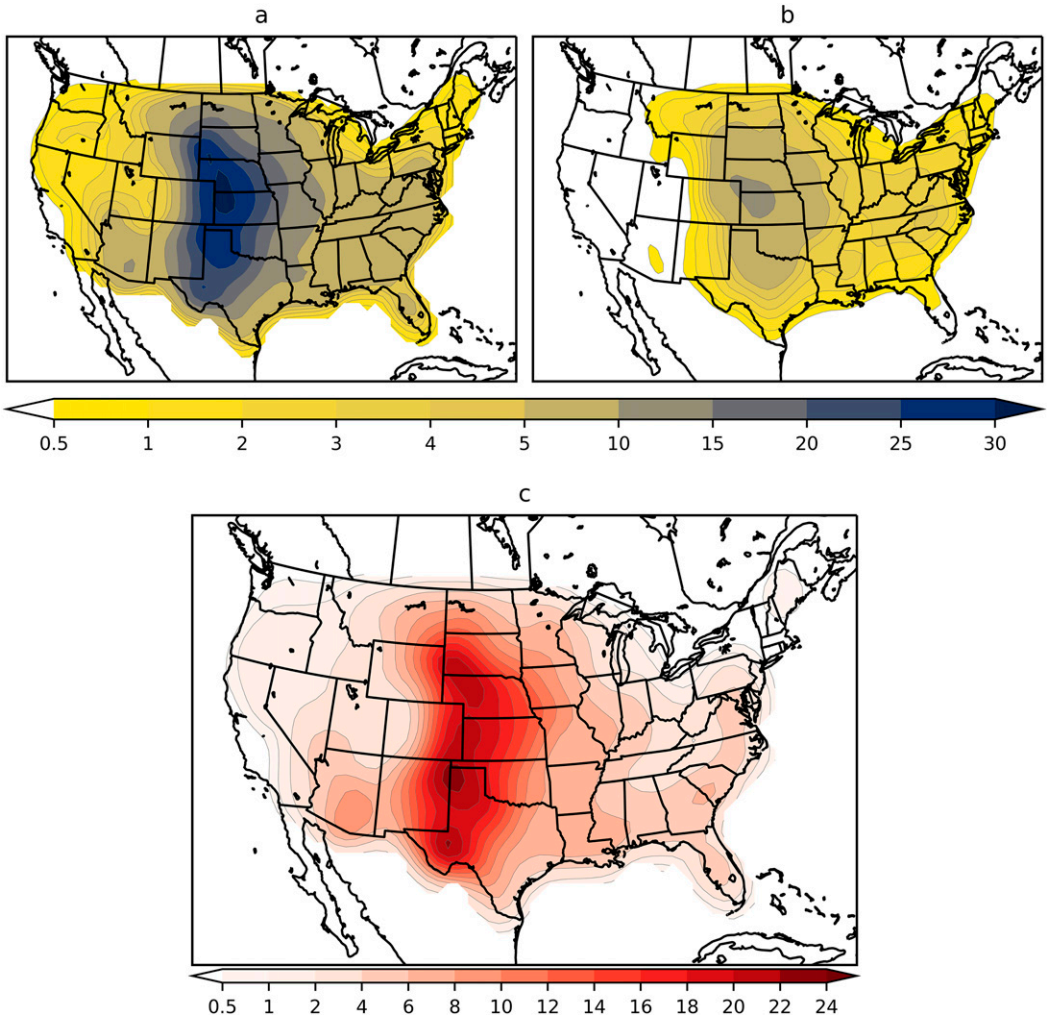

FIG. 1. Estimated average severe hail hours per year using (a) MESH and (b) *Storm Data*. (c) The difference between MESH and *Storm Data* [i.e., (a) minus (b)].

As previously discussed, hail may be underreported in some places. This needs to be kept in mind when comparing the MESH to *Storm Data*. While there are methods for correcting report bias (e.g., Potvin et al. 2019), these corrections were not used in this study as the raw difference between MESH and reports were of greatest interest. A comparison between the average hail days from MESH and *Storm Data* shows some very stark regional differences (Fig. 1c). The greatest differences occur across the High Plains into the central and southern plains. MESH estimates on the order of 20 more hail hours per year across much of the High Plains than what is recorded in *Storm Data*. This result suggests that severe hail is occurring more frequently and farther west than what is reported. Even with the MESH likely overestimating severe hail to some degree, it is still plausible that severe hail is going underreported due to relatively low population density in the High Plains. There is a less pronounced relative maximum in average severe hail day differences along and south of the Mogollon Rim in central Arizona. This is another situation in which low population density likely explains why MESH estimates a larger number of severe hail hours. Relative minima in severe hail hour differences exist within much of the Intermountain West (i.e., areas bounded by the Rockies, Sierra Nevada, and Cascades), the West Coast, the Appalachians, and the Northeast. Lack of radar coverage (on account of beam blockage or widely spaced radars) and low population density account for what is seen in Intermountain West and, to a lesser degree, within parts of the Appalachians. The West Coast, Appalachians, and Northeast, however, all tend to have environments that are less supportive of severe hail (Allen et al. 2015). Differences in significant hail between MESH and *Storm Data* are similar to severe hail in that the same areas of the High Plains and plains are where the larger values exist (Fig. 2c). Within the Black Hills and areas to the southwest is where the maximum difference occurs. This signal does not appear to be an artifact of the MESH as the grid cell containing the Black Hills had the most significant hail reports during the study period.

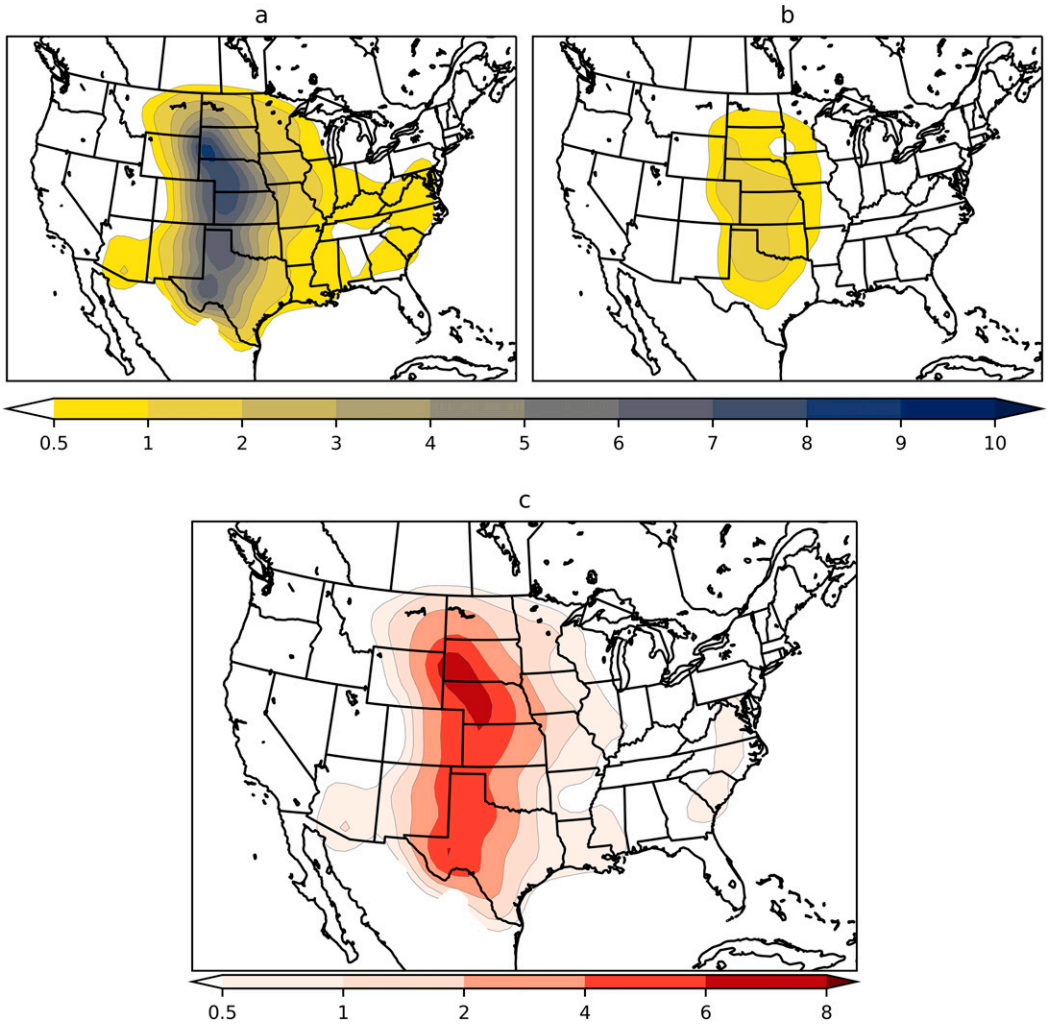

FIG. 2. As in Fig. 1, but for significantly severe hail.

Comparisons between MESH and *Storm Data* reports were made for each hour, which is seen in Fig. 3. Only a few hours are shown as a summary. Here, the MESH severe hail hours estimates are plotted in filled contours. Red contours represent the report-based hail estimates that share the same contour levels as the MESH. Overall, the MESH and report-based severe hail estimates highlight very similar geographical areas throughout the day. However, the MESH-based data show greater risk in Florida and Arizona during the afternoon whereas reports indicate very little risk. For the areas of highest hail risk, there is not an obvious temporal difference in where MESH and hail reports show maximum severe hail risk. For significant hail (not shown), the spatial pattern and temporal evolution are quite similar. The biggest difference in that case is that probabilities are much more concentrated within the High Plains and plains regions than for severe hail.

Previous work by Cintineo et al. (2012) on a daily MESH-climatology using the methods of Brooks et al. (2003) showed similar spatial patterns across the plains and portions of the western United States when compared to the results in this study. They also compared their MESH-based data with reports. Within the plains, the analysis produced similar results to this study. However, within some areas east of the Mississippi River, they found that severe reports were more common than MESH-based severe events. This regional difference is in contrast to what was found in this study and was likely due to their use of the 19 mm (≈0.75 in.) threshold for severe reports. That threshold choice was out of necessity as they analyzed data before the threshold was officially increased to 25.4 mm (1 in.). Due to the low predictive skill of MESH for significant hail, Cintineo et al. (2012) did not include a significant hail MESH climatology.

b. *Comparing day and night*

With an hourly climatology, comparisons between daytime and nighttime hours can be made. The difference between day and night severe hail hours per year for MESH (Fig. 4a) and *Storm Data* (Fig. 4b) were computed. Because of temporal smoothing as well as day and night duration being unequal, conclusions about how MESH or *Storm Data* differ between

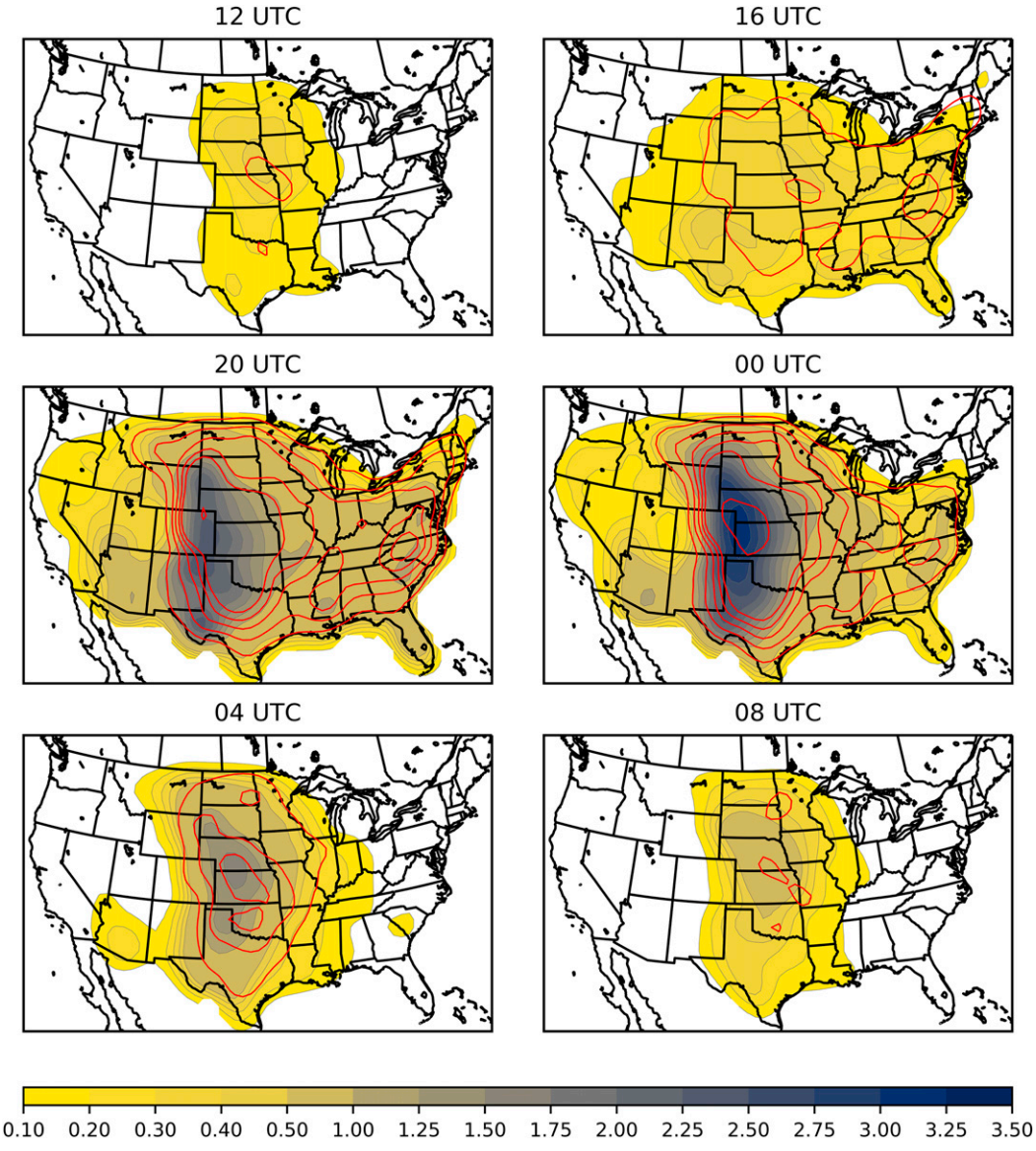

FIG. 3. A comparison of average severe hail hours for 1200, 1600, 2000, 0000, 0400, and 0800 UTC between MESH (color fill) and *Storm Data* (red contours). All units are severe hail hours per year for the labeled hour, and the contour levels are the same for both datasets.

both periods should be made with some caution. However, the main focus of this comparison is to see how MESH and *Storm Data* differ with each other. Strong differences are evident in the MESH diurnal climatology. During the day, more MESH-diagnosed severe hail occurs from the Texas Big Bend into the High Plains and southern Rockies of New Mexico and Colorado, where thunderstorm initiation over the higher terrain is a regular occurrence. For hail reports, there is a tendency for more reports to occur during the day over the contiguous United States with local maxima in the central plains and portions of the Piedmont region of the Carolinas and Virginia. East of the Mississippi River, MESH and reports share a relatively similar spatial pattern, though MESH does show slightly higher values in Mississippi, Alabama, Georgia, and central Florida. During the night, however, the MESH denotes more severe hail hours occurring in the plains, particularly in the Platte River Valley into north-central Kansas. Aside from a weak signal during the daytime in parts of the southern High Plains, significant hail does not show a substantial difference between daytime and nighttime in either dataset (not shown).

Comparing severe hail during the daytime versus the nighttime (Fig. 4) shows interesting patterns in both the MESH and *Storm Data* datasets. The most prominent difference is the slightly greater number of hail hours during the night in the MESH within the central and northern plains. A question to

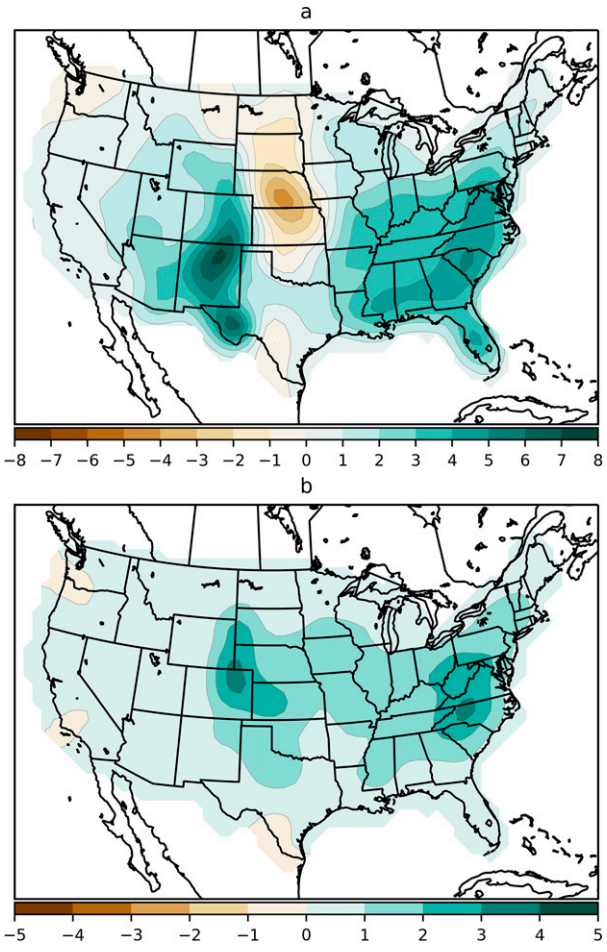

FIG. 4. Difference in average severe hail hours per year between day and night (i.e., day minus night) for (a) MESH and (b) *Storm Data*.

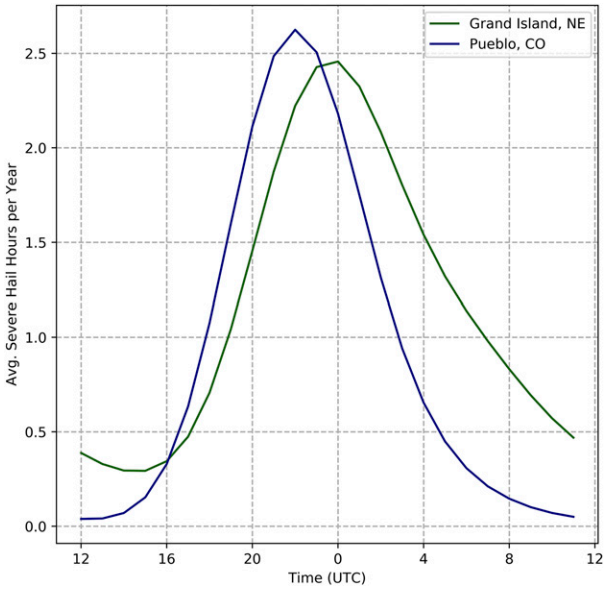

FIG. 5. Average severe hail hours per year calculated for each hour at the grid points nearest Grand Island, Nebraska (dark green), and Pueblo, Colorado (navy).

consider when trying to explain this difference is: when does severe hail occur during the night? To answer this question the KDE probabilities for each individual hour (i.e., all 1200 UTC during the year, and so on) can be summed to get the average hail hours per year for that individual hour. This is done for two locations: Pueblo, Colorado, and Grand Island, Nebraska (Fig. 5). The locations of these cities can be found in Fig. 6. Grand Island is chosen as it is within the area of the central plains that sees the most hail during the night according to MESH. Pueblo is chosen as it sees more hail during the afternoon according to MESH and will provide a good comparison. While both locations have similar magnitudes for the peak severe hail risk during the afternoon, there are marked differences from late afternoon into the later periods of the night. In particular, Grand Island, NE, shows a more steady drop-off after about 0100 UTC as opposed to the sharper drop-off seen at Pueblo. It is this decreasing but elevated (as compared to surrounding areas) severe hail risk across the central and northern plains that leads to a slightly greater risk occurring during the night in the MESH data. A secondary feature is that the highest risk during the night at Grand Island, occurs immediately after sunset and declines through the rest of the night. To be sure that the temporal smoothing applied to the data was not responsible for the trends shown here, a similar analysis was done to the raw MESH data. While not shown here, that analysis also exhibited the same temporal trends at both locations.

An important climatological feature of the central United States is the nocturnal precipitation maximum. Thunderstorms often develop along the Rocky Mountains and progress eastward into the plains with time, typically growing upscale with aid from the Great Plains low-level jet. This phenomenon has been well-documented in the literature (e.g., Easterling and Robinson 1985; Riley et al. 1987; Carbone et al. 2002; Carbone and Tuttle 2008). This particular convective evolution is the reason why the central and northern Plains regions are commonly impacted by nocturnal mesoscale convective systems (MCS; Haberlie and Ashley 2019; Cheeks et al. 2020). MCSs can produce severe weather, though hail is more common early in their life cycle (Maddox et al. 1982; Houze et al. 1990). It is significant that the MESH-based analysis produces a signal for severe hail during the night in the same locations where cellular storm modes often grow upscale into MCSs. This signal is not something that is seen in the *Storm Data* reports—rather, a greater number of hail hours occur during the day across nearly the entire domain.

Another contributing factor to this signal is the secondary convective initiation peak in the central plains during the night (i.e., commonly between 0400 and 0800 UTC) (Reif and Bluestein 2017; Stelten and Gallus 2017). These storms are largely driven by warm air advection associated with the nocturnal increase in the Great Plains low-level jet (Mead and Thompson 2011). Reif and Bluestein (2017) further showed that these storms tend to produce more hail reports than wind

FIG. 6. Reference for cities used in this analysis.

reports which is also consistent with the analysis by Bunkers et al. (2020). Both Figs. 4 and 5 indicate that the MESH-based analysis is picking up on these signals. A similar signal for higher hail occurrence during the night, present in both MESH and *Storm Data*, is seen in the middle Rio Grande Valley region of Texas. This is a location that is frequently impacted by either discrete supercells or MCSs moving off the higher terrain across the Mexican border shortly after sunset and into the night (Edwards 2006; Haberlie and Ashley 2019; Cheeks et al. 2020). Particularly with the central and northern plains, having the MESH data provides additional information during a time of day when hail reports decrease in frequency.

Significant hail exhibited very little difference between daytime and nighttime in both the MESH and *Storm Data* data (not shown). The only exception was a weak signal for greater daytime risk in portions of the southern High Plains. While it remains possible that significant hail could be under reported during the night, boundary layer cooling and associated changes in the environment will generally decrease the potential for significant hail after sunset. Within the study period, approximately 30% of significant hail reports occurred during the nighttime while MESH showed around 40%.

With regard to the discrepancies between MESH and hail reports during the night in the plains, NWS warning verification practices could potentially lead to underreported severe hail during the night as only one report is required to verify each warning (Amburn and Wolf 1997; Blair et al. 2011; Bunkers et al. 2020). Human observers are increasingly less likely to be available to observe severe hail after sunset. With the advent of mesonets across several states in the Great Plains, observing severe wind gusts with relatively dense networks of automated sensors is not going to be affected by time of day. While there are meteorological reasons for a decrease in hail potential during the night, as alluded to earlier, this factor also plays a role. The MESH, as with automated sensors for wind, is also not going to be affected by time of day. Given the work by Reif and Bluestein (2017) and Stelten and Gallus (2017), the MESH-based analysis is picking up on an observed phenomenon better than reports. Even if the MESH is predicting too much severe hail during the night, it is still providing useful information above what is present with reports alone.

c. *Comparison to updated MESH formulations*

Murillo and Homeyer (2019) developed updated MESH equations based on a larger sample of hail reports from a substantially larger geographic domain than what were used in Witt et al. (1998) ($MESH_{Witt}$). They produced two new power-law relationships between SHI and MESH: one a fit to the 75th-percentile of the sampled hail size distribution ($MESH_{75}$) and the other a fit to the 95th-percentile of the distribution

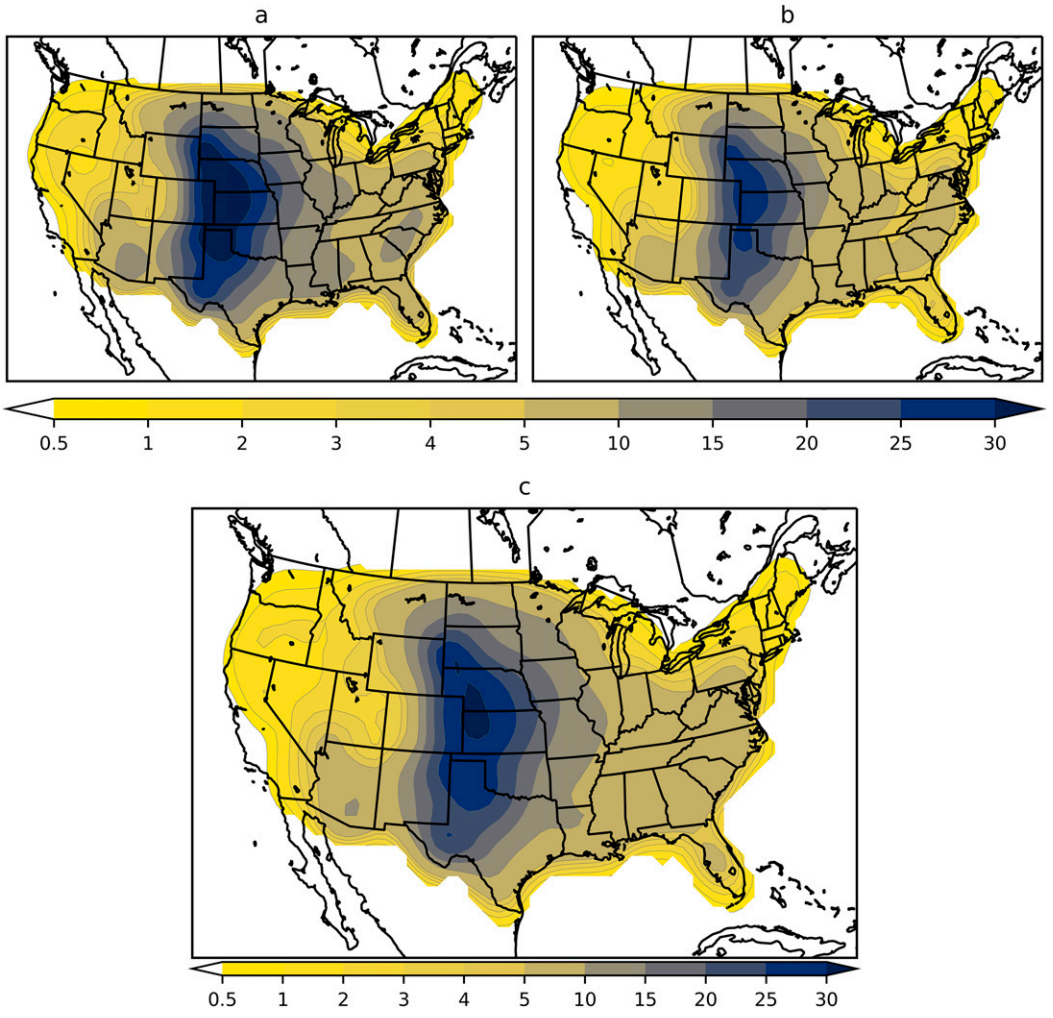

FIG. 7. Estimated average severe hail hours per year using (a) $MESH_{75}$, (b) $MESH_{95}$, and (c) $MESH_{Witt}$. The thresholds for severe hail used are 40, 64, and 29 mm, respectively.

($MESH_{95}$). Based on their analysis, the $MESH_{95}$ showed the best improvement over $MESH_{Witt}$ when comparing observed hail sizes with radar-based estimates. To compare the original MESH with the new formulations, the same procedure in section 3b was applied to $MESH_{75}$ and $MESH_{95}$. The thresholds for severe and significantly severe hail for the new MESH equations come from the analysis done by Murillo and Homeyer (2019). They found $MESH_{75}$ to have the best performance at 40 mm (47 mm) and $MESH_{95}$ at 64 mm (83 mm) for severe (significant) hail. It should be noted that these severe hail thresholds are substantially higher than those found by Cintineo et al. (2012) and Ortega (2018) for $MESH_{Witt}$. Figures 7 and 8 show the results of the analysis for severe and significantly severe hail, respectively. For all MESH formulations, for both severe and significantly severe hail, the same geographical locations are highlighted. This is not surprising as the formulations are all power-law relationships that have differing coefficients. The general trend is that the $MESH_{95}$ falls between $MESH_{75}$ and $MESH_{Witt}$ in terms of the magnitude of hail risk. The $MESH_{75}$ appears to be the outlier as the $MESH_{95}$ and $MESH_{Witt}$ are relatively similar to one another.

The performance in identifying severe and significantly severe hail is important when determining what, if any, MESH formulation should be used to verify forecast. Murillo and Homeyer (2019) calculated performance statistics for each MESH formulation and found no significant difference between each in terms of critical success index (CSI) aside from the threshold that produced the maximum CSI. For this work, a similar analysis was done using 2012–2019 hail reports data and verifying on the SPC 80 km LCC grid. This differs from the Murillo and Homeyer (2019) analysis in that there is no storm object tracking and their work used hail reports starting in 2010. Figures 9 and 10 show performance diagrams (see Roebber 2009) for severe and significantly severe hail, respectively. For reference, POD refers to probability of detection and FAR refers to false alarm ratio. As in Murillo and Homeyer (2019), all formulations do not vary significantly in

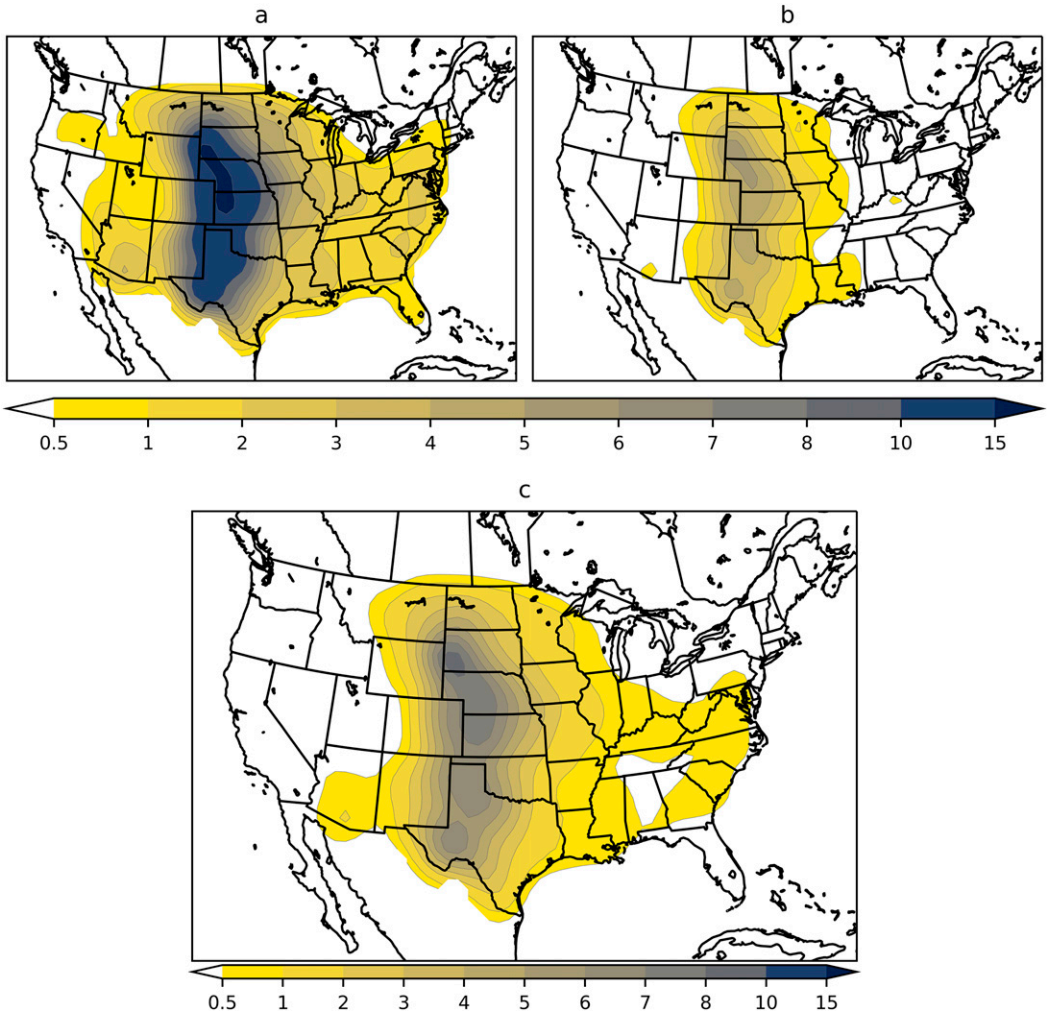

FIG. 8. As in Fig. 7, but for significantly severe hail. The thresholds for severe hail used are 47, 83, and 50.8 mm, respectively.

CSI where they overlap. The primary difference is that the $MESH_{95}$ is skewed toward higher POD at the same thresholds. Also of note is that $MESH_{Witt}$ has higher CSI than both $MESH_{75}$ and $MESH_{95}$ at the exact 25.4 mm (1 in.) and 50.8 mm (2 in.) thresholds. Due to the fact that $MESH_{95}$ simply has different power-law coefficients than $MESH_{75}$ and $MESH_{Witt}$, it is likely that adding additional, higher thresholds would lead to the $MESH_{95}$ having similar CSI to the other formulations at the cost of lower POD. As expected, CSI for all formulations is quite low for significantly severe hail. This has been the case in similar studies (Cintineo et al. 2012), though Murillo and Homeyer (2019) showed higher CSI using a storm-object-based verification approach.

While each formulation can perform similarly when choosing thresholds that maximize CSI, the question then becomes how reliable each MESH formulation is when predicting hail size. To assess reliability, a similar procedure to Wilson et al. (2009) was used where MESH values were compared to the averaged observed hail size that occurred at that particular MESH value. It should be noted that Wilson et al. (2009) used Severe Hazards Analysis and Verification Experiment (SHAVE) data. The SHAVE project provided high spatial and temporal resolution hail observations (see Ortega et al. 2009). This analysis bins values of MESH, on its native 1-km grid, into 1-mm increments. Hail reports are placed on the same grid and the average observed size at each MESH increment is calculated. Figure 11 shows the results of this analysis for each MESH formulation. What can be seen is that there are differences between the updated MESH formulations and the original $MESH_{Witt}$ near the important 25.4 mm (1 in.) threshold. At that threshold, both $MESH_{75}$ and $MESH_{95}$ more reliably predict hail size than $MESH_{Witt}$, but $MESH_{75}$ and $MESH_{95}$ are more likely to overestimate hail size between 25.4 mm (1 in.) and 50.8 mm (2 in.) than $MESH_{Witt}$. What is also interesting is how $MESH_{Witt}$ and $MESH_{75}$ tend to more reliably predict hail sizes near the other important threshold of 50.8 mm (2 in.). As you approach giant hail (101.6 mm or 4 in.) all formulations become less reliable, though small sample size and non-Rayleigh scattering of the radar beam also become issues.

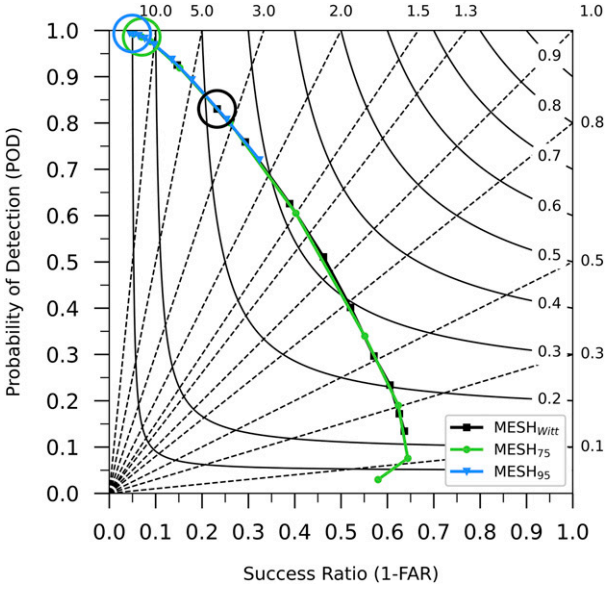

FIG. 9. Performance diagram for $MESH_{75}$ (green), $MESH_{95}$ (blue), and $MESH_{Witt}$ (black). The thresholds used for severe hail are 19.05, 25.4, 29, 35, 40, 45, 50.8, 55, 60, and 64 mm. The open circles highlight where the 25.4 mm (1 in.) threshold is for each similarly colored line. Dashed lines indicate bias scores that are labeled on the outward extensions of the lines. Solid lines indicate CSI which are labeled on the contours.

5. Discussion

While other MESH climatologies have been completed (e.g., Cintineo et al. 2012; Murillo et al. 2020), this is the first to include an investigation as to what the climatological hail risk is

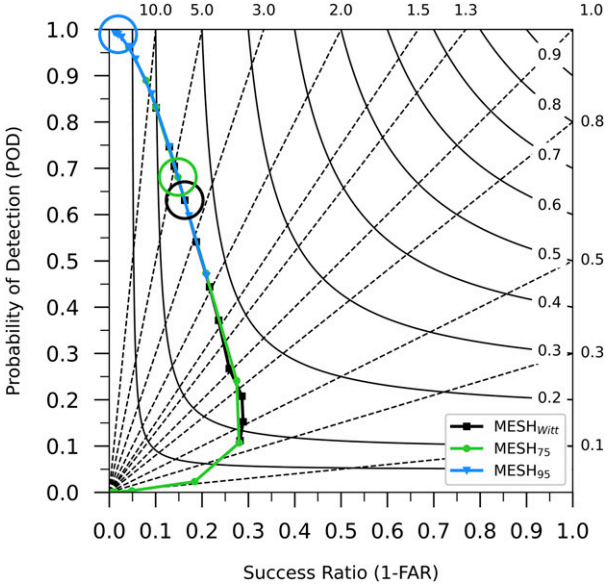

FIG. 10. As in Fig. 9, but for significantly severe hail. The thresholds used for significantly severe hail are 40, 45, 47, 50.8, 55, 60, 64, 70, 75, 80, and 84 mm. The open circles highlight where the 50.8 mm (2 in.) threshold is for each similarly colored line.

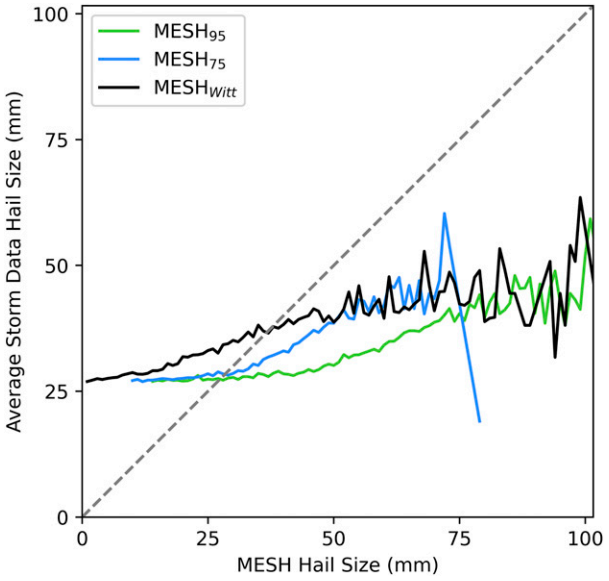

FIG. 11. Reliability diagram for $MESH_{75}$ (green), $MESH_{95}$ (blue), and $MESH_{Witt}$ (black). The gray dashed line represents perfect reliability.

on subdaily time scales. This allows for a similarly novel comparison as to the hail risk during the day and night. Risk for severe hail is certainly higher in the MESH-based climatology than one based on reports. Given that the design of the MESH algorithm is to increase the probability of detection of severe hail (Witt et al. 1998), it is not surprising that a climatology based on these data show higher estimates than that of one based on reports. The true risk for severe hail likely resides in between the estimates based on reports and MESH, which is why additional work to collect high-quality, high-resolution hail observations will be critical to improving our remotely sensed estimates of hail size. Products like the MESH, as this study has shown, can provide important value in areas with low population density as well as during the night. The ability for the MESH to do this, as well as avoid some of the issues with human reports, make it a valuable tool in forecast operations in terms of nowcasting, understanding baseline hail risk, and verification.

Potential operational applications

This work presents evidence that the MESH product has the ability to highlight where severe hail is occurring that may be missed by human observers. The greatest value from MESH is in locations with low population density and during the night. While the operational formulation of the MESH is still subject to certain caveats (see section 2), the spatial and the temporal coverage offers significant benefits over human reports. With the nonmeteorological artifacts noted in this work, the use of the MESH product to verify SPC probabilistic hail forecasts may be preferable to the public reports currently used to verify those hail forecasts. SPC forecasts are based on where storms will form in environments supportive of severe hail, which often includes areas of low population density. Using the

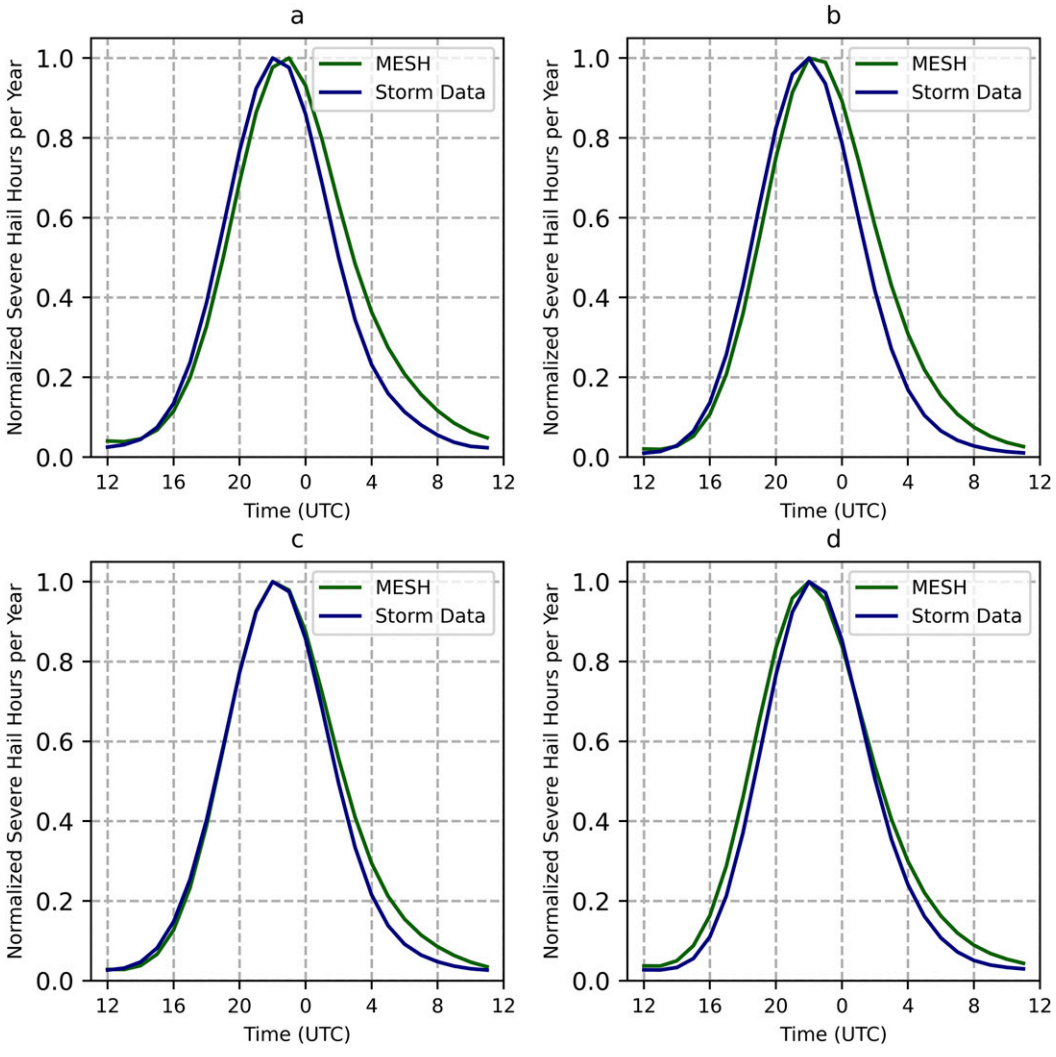

FIG. 12. Normalized severe hail hours per year for each hour at the grid point containing (a) Gillette, Wyoming; (b) Limon, Colorado; (c) Springer, New Mexico; and (d) Fort Stockton, Texas.

MESH to verify these forecasts would provide a more complete picture of where severe hail has likely occurred. Furthermore, the MESH algorithm can be improved upon. Such improvements are especially necessary for significant hail.

Given the performance and reliability results in section 4c, serious consideration has to be given to which MESH formulation to choose when using the data for climatological and verification purposes. When the appropriate thresholds are selected, both $MESH_{Witt}$ and $MESH_{95}$ show similar and reasonable climatological hail risk within the CONUS as well as similar skill in identifying severe and significantly severe hail (Figs. 9, 10, respectively). Though it performs similarly, the $MESH_{95}$ requires a greater degree of bias correction to do so than $MESH_{Witt}$. The main argument in favor of using an updated MESH formulation comes from its improved reliability around the severe hail size threshold (25.4 mm; Fig. 11). Considering that the goal is to accurately depict climatological hail risk and where severe hail is likely to have occurred, using the updated MESH formulations would be justified. The situation is less clear with significantly severe hail, however, as no one formulation is as reliable as with severe hail. From this analysis, though, the $MESH_{Witt}$ formulation is going to be the optimal, if imperfect, choice. What will ultimately be needed to improve hail size estimate algorithms are higher-resolution, comprehensive hail-fall datasets such as from SHAVE (Ortega et al. 2009) and the Hail Spatial and Temporal Observing Network Effort (HailSTONE) (Blair et al. 2017). SHAVE, in particular, has already been used to show the utility and deficiencies of MESH (see Wilson et al. 2009; Cintineo et al. 2012; Ortega 2018). Similar analysis with high-resolution observed hail data should be done with the MESH formulations from Murillo and Homeyer (2019) to verify the results within this analysis.

From Figs. 1 and 13, there is a clear westward shift in the probability of severe hail depicted by the MESH data.

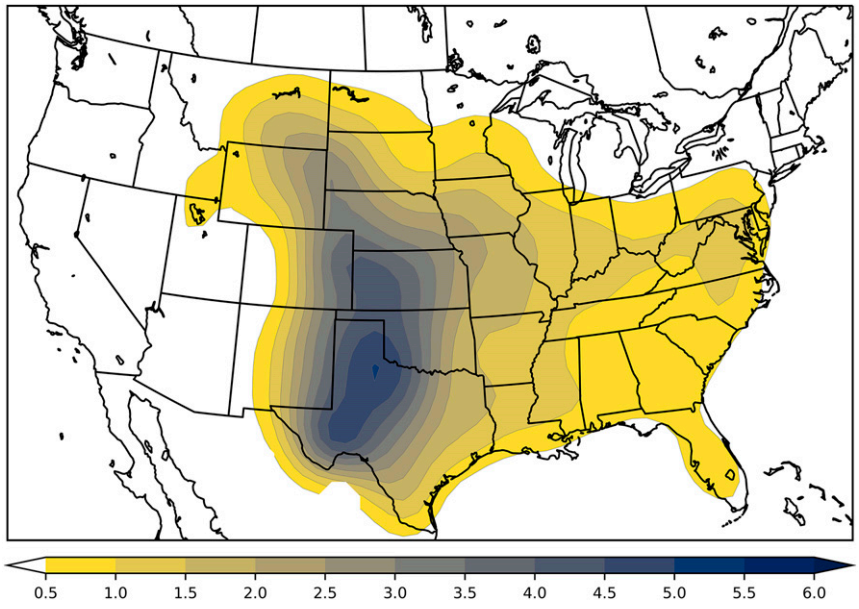

FIG. 13. KDE of MESH-diagnosed severe hail probability for the 4-h period ending at 0000 UTC 15 May.

This same signal is much less pronounced for significant hail (Fig. 2). The largest discrepancy between the severe MESH and reports is within the southern High Plains into the Texas Big Bend. Lack of population density is the most likely reason for the differences in this case. The central High Plains and parts of the northern High Plains still see discrepancies, though not quite as pronounced. Figure 1 shows that from Denver, Colorado to Cheyenne, Wyoming, there is greater agreement between MESH and *Storm Data*, where population density is greater. Elsewhere in the central and northern High Plains there is less agreement. Based on the spatial trends in MESH versus human reports shown herein, a forecaster has to weigh the meteorological threat with the potential of not being able to verify a warning or a watch with human reports in sparsely populated areas. However, use of MESH to verify these forecasts limits the uncertainty in human reporting and allows the forecaster to focus more on the threats posed by the weather, while still making a forecast that can be evaluated quantitatively.

A natural question to ask after seeing these results is: if there is greater risk farther west in MESH, is there also a difference in timing of that risk? This is an important question as the answer would have implications for watch and warning timing. To answer that question, a few grid points within the High Plains region were selected to compare the average daily hail risk between MESH and *Storm Data*. The process is nearly identical as to the one used in Fig. 5, but the data are normalized by their respective maximum daily risk values. The normalization is necessary as the goal is to compare the relative timing of when the greatest hail risk occurs. Figure 12 shows these analyses from locations roughly along a north–south transect in the High Plains spanning Wyoming to Texas. The locations of the chosen cities can be found in Fig. 6. Differences in timing were negligible for all locations with only Gillette, Wyoming, (Fig. 12a) showing a one hour offset between the datasets. Overall, the curves showed similar daily cycles with MESH tending to remain higher during the night in some locations. All this suggests that a timing difference in peak hail risk does not accompany the westward shift in severe hail risk seen in the MESH data.

As mentioned earlier, a forecaster will benefit from having knowledge of the baseline climatological risk of the hazard they are predicting. The probabilities for severe hail based on MESH in this work can provide this context to SPC forecasters when creating outlook probabilities for severe hail. Of course, having hourly probabilities will not be particularly useful when forecasting for a 24-h period. Krocak and Brooks (2020) found that more than 95% of severe weather reports (hail, wind, and tornado) within 40 km of a point occur within a 4-h window during a given convective day (1200–1200 UTC). Applying their methods to the MESH data, over 92% of severe MESH occurs within the same 4-h period. The reduction in the number of events captured is most likely related to only including hail events as well as the MESH data being hourly maximum values and potentially missing multiple severe events in one hour. Given these results, a 24-h probabilistic hail outlook can, by proxy, be represented by the maximum 4-h probability within a given convective day. For each hour in the quality-controlled MESH dataset, the maximum MESH values are taken for the 4-h period that ends with that particular hour. From here, the same process as described in section 3b is carried out and the 4-h period with the maximum probability can be extracted and used as a proxy for the full 24-h period. An example of the type of guidance a forecaster could use is in Fig. 13. Based on Fig. 13, a forecaster expecting storms from southwest Oklahoma into western North Texas during the

afternoon of 14 May should consider a 5% probability of severe hail as the baseline. Comparatively, the same analysis using *Storm Data* reports only suggested around a 3% probability (not shown). For the same day, a forecast of 5% severe hail probability for areas in the Southeast United States would be well above the climatological risk as indicated by hail reports. Because of the hourly nature of the data, other time windows of subdaily severe hail risk could be calculated for other products such as severe weather watches and mesoscale convective discussions.

Acknowledgments. The authors would like to acknowledge the reviewing of this manuscript, which helped to improve it. Makenzie Krocak was helpful in making sure the methods were consistent with her work. Computing for this project was performed at the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma (OU). The authors would like thank colleagues at NSSL and National Center for Atmospheric Research (NCAR) for providing supplemental data for this work. Rich Thompson (SPC) provided useful feedback during the writing process. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect the views of NOAA or the Department of Commerce.

Data availability statement. While the MRMS MESH is a product used in NWS forecast operations, there is no formal archive of the data. MESH data in this project were obtained by archiving the SPC operational MESH feed from NCEP with limited augmentation from NSSL and NCAR in the case of missing data. MRMS MESH data can be created using the WDSS-II software (http://wdssii.org) and the freely available NOAA Radar Data available via numerous cloud providers. Hail report data are available from the severe weather database at the SPC WCM website. These data are identical to what is found in the NCDC *Storm Data* publication.

REFERENCES

Allen, J. T., and M. K. Tippett, 2015: The characteristics of United States hail reports: 1955–2014. *Electron. J. Severe Storms Meteor.*, **10** (3), https://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/149.

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

——, and Coauthors, 2017: An extreme value model for U.S. hail size. *Mon. Wea. Rev.*, **145**, 4501–4519, https://doi.org/10.1175/MWR-D-17-0119.1.

Amburn, S. A., and P. L. Wolf, 1997: VIL density as a hail indicator. *Wea. Forecasting*, **12**, 473–478, https://doi.org/10.1175/1520-0434(1997)012<0473:VDAAHI>2.0.CO;2.

Ashley, W. S., A. J. Krmenec, and R. Schwantes, 2008: Vulnerability due to nocturnal tornadoes. *Wea. Forecasting*, **23**, 795–807, https://doi.org/10.1175/2008WAF2222132.1.

Blair, S. F., 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. *Electron. J. Severe Storms Meteor.*, **6** (7), http://www.ejssm.org/ojs/index.php/ejssm/article/view/87.

——, and Coauthors, 2017: High-resolution hail observations: Implications for NWS warning operations. *Wea. Forecasting*, **32**, 1101–1119, https://doi.org/10.1175/WAF-D-16-0203.1.

Brooks, H. E., C. A. Doswell, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640, https://doi.org/10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2.

Bunkers, M. J., S. R. Fleegel, T. Grafenauer, C. J. Schultz, and P. N. Schumacher, 2020: Observations of hail–wind ratios from convective storm reports across the continental United States. *Wea. Forecasting*, **35**, 635–656, https://doi.org/10.1175/WAF-D-19-0136.1.

Carbone, R. E., and J. D. Tuttle, 2008: Rainfall occurrence in the U.S. warm season: The diurnal cycle. *J. Climate*, **21**, 4132–4146, https://doi.org/10.1175/2008JCLI2275.1.

——, ——, D. A. Ahijevych, and S. B. Trier, 2002: Inferences of predictability associated with warm season precipitation episodes. *J. Atmos. Sci.*, **59**, 2033–2056, https://doi.org/10.1175/1520-0469(2002)059<2033:IOPAWW>2.0.CO;2.

Cheeks, S. M., S. Fueglistaler, and S. T. Garner, 2020: A satellite-based climatology of central and southeastern U.S. mesoscale convective systems. *Mon. Wea. Rev.*, **148**, 2607–2621, https://doi.org/10.1175/MWR-D-20-0027.1.

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, https://doi.org/10.1175/WAF-D-11-00151.1.

Davis, S. M., and J. G. LaDue, 2004: Nonmeteorological factors in warning verification. *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., P2.7, https://ams.confex.com/ams/11aram22sls/techprogram/paper_81766.htm.

Doswell, C. A., 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, https://doi.org/10.1175/WAF866.1.

Easterling, D. R., and P. J. Robinson, 1985: The diurnal variation of thunderstorm activity in the United States. *J. Climate Appl. Meteor.*, **24**, 1048–1058, https://doi.org/10.1175/1520-0450(1985)024<1048:TDVOTA>2.0.CO;2.

Edwards, R., 2006: Supercells of the Serranias del Burro (Mexico). *23rd Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., P6.2, https://www.spc.noaa.gov/publications/edwards/delburro.pdf.

Federer, B., and Coauthors, 1986: Main results of Grossversuch IV. *J. Climate Appl. Meteor.*, **25**, 917–957, https://doi.org/10.1175/1520-0450(1986)025<0917:MROGI>2.0.CO;2.

Haberlie, A. M., and W. S. Ashley, 2019: A radar-based climatology of mesoscale convective systems in the United States. *J. Climate*, **32**, 1591–1606, https://doi.org/10.1175/JCLI-D-18-0559.1.

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

Hitchens, N. M., and H. E. Brooks, 2014: Evaluation of the storm prediction center's convective outlooks from day 3 through day 1. *Wea. Forecasting*, **29**, 1134–1142, https://doi.org/10.1175/WAF-D-13-00132.1.

Houze, R. A., B. F. Smull, and P. Dodge, 1990: Mesoscale organization of springtime rainstorms in Oklahoma. *Mon. Wea. Rev.*, **118**, 613–654, https://doi.org/10.1175/1520-0493(1990)118<0613:MOOSRI>2.0.CO;2.

Jewell, R., and J. Brimelow, 2009: Evaluation of Alberta hail growth model using severe hail proximity soundings from the United States. *Wea. Forecasting*, **24**, 1592–1609, https://doi.org/10.1175/2009WAF2222230.1.

Krocak, M. J., 2017: A sub-daily severe weather climatology and its implications for forecasting. M.S. thesis, School of Meteorology, University of Oklahoma, 82 pp.

——, and H. E. Brooks, 2018: Climatological estimates of hourly tornado probability for the United States. *Wea. Forecasting*, **33**, 59–69, https://doi.org/10.1175/WAF-D-17-0123.1.

——, and ——, 2020: An analysis of subdaily severe thunderstorm probabilities for the United States. *Wea. Forecasting*, **35**, 107–112, https://doi.org/10.1175/WAF-D-19-0145.1.

Maddox, R. A., D. M. Rodgers, and K. W. Howard, 1982: Mesoscale convective complexes over the United States during 1981—Annual summary. *Mon. Wea. Rev.*, **110**, 1501–1514, https://doi.org/10.1175/1520-0493(1982)110<1501:MCCOTU>2.0.CO;2.

Mead, C., and R. Thompson, 2011: Environmental characteristics associated with nocturnal significant-tornado events in the Great Plains. *Electron. J. Severe Storms Meteor.*, **6** (6), https://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/84.

Melick, C. J., I. L. Jirak, J. Correia Jr., A. R. Dean, and S. J. Weiss, 2014: Exploration of the NSSL maximum expected size of hail (MESH) product for verifying experimental hail forecasts in the 2014 spring forecast experiment. *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 76, https://ams.confex.com/ams/27SLS/webprogram/Paper254292.html.

Murillo, E. M., and C. R. Homeyer, 2019: Severe hail fall and hailstorm detection using remote sensing observations. *J. Appl. Meteor. Climatol.*, **58**, 947–970, https://doi.org/10.1175/JAMC-D-18-0247.1.

——, ——, and J. T. Allen, 2020: A 22-year hail climatology using GridRad MESH observations. *20th Symp. on Meteorological Observation and Instrumentation*, Boston, MA, Amer. Meteor. Soc., 11.2, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/366946.

Ortega, K. L., 2018: Evaluating multi-radar, multi-sensor products for surface hail-fall diagnosis. *Electron. J. Severe Storms Meteor.*, **13** (1), https://ejssm.org/ojs/index.php/ejssm/article/viewArticle/163.

——, T. M. Smith, G. J. Stumpf, J. Hocker, and L. Lopez, 2005: A comparison of multi-sensor hail diagnosis techniques. *21st Int. Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, San Diego, CA, Amer. Meteor. Soc., P1.11, https://ams.confex.com/ams/Annual2005/techprogram/paper_87640.htm.

——, ——, and ——, 2006: Verification of multi-sensor, multi-radar hail diagnosis techniques. *Symp. on the Challenges of Severe Convective Storms*, Atlanta, GA, Amer. Meteor. Soc., P1.1, https://ams.confex.com/ams/Annual2006/techprogram/paper_104885.htm.

——, ——, K. L. Manross, K. A. Scharfenberg, A. W. A. G. Kolodziej, and J. J. Gourley, 2009: The Severe Hazards Analysis and Verification Experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530, https://doi.org/10.1175/2009BAMS2815.1.

Podlaha, A., S. Bowen, and M. Lörinc, 2020: Weather, Climate & Catastrophe Insight: 2019 Annual Report. Annual Rep. AON, 83 pp., http://thoughtleadership.aon.com/Documents/20200122-if-natcat2020.pdf.

Potvin, C. K., C. Broyles, P. S. Skinner, H. E. Brooks, and E. Rasmussen, 2019: A Bayesian hierarchical modeling framework for correcting reporting bias in the U.S. tornado database. *Wea. Forecasting*, **34**, 15–30, https://doi.org/10.1175/WAF-D-18-0137.1.

Reif, D. W., and H. B. Bluestein, 2017: A 20-year climatology of nocturnal convection initiation over the central and southern Great Plains during the warm season. *Mon. Wea. Rev.*, **145**, 1615–1639, https://doi.org/10.1175/MWR-D-16-0340.1.

Rhodes, B., 2019: Skyfield v1.16. Brandon Rhodes, accessed 21 January 2020, https://github.com/skyfielders/python-skyfield.

Riley, G. T., M. G. Landin, and L. F. Bosart, 1987: The diurnal variability of precipitation across the central Rockies and adjacent Great Plains. *Mon. Wea. Rev.*, **115**, 1161–1172, https://doi.org/10.1175/1520-0493(1987)115<1161:TDVOPA>2.0.CO;2.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Rothfusz, L. P., R. Schneider, D. Novak, K. Klockow-McClain, A. E. Gerard, C. Karstens, G. J. Stumpf, and T. M. Smith, 2018: FACETs: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99**, 2025–2043, https://doi.org/10.1175/BAMS-D-16-0100.1.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Stelten, S., and W. A. Gallus Jr., 2017: Pristine nocturnal convective initiation: A climatology and preliminary examination of predictability. *Wea. Forecasting*, **32**, 1613–1635, https://doi.org/10.1175/WAF-D-16-0222.1.

Stumpf, G. J., T. M. Smith, and A. E. Gerard, 2002: The Multiple-Radar Severe Storm Analysis Program (MR-SSAP) for WDSS-II. *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 4.6, https://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_46790.htm.

Waldvogel, A., B. Federer, W. Schmid, and J. F. Mezeix, 1978: The kinetic energy of hailfalls. Part II: Radar and hailpads. *J. Appl. Meteor.*, **17**, 1680–1693, https://doi.org/10.1175/1520-0450(1978)017<1680:TKEOHP>2.0.CO;2.

Wendt, N. A., I. L. Jirak, and C. J. Melick, 2016: Verification of severe weather proxies from the NSSL-WRF for hail forecasting. *28th Conf. on Severe Local Storms*, Portland, OR, Amer. Meteor. Soc., 110, https://ams.confex.com/ams/28SLS/webprogram/Paper300913.html.

Wilson, C. J., K. Ortega, and V. Lakshmanan, 2009: Evaluating multi-radar, multi-sensor hail diagnosis with high resolution hail reports. *25th Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, Amer. Meteor. Soc., P2.9, https://ams.confex.com/ams/pdfpapers/146206.pdf.

Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. W. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2.

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.