UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

Examining the Predictability of Tornadic and Nontornadic Non-Supercellular MCS Storms using GridRad-Severe Radar Data and Machine Learning Techniques

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

Amanda M. Murphy
Norman, Oklahoma
2023

Examining the Predictability of Tornadic and Nontornadic Non-Supercellular MCS Storms using GridRad-Severe Radar Data and Machine Learning Techniques

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Cameron Homeyer, Chair

Dr. Marcela Loria-Salazar

Dr. Amy McGovern

Dr. Corey Potvin

Dr. Leslie Flanagan

© Copyright by Amanda M. Murphy 2023
All Rights Reserved.

Acknowledgments

Thank you so much to Dr. Cameron Homeyer, who allowed me to work on this incredibly interesting project and be a part of the GridRad-Severe dataset creation. His guidance has been an invaluable part of my graduate school journey. Also thank you to my committee–Drs. Marcela Loria-Salazar, Amy McGovern, Corey Potvin, and Leslie Flanagan–for their scientific expertise and overall support throughout this process. Thank you also to Drs. Elisa Murillo and Matthew Brown for their insight and continued collaboration on the social vulnerabilities portion of this work; Drs. Ryan Lagerquist and Randy Chase for their invaluable guidance on model architecture and hyperparameter testing for the ML portion; and Kiley Allen for her assistance in building the GR-S dataset. Finally, I am constantly buoyed by the support of my friends and family, without whom this dissertation would not be possible.

Figures herein were designed to be equally interpretable to individuals with full color vision and individuals with color vision deficiency. This material is based upon work supported by the National Science Foundation under grant ICER-2019758 and the NOAA Weather Program Office Joint Technology Transfer Initiative grant NA20OAR4590356 to the University of Oklahoma. This work is part of the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

List of Tables

List of Figures

Abstract

Many studies have aimed to identify novel storm characteristics that are indicative of current or future severe weather potential using a combination of ground-based radar observations and severe reports. However, this is often done on a small scale using limited case studies on the order of tens to hundreds of storms due to how time-intensive this process is. Herein, we introduce the GridRad-Severe dataset, a database including $\sim$100 severe weather days per year and upwards of 1.3 million objectively tracked storms from 2010-2019. Composite radar volumes spanning objectively determined, report-centered domains are created for each selected day using the GridRad compositing technique, with dates objectively determined using report thresholds defined to capture the highest-end severe weather days from each year, evenly distributed across all severe report types (tornadoes, severe hail, and severe wind). Spatiotemporal domain bounds for each event are objectively determined to encompass both the majority of reports as well as the time of convection initiation. Severe weather reports are matched to storms that are objectively tracked using the radar data, so the evolution of the storm cells and their severe weather production can be evaluated. Herein, we apply storm mode (single cell, multicell, or mesoscale convective system) and right-moving supercell classification techniques to the dataset, and revisit various questions about severe storms and their bulk characteristics posed and evaluated in past work. Additional applications of this dataset are reviewed for possible future studies.

Given this large dataset of severe storms, questions about storm structure of very specific storm types can be investigated using what is still a large subsample of the total GridRad-Severe dataset. This study compares populations of tornadic non-supercellular MCS storm cells to their nontornadic counterparts, focusing on nontornadic storms that have similar radar characteristics to tornadic storms. Comparison of single-polarization radar variables during storm lifetimes show that median values of low-level, mid-level, and column-maximum azimuthal shear, as well as low-level radial divergence, enable the

highest degree of separation between tornadic and nontornadic storms. Focusing on low-level azimuthal shear values, null storms were randomly selected such that the distribution of null low-level azimuthal shear values matches the distribution of tornadic values. After isolating the null cases from the nontornadic population, signatures emerge in single-polarization data that enable discrimination between nontornadic and tornadic storms. In comparison, dual-polarization variables show little deviation between storm types. Tornadic storms both at tornadogenesis and at 20-minute lead time show collocation of the primary storm updraft with enhanced near-surface rotation and convergence, facilitating the non-mesocyclonic tornadogenesis processes.

With this additional knowledge about the structure of tornadic vs. nontornadic storms and which radar variables best differentiate the two, machine learning methods can be used to learn the differences between these storm type at various lead times and improve tornado predictability. A convolutional neural network was trained on tornadic and nontornadic data where the nontornadic data were either sampled from storms that have similar radar characteristics to tornadic storms as in the PMM analyses or sampled from the entire population of non-supercellular MCS storms. These models were then tested on independent data from 2020-2021, again either including all tornadic storms and sampling nontornadic cases as in the PMM analyses or including all tornadic and nontornadic storms. Models that were tested on all tornadic and nontornadic storms, whether they were trained and validated on datasets including sampled strong nontornadic storms or a sample of all nontornadic storms, both performed well below the baseline performance metrics from the NWS. However, when the model was trained, validated, and tested using samples of all tornadic storms and only strong nontornadic storms, model test performance far exceeded the baseline NWS metrics. Performance metrics include a probability of detection (POD) of 79%, a false alarm ratio (FAR) of 58%, and a CSI of 0.38. Compared to the NWS metrics of 49%, 75%, and 0.2, respectively, this model shows clear promise as a supplemental forecasting tool for scenarios where a storm is identified as (at least) borderline tornadic.

However, further analyses of the model performance scaled to account for the true proportion of tornadic vs. nontornadic storms shows that it was the unnatural ratio of tornadic to nontornadic storms, and not the focus on strong nontornadic storms, that was the cause for the improved model performance.

Finally, a brief analysis of the underlying populations and their demographic characteristics in the vicinity of tornadoes are examined. Special attention is given to non-supercellular MCS storms, as well as discrete supercells, whose tornadoes are often a main focus of tornado research in the U.S. Analyses show that groups making up $\sim$3% or less of the CONUS mean population typically have lower relative population densities in the vicinity of storms. The Black or African American Alone demographic has higher relative populations in the vicinity of all tornadoes compared to their CONUS mean population density, as do all Non-Hispanic categories (Not Hispanic, Non-Hispanic White and Non-Hispanic Black). Comparing population densities near specific types of tornadoes (i.e., mode and combination of mode and human impact) to their densities near all tornadoes, the White Alone demographic has population densities near the CONUS mean for supercellular tornadoes, but that density jumps 6-7 percentage points in the vicinity of deadly supercellular tornadoes when examining underlying population density by deadly event and by death, suggesting that the deadliest supercellular tornadoes occur in predominantly White areas. On average, populations in the vicinity of all tornadoes have $\sim$75-80% higher Black or African American Alone and Non-Hispanic Black densities when compared to the CONUS mean, with those demographics' relative densities only increasing when isolating MCS tornadoes and deadly MCS tornadoes, suggesting that the deadliest MCS tornadoes preferentially occur in areas with relatively higher Black or African American Alone and Non-Hispanic Black populations. One particularly striking result is that the mean Social Vulnerability Index (SVI) of populations near all tornadoes is just barely above the CONUS mean (0.52 vs. CONUS mean of 0.51), but is slightly lower for supercellular tornadoes (0.49) and higher for MCS tornadoes (0.57). Therefore, MCS tornadoes tend to occur in

areas that are less resilient to natural disasters than both the CONUS mean and areas in the vicinity of supercellular tornadoes. For both MCS and supercellular tornadoes that were associated with deaths or injuries, the local SVI is higher, likely pointing to the applicability of SVI in identifying areas less resilient to natural disasters.

Chapter 1

Introduction

1.1 GridRad-Severe

Severe weather, including tornadoes, severe hail, and severe wind, has substantial impacts across the U.S. each year. NCEI (2023) reported that severe weather accounted for approximately 8.5 billion dollars (inflation adjusted) in annual losses and nearly 2,000 deaths from 1980-2022. Additionally, there have been 163 severe weather events each totaling 1 billion dollars (inflation adjusted) in losses or more, including derechos, hail storms, and tornado outbreaks, and 38 events with 10 or more casualties over that same period. It remains important to analyze these, and other such high-end severe weather events, to further improve our resilience to them.

Examining a severe weather event is inherently multifaceted, and the data used depend primarily on the type of analysis. These data often include synoptic-scale and mesoscale data starting in the days to hours preceding an event (e.g., Rockwood and Maddox 1988; Coniglio et al. 2011; Hurlbut and Cohen 2014; Vaughan et al. 2017), radar and satellite data to examine storm-scale features and evolution during the event, and storm reports in the aftermath to evaluate impacts. Radar data can be incredibly useful to understand the physical and kinematic structure of severe vs. non-severe storms. In particular, such data have provided insight into the intensity of precipitation, horizontal and vertical extents, wind speeds, flow patterns, rotational velocities, and precipitation distributions associated with a storm (e.g., Byers and Braham 1949; Browning 1964; Brown et al. 1978; Lemon and Doswell 1979; Wurman et al. 1996; Parker and Johnson 2000). Radars have been used for several decades to understand tornadic storms and tornadogenesis (e.g., Lemon and Doswell 1979; Ryzhkov et al. 2002, 2005; Kumjian and Ryzhkov 2008; Kurdzo et al. 2017; Homeyer et al. 2020), estimate hail size in a storm (e.g., Witt et al. 1998; Murillo and Homeyer 2019), and

better understand severe straight-line wind events (e.g., Fujita and Byers 1977; Theodore Fujita 1990; Wakimoto 2001; Klimowski et al. 2003). With the advent of dual-polarization radar and integration of such radars into the operational network of S-band radars in the U.S. (NEXRAD network) in 2013, a wealth of additional information can be inferred from these data including improved hydrometeor classification, detecting the presence and size of hail, convective updraft and vertical wind shear identification, and detection of tornadic debris (Kumjian 2013). Radar datasets therefore remain powerful tools to further understand the structure of severe storms and any unique identifying characteristics that can be used in real-time for warning decisions.

One such radar analysis technique that has been performed both manually and objectively is storm mode classification. Accurate identification of storm mode allows for further insight into the potential for various types of severe weather. Common storm classifications include single cell storms, multicellular storms, and mesoscale convective systems (MCSs). For smaller-scale studies, subjective (manual) identification is often performed, as it is not prohibitively time intensive for so few samples. However, for studies using larger databases of storms, objective methods are a practical and often necessary solution for classification. Various studies use observed or simulated column-maximum reflectivity and a 30-40 dBZ threshold (e.g., Trapp et al. 2005; Snively and Gallus 2014; Thielen and Gallus 2019) to define contours encapsulating convective elements. Using radar data analyzed on the order of minutes to hours, these techniques often incorporate constraints for aspect ratio (i.e., the length to width ratio; e.g., Bluestein and Jain 1985; Fowle and Roebber 2003; Gallus et al. 2008; Smith et al. 2012; Snively and Gallus 2014; Thielen and Gallus 2019), maximum contour dimension (e.g., Bluestein and Jain 1985; Parker and Johnson 2000; Trapp et al. 2005; Gallus et al. 2008; Smith et al. 2012; Snively and Gallus 2014; Thielen and Gallus 2019), enclosed area (e.g., Fowle and Roebber 2003), and storm persistence/duration (e.g., Geerts 1998; Pinto et al. 2015; Feng et al. 2018, 2019).

Beyond determining a storm's mode, one common approach to analyzing their severe weather potential is through case studies, especially using radar observations and severe reports. Case studies of severe weather events and their radar presentations abound in the literature, providing valuable fine-scale insight into the inner workings of severe thunderstorms. For example, a case study examination of a tornadic supercell in Oklahoma on 3 May 1999 led to the discovery of the polarimetric radar tornadic debris signature (TDS) by Ryzhkov et al. (2002), prompting analyses of other supercells for potential analogous signatures in Ryzhkov et al. (2005) and the eventual inclusion of the polarimetric TDS in modern guides on polarimetric radar utility (e.g., Kumjian 2013). Additionally, Fujita and Byers (1977) examined the meteorological conditions surrounding an airplane crash and detected thunderstorm winds that were much stronger than anything previously observed, coining these winds a "downburst." This observation prompted numerous field campaigns targeting downburst-producing storms (e.g., Fujita and Wakimoto 1982; McCarthy et al. 1982; Wilson et al. 1988), leading to a more comprehensive understanding of the phenomenon today and likely saving many lives (Wilson and Wakimoto 2001). Working with case studies allows for very detailed analyses of severe storms, but the conclusions from such studies are limited in generalizability given their relatively small sample sizes.

Large-scale studies of severe weather in the literature date as far back as the 1940s with the Thunderstorm Project (Byers and Braham 1949), and climatological studies have driven many of the scientific community's advances in severe weather knowledge. Tornadoes in the U.S. are most common in early summer (Brooks et al. 2003) and in the late afternoon to early evening (Ashley et al. 2008), mostly in the Great Plains and into the Southeast (Coleman and Dixon 2014; Gensini and Brooks 2018; Krocak and Brooks 2018). The majority of reported tornadoes are weak (EF-0 to EF-1; e.g., Brooks and Doswell 2001; Trapp et al. 2005) although the less frequent significant tornadoes (EF 2+) are responsible for nearly 90% of fatalities (e.g., Anderson-Frey and Brooks 2019). Climatologies of severe hail like those by Cintineo et al. (2012), Murillo et al. (2021), and Wendt and Jirak (2021)

are also typically built using reports, despite well-known reporting limitations (see Allen and Tippett 2015, and references therein). Allen and Tippett (2015) examined a 60-year record of over 260,000 hail reports and found that the majority of all hail reports occurred in the late afternoon to early evening primarily during the late spring, with a maximum in the Great Plains that is slowly shifting northward. Studies of severe winds often focus attention on MCSs, which can produce uniquely widespread and damaging severe winds, including and especially from derechos (Johns and Hirt 1987; Corfidi et al. 2016). Coniglio and Stensrud (2004) found that higher-end derecho events in their 16-year climatology favor the southern plains and Midwest. Derechos in the eastern two-thirds of the CONUS tend to occur more in the summer months (Coniglio and Stensrud 2004) in the late evening to overnight (Bentley and Mote 1998). Recent work using machine learning to classify and track MCSs and quasi-linear convective systems (QLCSs) by Ashley et al. (2019) showed, using their developed MCS climatology, that nearly a third of all MCSs in their 22-year dataset were also QLCSs, and QLCS storms were linked with 28% of all severe wind reports in the central and eastern U.S. Climatological studies allow for a large-scale view of severe weather to develop mental models for how, when, and where severe hazards are likely to occur.

Several studies have investigated compelling scientific questions about severe storms beyond just their climatological distributions using a large record of observations. For example, studies such as Homeyer et al. (2020), Loeffler et al. (2020), and Van Den Broeke (2020) use radar data from tens to hundreds of tornadic and nontornadic supercells to examine tornadogenesis predictability; Blair et al. (2011) and Gutierrez and Kumjian (2021) examined radar signatures within tens to hundreds of giant and gargantuan hail-producing storms; and Bluestein and Jain (1985) and Schiesser et al. (1995) looked at mesoscale structures within radar data from dozens of severe MCSs. Past studies such as these, despite using larger datasets than the more numerous case study analyses, rarely contain more than a few hundred storms and are commonly limited in both their spatiotemporal extent

and temporal resolution of observations. Furthermore, any larger-scale studies done before 2013 have limited to no access to polarimetric radar data, which further limits understanding of storm microphysics that may be relevant to severe events. Therefore, there exists a need to expand these studies using a longer temporal record and analyze a much larger population of storms that are more spatiotemporally diverse, have higher temporal resolution data, and that occurred within the observational range of one or more polarimetric radars.

Recognizing the contributions of the aforementioned prior work to our understanding of severe storms, and with the increasing record of observations and emergence of a national polarimetric radar network, a clear incentive exists to create a modern database of radar data and severe weather reports to evaluate the characteristics of storms that produce severe weather. This paper aims to fill the knowledge gaps outlined herein using the newly developed GridRad-Severe database (hereafter abbreviated GR-S): a database including gridded multi-radar data covering the majority of the CONUS, objective storm tracks, and storm reports. Herein, we outline the creation and utility of GR-S as well as how well it reflects the spatiotemporal distribution of all storm reports in the U.S. Additionally, we introduce objective storm mode and supercell classification techniques to aid in data analysis, and examine initial findings of the GR-S dataset that complement and expand upon past studies. Finally, we directly compare GR-S findings with results from select seminal papers to demonstrate its ability to replicate and extend prior key findings.

1.2 Statistical characteristics of tornadic and nontornadic mesoscale convective system storms

Developing a large record of storms not only allows for bulk examination of severe storms, but also creates the opportunity to segment the data to examine very specific types of storms while retaining a large sample size. Although case study analyses of specific types of storms, such as discrete supercells or non-supercellular MCS storms that are associated

with severe reports, can elucidate radar signatures conducive to the strengthening, maintenance, and decay of that specific storm, examining the radar characteristics of a large number of such storms can confirm whether or not the patterns observed in case studies are indicative of the behavior of most if not all storms of that type. The GR-S dataset therefore can facilitate a deeper understanding of specific types of storms using fine-scale data typical of case study analyses and sample sizes typical of statistical analyses.

Improving the predictability of tornadoes and tornadic storms is crucial to the National Weather Service (NWS) mission statement of protecting life and property (Uccellini and Hoeve 2019). From 1986 to 2011, tornadoes warned in advance had an average lead time of 18.8 minutes, and the total frequency of such advanced warnings increased from $\sim$25% in 1986 to more than 50% in recent years (Brooks and Correia 2018). Tornado probability of detection (POD; i.e., percent of tornadic storms warned in advance of tornadogenesis) is $>$ 80% for tornadoes from supercells but less than 50% for tornadoes from nonsupercellular storms, which are warned on average $\sim$2 minutes later than supercellular tornadoes and are three times more likely to be warned at negative lead times (i.e., after the time of tornadogenesis; Brotzge et al. 2013). Nonsupercellular tornadoes are produced not by a persistently rotating mesocyclonic storm, but rather by the tilting of vorticity into the vertical by updrafts or downdrafts to create counterrotating mesovortices at the leading edge of a storm (e.g., Weisman and Davis 1998; Trapp and Weisman 2003; Flournoy and Coniglio 2019). In particular, Flournoy and Coniglio (2019) and Gibbs (2021) note in their reviews of quasi-linear convective system (QLCS) tornadogenesis that line-normal low-level shear plays an important role in mesovortex genesis for QLCSs when cold pool and updraft strength are sufficiently balanced by RKW theory (see also Weisman and Trapp 2003; Rotunno et al. 1988). Knowing that QLCS storms are a type of mesoscale convective system (MCS; e.g., Agee and Jones 2009), there is clear incentive to examine methods to improve non-supercellular MCS tornado warnings.

Climatological studies have found that anywhere from 18-27% of tornadoes are associated with MCSs or QLCSs (Trapp et al. 2005; Ashley et al. 2019; Murphy et al. 2023). Murphy et al. (2023) also found that approximately half of sub-significant severe tornadoes were associated with MCSs, and MCSs accounted for the majority of tornado-producing storms during the overnight hours. Nighttime tornadoes can pose a greater threat to life and property since people are asleep and are least confident in their ability to receive warnings even when a tornado is detected at positive lead times (Krocak et al. 2021). The negative impacts of a low POD for non-supercell QLCS storms are likely to only amplify with time, since the proportion of QLCS-attributed tornadoes is increasing (Ashley et al. 2019). Acknowledging that improved warnings do not operate in a vacuum and that any behavioral response to warnings is impacted by both meteorological and social factors (e.g., Agee and Jones 2009; Hoekstra et al. 2011; Mason et al. 2018; Ripberger et al. 2019; Ernst et al. 2021; Trujillo-Falcón et al. 2021), greater understanding of how MCS non-supercellular tornadoes form and what storm and environmental factors delineate a tornadic from a non-tornadic MCS cell can aid in improving warning metrics and saving life and property.

Often, attempts to nowcast a storm's tornadic potential rely on local environmental analyses and comparisons to historical environmental analogs. Coffer et al. (2019) note that such analyses go back as far as the 1940's and 1950's, with modern techniques at analyzing storm environments mainly relying on model-derived proximity soundings given the relative dearth of proximity soundings near severe and/or tornadic storms. Thompson et al. (2012) examined differences between QLCS and right-moving supercellular tornado environments using hourly Rapid Update Cycle (RUC) analysis data, finding that "environmental differences between the supercells and linear modes were relatively small." Anderson-Frey et al. (2016) also examined QLCS and right-moving supercellular tornadic environments using RUC data, again finding a great deal of overlap between supercellular and QLCS tornadic environments' derived parameters. In addition to overlap in environments for various storm modes, hits and misses for tornado warnings within environmental

parameter spaces of each storm mode also have considerable overlap, "perhaps highlighting the difficulty in distinguishing between storms that will be tornadic and those that will not be tornadic in similar environments" (Anderson-Frey et al. 2016). Lyza et al. (2022) note in their analyses of supercells during the April 2011 tornado outbreak that tornado-producing supercells were occasionally located near "dormant" cells, motivating the need to look beyond storm background environments and towards storm-scale environmental heterogeneities and "internal stochastic processes of supercells" to determine whether or not a storm will produce a tornado. Additionally, environmental reanalysis data have shown limitations in representing the low-level environment of storms (Taszarek et al. 2021; Coniglio and Jewell 2022), reducing their utility for tornadogenesis prediction.

Another way to anticipate potential tornadogenesis is by analyzing a storm's presentation on radar and comparing it to historical radar data. While not inherently a predictive tool, weather radar can give insight into storm dynamics and physics, thereby helping to determine if a storm is developing in a manner consistent with past tornadic storms. The Warn-On Forecast program (Stensrud et al. 2009, 2013; Heinselman et al. 2023) has used simulated radar reflectivity along with other model parameters to attempt to forecast which storms may become tornadic in the next few hours. When examining its ability to forecast all thunderstorms, Skinner et al. (2018) showed that such studies' average forecast critical success index (CSI or "skill"; Schaefer 1990) ranged from 0.7 at 20-minute forecast time to 0.4 at near 3-hour forecast times. Focusing on observed radar signatures, Homeyer et al. (2020) analyzed hundreds of tornadic and nontornadic supercell storms in the United States to elucidate common radar signatures of both types of storms, both at peak intensity and at 20 minute lead times. Using a probability-matched mean approach, which produces maps of average storm radar moments and/or variables at selected altitudes without artificially dampening the full range of values typical of each individual storm (PMM; Ebert 2001), they found tornadic supercell storms to have increased vertical alignment of

the mesocyclone and differing orientation of low-level polarimetric radar signatures relative to the direction of storm motion. The aforementioned study by Lyza et al. (2022) also looked at radar characteristics of supercells and found that mean azimuthal shear (rotation) discriminated well between nontornadic, pretornadic, and tornadic supercells. These studies on how radar observations of supercells relate to tornado potential motivate applying a similar technique to non-supercell storms. For example, Loeffler and Kumjian (2018) examined low-level differential reflectivity (Z_{DR}) and specific differential phase (K_{DP}) signatures in tornadic and nontornadic non-supercell storms with varying storm mode, finding statistically significant differences in the way the maxima of the two were oriented relative to storm motion in tornadic vs. nontornadic storms. By examining the radar presentation of non-supercell MCS storm cells specifically, potential differences between tornadic and nontornadic cells can be elucidated for improved short-term forecasting and nowcasting of tornadoes.

Herein, we present a PMM analysis of thousands of tornadic and nontornadic non-supercellular MCS storm cells identified in the 10-year GridRad-Severe dataset to determine if there are radar signatures that enable discrimination between storm types at both peak intensity and at a given lead time to first tornadogenesis. We first summarize the technique used to identify an appropriate null population of nontornadic storms. Differences in PMM fields of radar variables and moments are then evaluated at multiple altitudes above ground level (AGL) for each storm type. Finally, these signatures are summarized and directions for future research are discussed.

1.3 Machine learning and severe weather prediction

Creation of GR-S offers an opportunity to leverage this large severe weather dataset and machine learning techniques to possibly further improve severe storm prediction and nowcasting. After determining the typical radar-derived characteristics of tornadic and nontornadic non-supercellular MCS storms and what radar moments or variables best discriminate between the two, that relevant information can be ingested into a machine learning model for training and validation. The skill of the model during testing can be evaluated to determine how well it is able to predict whether or not a storm will produce a tornado.

The use of machine learning in meteorology has been steadily increasing over the past 30 years, increasing exponentially in total number of atmospheric science/meteorology publications and linearly in percent contribution to all atmospheric science/meteorology publications (Chase et al. 2022). Recent creation of a dedicated artificial intelligence journal under the American Meteorological Society journal umbrella (McGovern and Broccoli 2022) and the NSF AI2ES institute (McGovern et al. 2022) further highlight the growing interest in machine learning and utilization of machine learning techniques in meteorological research. Publicly available information on the implementation of machine learning in meteorological research (McGovern et al. 2019; Chase et al. 2022, 2023) makes understanding and applying machine learning in atmospheric science research more accessible than ever before.

Artificial intelligence and machine learning are often conflated when discussing their use in meteorology, but artificial intelligence is a larger umbrella containing machine learning (Boukabara et al. 2021). Chase et al. (2022) define machine learning as "any empirical method where parameters are fit (i.e., learned) on a training dataset in order to optimize (e.g., minimize or maximize) a predefined loss (i.e., cost) function." These include techniques ranging from simple linear regression to deep neural networks. Typical applications of machine learning in meteorology use supervised learning, where the model is trained on a set of inputs that have a corresponding set of outputs. The outputs can either be a range of

values or some binary indicator of whether or not an outcome will occur; these are called regression and classification tasks, respectively. More information about machine learning in meteorology is available in guides written by Chase et al. (2022) and Chase et al. (2023).

A number of problems have been investigated using machine learning, including the nowcasting of radar echo intensity (Cuomo and Chandrasekar 2021), precipitation (Ravuri et al. 2021), lightning (Cintineo et al. 2022), convective mode (Jergensen et al. 2020), and severe weather (Cintineo et al. 2020; Lagerquist et al. 2020; Flora et al. 2021; Mecikalski et al. 2021). McGovern et al. (2023) present a comprehensive review of machine learning applications for convection, highlighting "the challenges in developing ML approaches to forecast these phenomena across a variety of spatial and temporal scales." To combat the fact that machine learning models are sometimes considered "black boxes" where information goes in, predictions come out, and little is known about how the model came to that conclusion (particularly when using deep neural networks), McGovern et al. (2019) discuss a number of "model interpretation and visualization (MIV)" techniques, such as saliency maps, gradient-weighted class-activation maps, backward optimization, and novelty detection for deep learning methods. Another term for such techniques to understand the decision-making process of artificial intelligence systems is explainable AI (XAI).

As previously discussed, forecasting and/or nowcasting tornadogenesis remains a difficult problem in meteorology. To examine new ways to improve tornadogenesis nowcasting, some recent efforts have turned to machine learning, particularly deep learning, to improve predictions. Lagerquist et al. (2020) use a convolutional neural network (CNN) to examine both radar images and local environmental soundings to determine whether or not a given storm would become tornadic within the next hour. Developed based on studies examining the human brain's visual cortex, CNNs are built specifically for image recognition and have achieved "superhuman performance on some complex visual tasks" (Géron 2019). Lagerquist et al. (2020) found that a model trained on GridRad data acheived a CSI of 0.31

and area under the receiver operating characteristic (ROC) curve (AUC) of over 0.9. However, skill was not equally impressive across all storm types; the worst predictions included false positives for strongly rotating nontornadic supercells and false negatives for tornadic storms embedded within a larger line. Sandmæl et al. (2023) propose using machine learning operationally to identify tornadic storms, going beyond identification of tornado vortex signatures (TVSs) using the Tornado Detection Algorithm (TDA; Mitchell et al. 1998) and instead using a random forest technique on single-radar data to identify tornadic storms. Gensini et al. (2021) instead focused solely on environmental variables, using random forest classification to achieve an AUC of 0.78 and CSI of 0.23 when trying to differentiate between tornadoes and significant tornadoes. Steinkruger et al. (2020) took machine learning for tornado prediction one step further, developing a machine learning model that both determines tornado potential and makes tornado warning decisions–meaning, not only determining if there is tornadic potential but whether or not to warn a storm. Overall, machine learning models show promise for not only improving tornado detection, but also being able to do so in a real-time operational setting.

Given the documented difficulty of predicting QLCS tornadogenesis using a CNN in Lagerquist et al. (2020), herein, we use a CNN trained, validated, and tested on GR-S data to determine the model's utility in predicting non-supercellular MCS storms' tornadogenesis potential within the next 20 minutes. First, the model architecture and input data are described, as well as the method for selecting optimal hyperparameters. The model is then trained, validated, and tested on GR-S data, and results are discussed. Model performance is directly compared to performance metrics from the NWS to evaluate not only the skill that the models have in correctly identifying tornadic storms, but also any skill above what is already regularly acheived in NWS operations. This may highlight scenarios in which ML could be useful in operational settings during severe weather.

1.4 Societal impacts of severe weather

Discussions so far have focused on improvements in weather forecasting–data availability, analysis techniques, and machine learning applications–that can then improve non-supercellular MCS tornado warnings and the outcomes of those impacted by these tornadoes. However, increased warning performance is only part of the solution. No matter the quality of a warning, the human response to a warning is a large component in the impact that the local population experiences. Uccellini and Hoeve (2019) note this in a recent discussion on the Weather-Ready Nation (WRN) initiative, stating that "to meet the NWS mission of saving lives and property and enhancing the national economy, the NWS must improve the accuracy and timeliness of forecasts and warnings, and must directly connect these forecasts and warnings to critical life- and property-saving decisions through the provision of impact-based decision support services (IDSS)." They note the need for the NWS to move beyond just product generation and actively improve information dissemination to "ensure the 'message delivered equals the message received'".

Bridging the gap from "message delivered" to "message received" requires an understanding of the social characteristics of the populations you are trying to communicate with. To date, there have been a plethora of case studies examining warning comprehension and response, as well as severe weather preparedness, risk perception, vulnerability, and impact based on the underlying population (e.g., Chaney and Weaver 2010; Ahlborn et al. 2012; Burke et al. 2012; Chaney et al. 2013; Chiu et al. 2013; Kousky 2013; Silver and Andrey 2014; Paul and Stimers 2014; Jauernic and Broeke 2017; Sherman-Morris et al. 2022). Examining 200 years of tornado-related deaths, Agee and Taylor (2019) found that Arkansas, Mississippi, Alabama, and Oklahoma had an above-average susceptibility to tornado deaths when normalized by population from 1916-2017. Ashley (2007) comes to a similar conclusions, noting that "most tornado fatalities occur in the lower–Arkansas, Tennessee, and lower–Mississippi River valleys of the southeastern United States" from 1880-2005. They postulated that "the relative maximum of fatalities in the Deep South and minimum in

the Great Plains may be due to the unique juxtaposition of both physical and social vulnerabilities," including higher mobile home density, poverty, population, and population growth in the Southeast compared to the Great Plains. Given that the GR-S dataset includes tracked storms, storm classifications, and matched tornado reports, tornadoes from non-supercellular MCS storms can be spatially matched to recorded demographic characteristics and social vulnerability metrics to quantify the impact of MCS tornadoes on underlying populations with varying socioeconomic characteristics.

Beyond gaining a deeper understanding of how non-supercellular MCS storms appear on radar and how nowcasting of these storms can be improved, it is also incredibly important to understand the underlying demographics of the populations that these storms impact, and how the demographic makeups of these areas and the local social vulnerabilities compare to CONUS-mean characteristics. Herein, we use the large sample of GR-S tracked storms along with publicly-available U.S. Census data to objectively match locations of non-supercellular MCS tornadoes to the local population and its demographic characteristics. By investigating the local demographics and social vulnerabilities, and comparing those data to CONUS-mean data, we can further understand any potential disproportionate impacts of non-supercellular MCS tornadoes on different underlying demographics and understand how our current deficiencies in MCS tornado prediction capabilities may be disproportionately impacting different demographic groups.

Chapter 2

Creation of GridRad-Severe[1]

2.1 Radar data

Radar data sourced from the nationwide NEXRAD network (NOAA/NWS/ROC 1991; Crum and Alberty 1993) were used to create GridRad data using version 4.2 of the public algorithm (Homeyer and Bowman 2022). GridRad data are merged volumes of individual radar observations across the CONUS, binned on a regular longitude-latitude grid. This includes single-polarization radar moments such as radar reflectivity at horizontal polarization (Z_{H}) and radial velocity spectrum width (σ_V) before 2013, and additional dual-polarization variables such as differential radar reflectivity (Z_{DR}), co-polar correlation coefficient (ρ_{hv}), and specific differential phase (K_{DP}) following the polarimetric upgrade of the radar network. Derived kinematic variables were also calculated on the native grid of each radar and binned into GridRad volumes and include radial divergence and azimuthal shear of the radial velocity. The spatial resolution of GridRad data is $\sim$0.02 $\times$ $\sim$0.02 degrees longitude-latitude (48 grid points per degree), and 0.5-km vertical resolution up to 7 km above mean sea level (AMSL), after which the vertical resolution coarsens to 1-km up to 22 km AMSL. Temporal resolution of the data is 5 minutes. More technical details about the creation of GridRad data can be found in Homeyer and Bowman (2022). GridRad is one of a few commonly used merged CONUS radar products (notable alternatives include NOAA's Multi-Radar Multi-Sensor [MRMS] and Multi-Year Reanalysis of Remotely Sensed Storms [MYRORSS]) and is unique in its breadth of merged radar variables and merging methods that aim to provide high-fidelity echo top heights and internal storm structure. Herein, we only create GridRad data for severe events within the CONUS (specifically, domains spanning 24 to 50° N and 125 to 66° W).

[1]The work within this chapter has been published in the peer-reviewed literature as Murphy et al. (2023).

The GR-S database includes radar data from 2010-2019 inclusive, with future years expected to be added over time. This dataset starts in 2010 due to both good NEXRAD coverage and being after the NEXRAD transition to super resolution (Torres and Curtis 2007). Since the tornado rating scale changed from F to EF in 2007 (Doswell et al. 2009; Edwards et al. 2013) and the severe hail size threshold changed from 0.75" to 1.0" in early 2010 (before the first date in this dataset; Allen and Tippett 2015), all reports classified as severe herein are based on a uniform threshold for hail ($\geq$ 1.0") and wind (gusts $\geq$ 50 kts) and a uniform damage rating scale for tornadoes (the Enhanced Fujita or EF scale). Significant severe reports are those meeting or exceeding EF-2 for tornadoes, 2.0" for hail, and 65 kt gusts for wind (Hales Jr. 1988).

2.2 Storm report data

Storm report data are sourced from NOAA's Storm Events Database (SED) hosted at the National Centers for Environmental Information (NCEI/NOAA 2022) from 2010 through 2019 inclusive, including tornado, hail, and wind report data. Each SED report includes a unique event ID, start and end date and time, initial and final event coordinates (longitudes and latitudes), and magnitude (EF rating for tornadoes, maximum diameter for hail, and maximum wind speed for severe wind). Tornado reports also include tornado path length and width.

2.3 Event definition

Defining a GR-S event occurs in a few distinct steps. First, high-end severe days are identified using the tornado, hail, and wind reports from the SED. To be consistent with SPC severe days and most prior work, GR-S events begin at 12 UTC on the event day and end at 12 UTC on the following day. We identify days as high-end severe days if the number of tornado, hail, or wind reports exceeds 8, 45, or 120, respectively. These primary

thresholds were chosen because they result in a nearly balanced dataset of high-end tornado, hail, and wind days each year, with approximately 100 days per year being labeled as high-end severe days (i.e., GR-S events). The thresholds also approximately correspond to the 85th percentile of daily report counts for each hazard. Once a day is included in the GR-S database (via the aforementioned primary threshold), secondary thresholds are used to determine what severe types will be used for domain definition on that day, roughly corresponding to the upper quartile of daily report counts. This is done to maximize the diversity and breadth of severe weather that is analyzed on the selected GR-S event days; if a day is already included in the dataset, it makes sense to analyze not only the severe hazard that happened enough times to warrant the day's inclusion in the dataset, but also any other severe hazards that, while not prolific enough to pass that high primary threshold, still frequently occurred on that day. The secondary thresholds are roughly one-half of the primary thresholds: greater than 4 tornado reports, 22 hail reports, or 60 wind reports. Isolating data in these ways results in a large dataset focused solely on high-end severe weather events with a high level of spatiotemporal detail.

After high-end severe days are identified, the locations and times of the SED reports are used to constrain the spatiotemporal domains of each day's GridRad data. Domain bounds are objectively determined using the latitudes, longitudes, and times of reports for each day's selected report types (i.e., tornado, hail, and/or wind). These space and time bounds are independently created for each report type and the final domain results from retaining the extrema of each objective report domain. First, the mean latitude and longitude to the nearest half degree of a given report type is found. An initial box encompassing this point is created with bounds $\pm$ 5 degrees longitude and latitude from the mean report location. This $10^\circ \times 10^\circ$ box then recursively expands by 0.5 degrees in all directions until $\geq 90\%$ of reports for that report type are encompassed. Using only those reports enclosed in the objectively identified box, time boundaries are created in a similar manner, starting with $\pm$ 2 hours centered on the mean report time rounded to the nearest half-hour. The time boundaries expand recursively by half-hour increments in each direction until $\geq 90\%$ of the reports within the spatial boundaries are captured (i.e., $\geq 81\%$ of the total number of reports). Note that the time bounds of a GR-S event (1200-1200 UTC) differ slightly from the SPC definition of a single day (1200-1159 UTC). Once these time boundaries are determined, they are trimmed such that the start time is no more than 30 minutes before the first report and the end time is no more than 30 minutes after the last report. This ensures that the GridRad data created are focused on the times when severe weather was occurring. Figure 2.1 shows how often a location was encompassed within a GR-S domain. Overall, the GR-S domains were generally focused on the region from 100°-80° W and 30°-45° N, offset slightly southeast of the center of the CONUS. As the year cycles from spring through winter, the GR-S domains on average shift in a clockwise manner, consistent with seasonality in the SED report data (not shown).

As stated previously, this creation of space and time bounds is done independently and objectively for each report type that exceeds its secondary threshold. For days where more than one report type exceeds its secondary threshold, both a composite spatial domain and

a composite temporal domain are created using the individual spatial and temporal domains for each report type, retaining the overall maximum and minimum latitudes, longitudes, and times to create a composite domain that encompasses all individual domains. Once these spatiotemporal bounds are determined, the time bounds are limited to begin no earlier than 15 UTC on the event day and end no later than 12 UTC the following day, to limit analyses to the 12 UTC-12 UTC period used for SED reports. We use 15 UTC instead of 12 UTC as the start time limit because all GR-S day temporal domains are then extended backward by 3 hours to attempt to capture convection initiation (CI). This is motivated by prior work such as Bluestein and Parker (1993), which found in their dryline study in Oklahoma that the time between CI and the first tornado report for a storm was approximately 2-3 hours (see their Table 2). A flowchart describing these methods is shown in Fig. 2.2, and an additional schematic showing an example GR-S case and spatiotemporal domain selection is shown in Fig. 2.3. Panel (a) shows the process of spatial domain creation, including recursive 0.5° spatial domain expansion in all directions until 90% of reports are encompassed for all report types, and creation of a composite domain using the extrema of individual domains. Panel (b) shows the process of temporal domain creation. Subpanels (i)-(iii) show hail temporal domain creation, (iv)-(vi) show tornado temporal domain creation, and (vii) shows composite domain creation. Individual steps in the temporal domain creation process are shown separately, including (i, iv) identification of mean time of reports within individual spatial boundaries rounded to the nearest half-hour, (ii, v) recursive 30 minute expansion of the temporal domain in both directions until 90% of spatially-encompassed reports are encompassed within temporal domains, and (iii, vi) identification of report-based domains after trimming domains to no earlier (later) than 30 minutes before (after) the first (last) temporally-encompassed report of that type. Subpanel (vii) shows (gray) the compositing of the two domains, and trimming such that the temporal domain starts no earlier than 15 UTC (unnecessary in this example) before (black) subsequent expansion

back by 3 hours. This is a hypothetical case and is not meant to represent any specific day within the GR-S dataset.

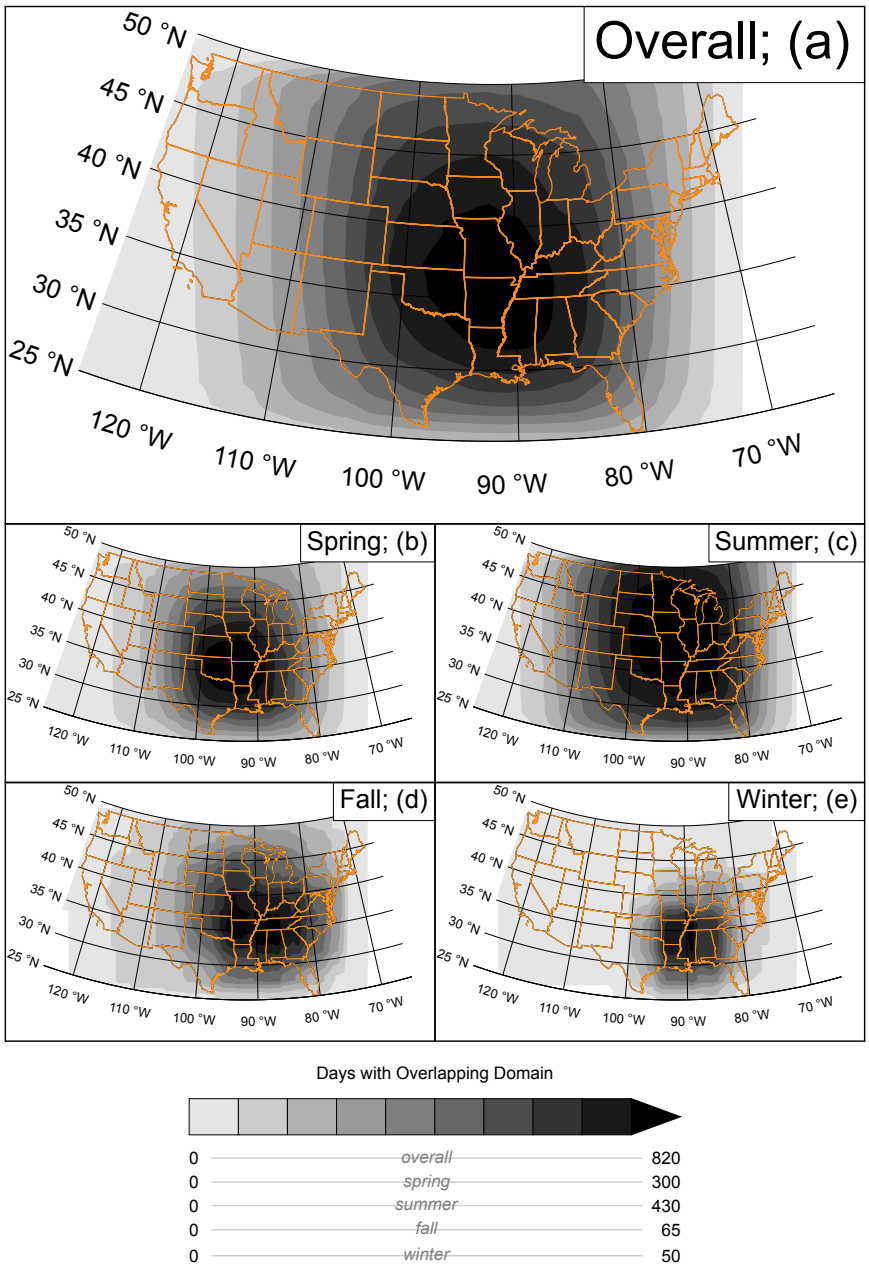

Figure 2.1: Contour plots of the number of times a point was encompassed within a GR-S domain, gridded to an approximate 80 $\times$ 80 km grid. Plots include (a) all GR-S days; (b) days in March, April, and May (spring); (c) days in June, July, and August (summer); (d) days in September, October, and November (fall); and (e) days in December, January, and February (winter). Colorbar limits are individual to each panel, and listed below the colorbar at the bottom. State borders are highlighted in orange.

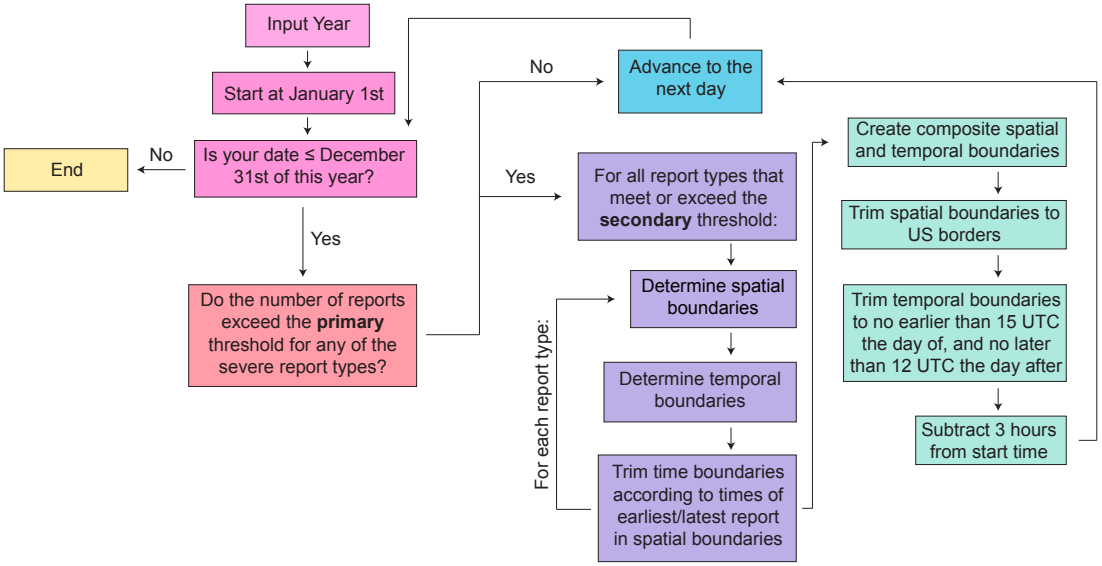

Figure 2.2: Flowchart explaining the creation of GR-S data for a given year.

2.4 GR-S storm tracks

Each GR-S event includes the 5-min GridRad volumes outlined in section 2.1 and a comma-delimited storm track file that includes official storm reports matched with each storm. The storm tracks for each event are identified using an echo-top altitude-based tracking method from Homeyer et al. (2017), with modification to resolve premature termination of tracks during storm splits and mergers as outlined in Lagerquist et al. (2020). In summary, the GridRad storm tracking algorithm identifies point locations of $Z_{\mathrm{H}} = 30$-dBZ echo-top altitude maxima and links them in time (5-min intervals for GR-S). Echo-top maxima are required to reach at least 4 km AMSL and be embedded within echoes classified as convection by the Storm Labeling in 3 Dimensions algorithm (SL3D; Starzec et al. 2017) to be tracked. Cells in subsequent time steps are linked in time if they are located within 15 km of each other (for neighboring 5-min volumes only). In cases where more than one echo-top maximum is located within 15 km of a previously defined storm, the closest one is matched during tracking. Finally, the tracking algorithm only retains tracks that are at least 15 minutes in duration–or, equivalently, are identified in at least three consecutive 5-minute GridRad volumes. To resolve storm splits and storm mergers, colinear storm tracks with closely located or overlapping initial and final locations and times are combined into one track. This combination ensures that cyclic updraft cycles, as seen commonly in supercells, are not split into multiple short-duration storm tracks. Two passes are made in this attempt to combine broken storm tracks: i) joining tracks with end and start times separated by one 5-min GridRad analysis (gap storms) so long as their end and start locations differ by $\leq$ 15 km, and ii) joining storms with start and end times that fall within one 5-min GridRad analysis of each other, so long as the minimum distance between track locations during the overlapping period is $\leq$ 15 km. In the former, the location during the gap is determined using linear interpolation between the end and start locations of the combined tracks. In the latter, tracks are combined at the closest point of coincidence during overlap such that the

point of the second (later) track through its remaining path is appended to the first (earlier) track.

After the initial storm tracking and track combination algorithms are applied to each GR-S event, the resulting 5-minute storm tracks are linearly interpolated to 1-minute resolution for spatiotemporal collocation with SED reports. To match the SED reports to the storm tracks, the closest tracked storm to a report at the report time is matched, so long as it lies within 30 km of the report location. While a maximal 30-km radius for report matching may be considered generous, it is noted that nearly all matched reports fall within 10-15 km of the objectively tracked storm centers and manual validation efforts in the past have demonstrated broad reliability of this approach (e.g., Homeyer et al. 2020). Figure 2.4 shows how storm reports are matched to storm tracks for the 14 April 2011 GR-S event. Matching reports with storm tracks in this way allows for individual storms to be classified as sub-severe or severe, and their individual characteristics examined in a bulk sense. An important limitation to accurate report matching is the spatiotemporal accuracy of the reports themselves; many studies (e.g., Trapp et al. 2006; Allen and Tippett 2015, for severe wind and hail reports, respectively) have reported on the imperfect nature of human-reported severe weather. These limitations are an important consideration in any work using storm reports as a method of validation.

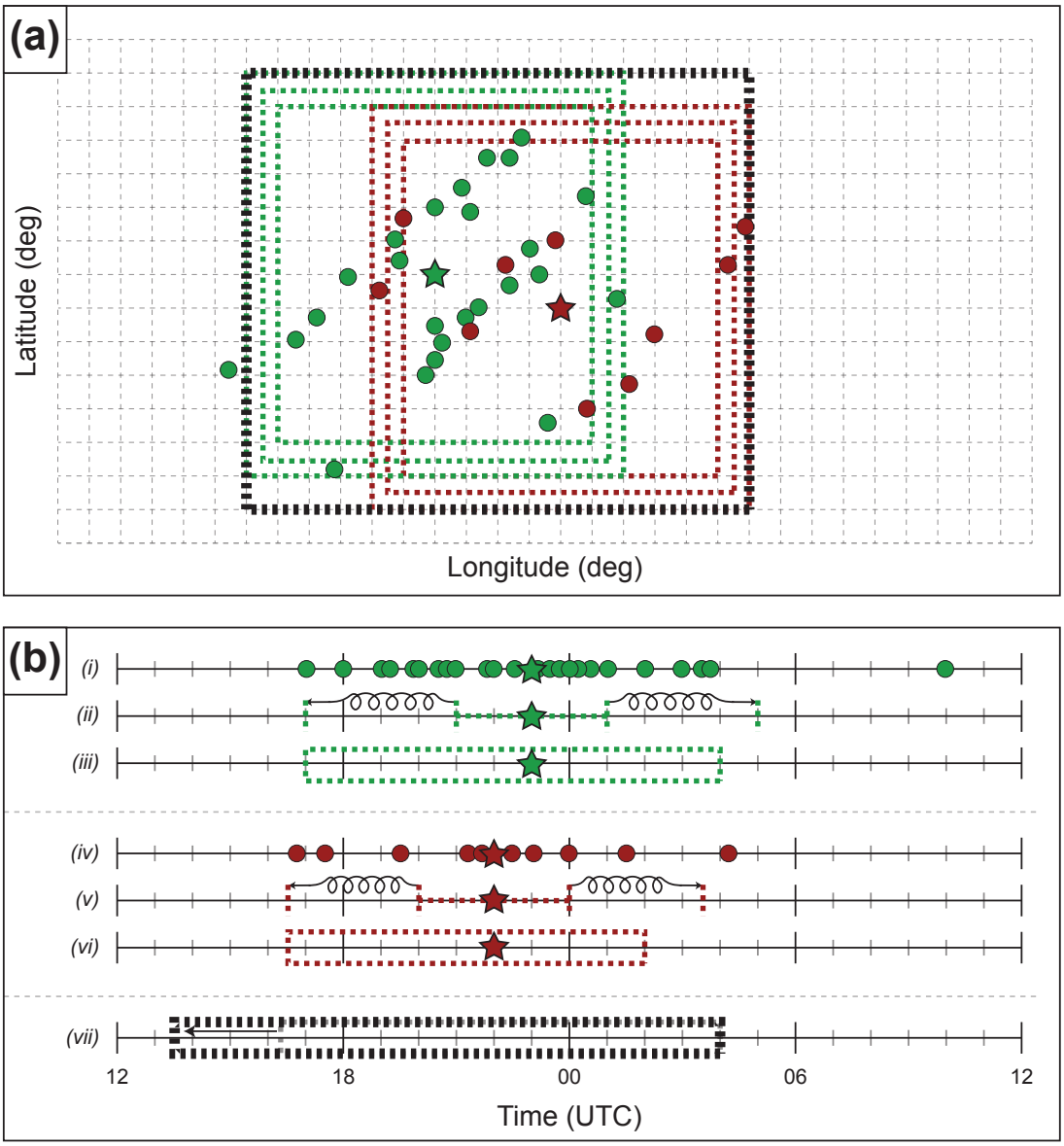

Figure 2.3: Schematic showing GR-S spatiotemporal domain selection. On all panels, red (green) features correspond to tornado (hail) report-based domain selection. Dots represent severe reports, stars represent the spatial or temporal mean location, red and green dashed boxes represent report-based domains, and black dashed boxes represent composite domains.

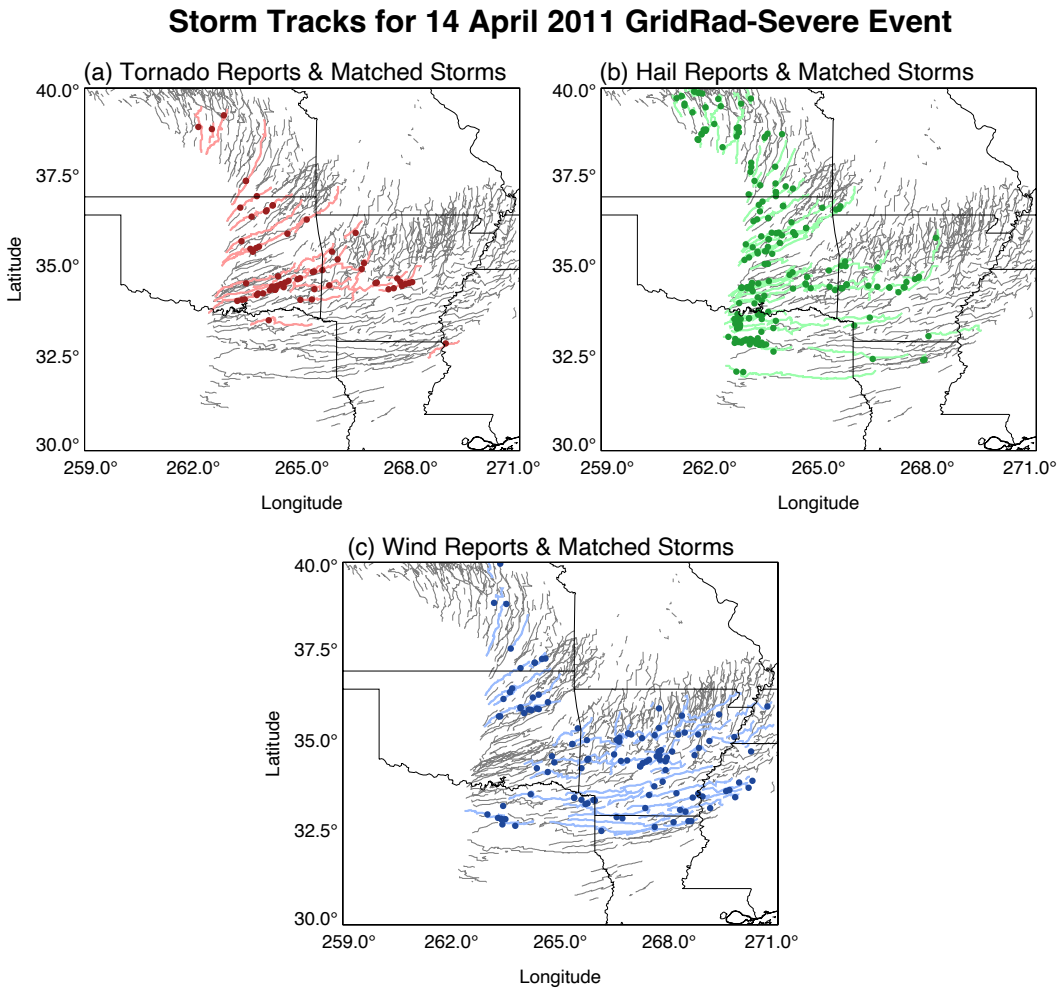

Figure 2.4: GR-S storm tracks for the 14 April 2011 event, with severe reports superimposed as dark circles and storm tracks matched with reports in a lighter shade of the same color for (a) tornadoes, (b) hail, and (c) wind. Storm tracks not matched with a report are shown in gray.

The final archived 1-minute, comma-delimited storm track files contain information on storm location, motion, radar characteristics, and storm-matched severe reports (summarized in Table 2.1). In the case of severe hail or wind reports, the event information is linked to the midpoint time of the report so that no report gets counted more than once. For tornado reports, most of the report information is linked to the initial report time (i.e., tornadogenesis).

While enabling unique analyses of the GR-S data, limitations to the objective storm tracking methods do exist. Common drawbacks include unresolved storm splits and mergers and poorly tracked initiation phases of some storms (before $Z_H = 30$ dBZ exists and/or before the 30-dBZ echo top exceeds 4 km AMSL, which is typically <15 minutes in severe storms). In particular, not resolving the initiation of storms until they reach a threshold strength prevents full analysis of storm development, potentially missing important signatures present in early storm development that may indicate the future severity of the storm.

2.5 Storm mode classification

Given that many past studies have classified storm mode subjectively, such efforts have often focused on small spatiotemporal domains. An objective method for storm mode classification is presented herein for use with the GR-S dataset. This storm mode classification relies on closed radar echo contours above a given Z_H threshold encompassing objectively tracked storms to be able to classify each track within a contour as a part of a single cell storm, multicell storm, or an MCS. By classifying the mode of each objectively tracked storm, characteristics of each storm can be analyzed in concert with storm-matched SED reports to potentially link storm mode and storm-scale characteristics with the presence of (or lack thereof) severe weather.

To classify storm mode, 30-dBZ contours are identified using column-maximum reflectivity (Z_{Hmax}) from each 5-min GridRad data file. For each closed contour, the area and maximum dimension are calculated, and the number of tracked storms within each contour is counted. Tracked storms are then identified as single (i.e., discrete) cell if either 1) only a single tracked storm exists within a contour, or 2) a relatively small contour (< 3000 km^2) encompasses no more than two tracked storms. Two tracked storms within a small contour are both classified as single cell to account for storms with cyclic updraft generation (e.g., supercells) that can have more than one updraft at a single time, but do not persist as such.

Homeyer et al. (2020) found in their composite analyses of supercells that the average supercell size for Z_{Hmax} = 30 dBZ is approximately 3000 km^2 (their Fig. 3). The 3000 km^2 contour area threshold is therefore chosen to encompass the size of some of the largest single cell storms expected. For larger contours ($\geq$ 3000 km^2) containing 2 tracked storms or any contour containing 3 or more tracked storms, those storms are classified as either multicell or MCS. The delineation between the two is made using the contour maximum dimension; if the maximum dimension is $\geq$ 100 km, it is classified as an MCS (Houze Jr. 2004).

Since the storm mode classification algorithm is independently run on each time step, a tracked storm may have a complex, time-varying storm mode classification. We do not employ a minimum time that a tracked storm has to be identified as a consistent storm mode to retain such a classification. This is done because, instead of classifying a storm based on its mean storm mode, we want to enable investigation into how each storm evolves in time and, if possible, how changes in its severity accompany changes in storm mode. Figure 2.5 shows select times during the evolution of the 14 April 2011 GR-S event in the southern plains. Over time, many single cell storms that initiated early in the event merge into what eventually becomes a large MCS. In concert with matched storm reports (e.g., Fig. 2.4), storm mode can be linked with a storm's severity for bulk analyses of the two characteristics.

Information in GR-S Event Storm Track Files	
Storm Information	Report Information*
Storm Number	Binary Report Flag
Storm Date & Time	Report Number
Storm Longitude	Report Longitude
Storm Latitude	Report Latitude
Eastward Storm Motion	Report Magnitude
Northward Storm Motion	Instantaneous Tornado Count
Echo Top Altitudes	Max Instantaneous Tornado Rating
Column-Maximum Z_{H}	Tornado End Date & Time
	Tornado Width
	Tornado Length

Table 2.1: Storm attributes included in GR-S event track files, separated for radar-based storm information (left column) and matched SED report information (right column). *Report flag (0 or 1), number, longitude, latitude, and magnitude are given for all report types (tornado, hail, and wind), where for tornadoes the report information is listed only at the time of tornadogenesis. The additional SED report information for tornado end time, width, and length are also only given at the time of tornadogenesis, while the instantaneous tornado count and maximum rating are based on all reports valid at each 1-min storm track time.

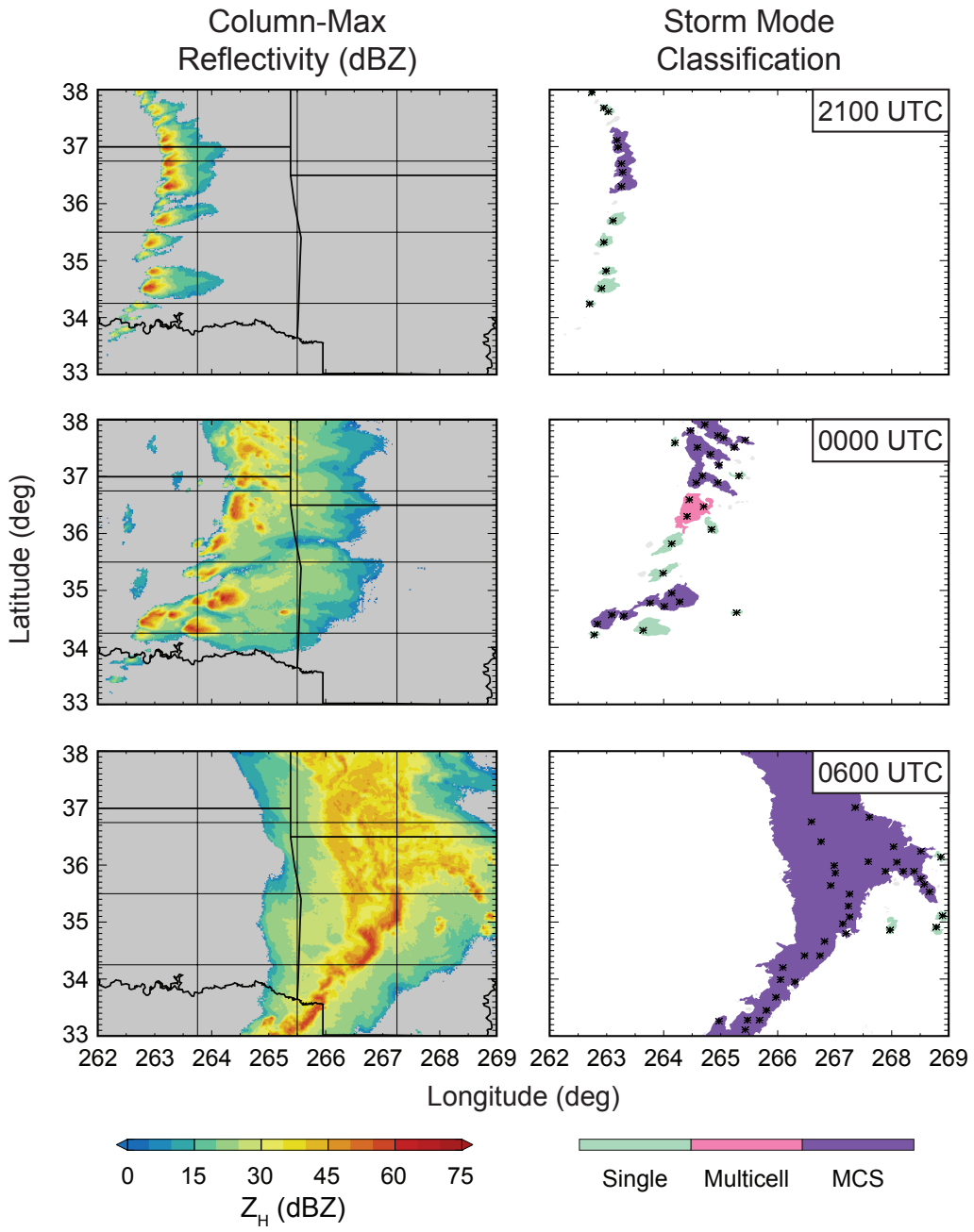

Figure 2.5: Select images of (left) column-max reflectivity and (right) storm mode classification from a limited spatial domain within the 14-15 April 2011 GR-S case. For storm mode classification images, the interior of identified 30-dBZ contours are shaded in green, pink, or purple to denote single cell, multicell, or MCS classification, respectively. Tracked storms within such contours are indicated by black asterisks. Identified 30-dBZ contours that do not encompass any storm tracks are shaded in light gray.

2.6 Mesocyclonic updraft classification

Another important severe storm characteristic is whether or not a storm's updraft was rotating when the storm produced a severe report. Updrafts that have sufficient persistent rotation are defined herein as mesocyclonic, and otherwise as non-mesocyclonic. Mesocyclonic updrafts are classified using the methods for right-moving supercell identification outlined in Homeyer et al. (2020), originally based on work by Sandmæl (2017). Namely, five criteria are used to objectively identify updrafts as mesocyclonic: 1) maximum midlevel (4-7 km AMSL) azimuthal shear exceeds $4 * 10^{-3}$ s^{-1} for at least 40 minutes; 2) maximum midlevel azimuthal shear meets or exceeds $5 * 10^{-3}$ s^{-1}; 3) maximum column-max azimuthal shear meets or exceeds $7 * 10^{-3}$ s^{-1}; 4) maximum column-max radial divergence meets or exceeds $1 * 10^{-2}$ s^{-1}; and 5) maximum column-max velocity spectrum width meets or exceeds 13 m s^{-1}. The sixth criterion used in Homeyer et al. (2020)–max 40-dBZ echo top altitude meets or exceeds 11 km–was not applied in this study. This was done to enable reliable classification of wintertime convection, which often has lower echo tops.

It is important to note that the mesocyclonic updraft classification criteria were based on right-moving supercell identification criteria, so left-moving (mesoanticyclonic) storms are not independently examined herein. Both anecdotal evidence and prior research show a dearth of left-moving supercells compared to right-movers. Bunkers et al. (2006) examined long-lived supercells and found that, of 184 long-lived supercells in their dataset, only 4 were left-movers. An approximate ratio of left- to right-moving supercells is, to our knowledge, unknown. Future work may focus on classification of left-moving supercells using GR-S or an alternative dataset and further investigation of their characteristics compared to right-movers.

The result of these classification techniques is that each individual tracked storm has a time-varying storm mode classification and a binary mesocyclonic updraft classification. This does mean that, for example, cells that otherwise have a single predominant storm

mode may have varying classifications throughout their lifetimes. For those interested in examining the predominant storm mode of a given cell, statistical examinations of the prevailing storm mode (such as examining the median or mode) of the high-frequency and time-varying storm mode classifications can be performed. Investigation of overarching storm mode classifications will be useful (and will be further discussed) in Chapters 4-6. Additionally, this mesocyclonic updraft classification scheme can, and does, result in classifications such as multicell or MCS storms with mesocyclonic updrafts. Examples of these types of storms may include supercells that share the same 30-dBZ precipitation shield, mergers of a discrete supercell with a QLCS, supercells present in the early organizing stages of a QLCS (e.g., Weisman and Trapp 2003) or generation of vortices that meet mesocylonic rotation criteria (e.g., DeWald and Funk 2002). The identification of multicells and storms within MCSs that meet mesocyclonic updraft criteria in GR-S data allow for filtering of these storms if desired. Herein, these storms and their attendant severe weather are retained and examined in Section 3.1, but are removed in Section 3.2 where GR-S MCS storms that produced severe weather are compared to objectively identified severe MCSs and QLCSs. While the presence of supercell-like updrafts within MCSs has been noted in the literature (e.g., Ashley et al. 2023), care should be taken when using these data to compare to past studies that may or may not have included embedded supercells in their analyses.

Chapter 3

Statistical overview of GridRad-Severe data[1]

3.1 Results

In order for conclusions in this study to be representative of the total climatology of severe weather and therefore broadly applicable, the distribution of GR-S storm matched reports must be representative of the complete SED database. Representativeness herein includes capturing not only the majority of SED reports, but also the correct spatiotemporal distribution. Looking at only the number of total SED reports vs. GR-S storm-matched reports, Table 3.1 summarizes the percent of severe and significant severe reports captured by GR-S for 2010-2019. Of all SED reports during those ten years, the GR-S data retains ∼63–77% of total reports and ∼68–91% of significant severe reports, with percent matched highest for tornado reports and lowest for wind reports. Focusing only on SED reports that existed within the spatiotemporal bounds of the GR-S domains, the range of retained reports increases to ∼88–94% for all reports and ∼91–98% for significant severe reports, again with percent matched highest for tornado reports and lowest for wind reports. This means that for all SED reports within the spatiotemporal bounds of the GR-S domains, the GR-S storm tracking and report matching procedure matches approximately 9 out of every 10 reports to a storm. Examining the total number of reports captured, the GR-S database retains 164748 out of the total 249600 SED reports during the 10-year period (∼66%). This is expected since the GR-S database only includes ∼100 days per year, with data only within limited spatiotemporal domains. However, it is encouraging that, if a report exists within a GR-S domain, it is highly likely that it will be matched with a GR-S tracked storm. Therefore,

[1]The work within this chapter has been published in the peer-reviewed literature as Murphy et al. (2023).

the domain selection criteria coupled with the matching algorithm are both capturing a majority of SED reports and effectively matching reports within GR-S bounds to objectively tracked storms.

Comparison of SED and GR-S Storm-Matched Reports			
Report Type	(1) % of SED Total	(2) % of SED within GR-S Bounds	(3) Total GR-S Reports
Tornado Initiations	76.83%	94.16%	10542
Hail Reports	70.60%	94.33%	56025
Wind Reports	62.72%	87.81%	98181
Sig. Tornado Initiations	90.85%	97.69%	1608
Sig. Hail Reports	79.85%	94.68%	5842
Sig. Wind Reports	67.80%	90.56%	7179

Table 3.1: Comparison of SED and GR-S matched reports from 2010-2019. For each tornado/hail/wind, data include (1) percent of GR-S matched reports compared to all SED reports of that type and over that period; (2) percent of GR-S matched reports compared to SED reports of that type and over that period, confined within the corresponding day's GR-S spatiotemporal bounds; and (3) the total GR-S matched reports in the dataset. Data are also shown isolating significant severe reports.

Capturing the majority of severe reports is only one facet of examining the representativeness of the GR-S data. Equally important is the distribution of the reports–spatial and temporal, for both severe and significantly severe events–and whether those distributions match the full SED report climatology. Figure 3.1 shows the breakdown of total reports by month for both GR-S matched reports and SED reports, with lines showing the percent contribution of tornado/hail/wind reports to the total reports in each month. Tornado

data in this figure and for all future analyses are focused on the time of tornado initiation. The difference in total data points represented on each plot is captured by the y-axes, which show the mean 34% decrease in reports when comparing SED to GR-S data. The monthly distribution of reports is similar between the GR-S and SED data, although month-to-month variations in percent of SED reports captured within GR-S are visible. Percent differences between GR-S and SED reports per month range from 11-56%, with some of the lowest percent differences in the late spring and early summer (below 19% from April to June, inclusive). This is potentially due to a preference for higher-end severe days to occur in the spring to early summer, so a higher percentage of all severe weather days in that period would be captured by the GR-S domain selection criteria. Notably, April through June alone make up 51.8% of all GR-S days in this dataset. If only SED reports within the GR-S bounds are considered (not shown), those percent differences range from 8-28%, and are lowest in the spring and summer (8-11%) and highest in fall and winter (11-28%). Therefore, the storm tracking technique is most effective at tracking severe convection and matching reports to those storms during the maximum of the annual cycle in severe weather (the early-mid warm season) and least effective during the the cool season. Despite these differences, the GR-S bar graph still closely resembles the SED bar graph, showing that GR-S is capturing the overall distribution of severe reports quite well. The percent contribution of tornadoes, hail, and wind to each month's overall report count also show very similar values between the GR-S and SED data, demonstrating a monthly GR-S report type balance that is representative of the underlying SED report data. Overall, this analysis shows that the GR-S database captures the annual cycle of all SED reports well.

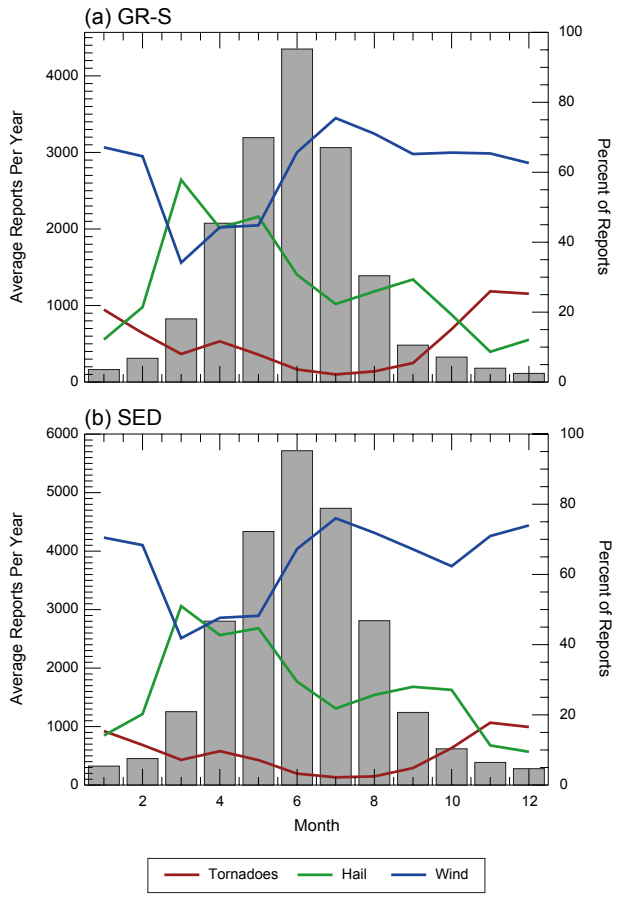

Figure 3.1: Comparison of (a) GR-S and (b) SED average annual storm reports, broken down by month for 2010-2019. Lines indicate the percent contribution of various severe report types (tornado, hail, and wind) to the total number of reports in a given month. Pearson correlation coefficients comparing GR-S and SED lines for each severe hazard exceed 0.95.

In addition to the annual cycle of reports, capturing the spatial distribution of reports is also very important. Figure 3.2 shows the gridded number of all SED reports from 2010-2019 for each severe report type juxtaposed with the gridded number of reports retained in the GR-S storm tracks. More reports are expected on the SED maps, again since GR-S events only include $\sim$100 days per year of severe weather. Qualitatively, maps of GR-S matched reports and SED reports have similar spatial distributions, confirming that the distribution of reported severe weather is well captured in the GR-S database. In combination with Table 3.1, Fig. 3.2 gives confidence that GR-S is capturing the majority of reports in a consistent way across most of the CONUS. This can be more directly examined in Fig. 3.3, which shows both the total report difference between the GR-S database and the SED record and the percent difference between the two. These plots further demonstrate that reports are well matched across the eastern two-thirds of the country where reports are more frequent (Fig. 3.2). The areas with high percent differences are commonly found in locations with low report counts for both GR-S and SED (e.g., compare total number of reports in Fig. 3.2f to percent difference in Fig. 3.3f for severe wind in the western CONUS), meaning that while the percent of reports missed in these regions may be quite high, the total number of reports missed is quite low. Therefore, based on the results shown in Figs. 3.1–3.3, we can confidently say that the storm-matched reports within the GR-S database are a representative sample of the total climatology within the SED dataset.

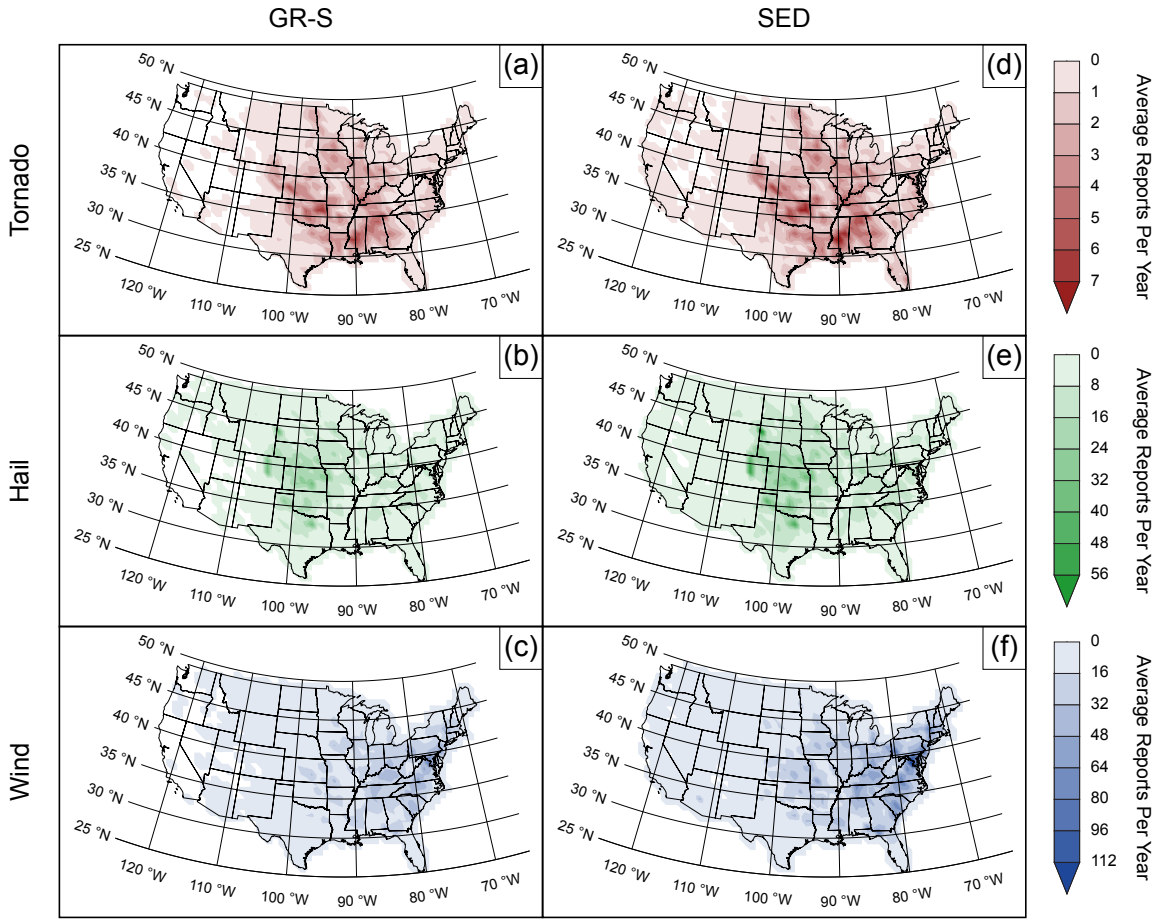

Figure 3.2: Contour plots of average annual number of (a-c) GR-S and (d-f) SED reports for (a,d) tornado, (b,e) severe hail, and (c,f) severe wind reports from 2010-2019, inclusive. Data are gridded on an approximately 80 $\times$ 80 km grid. Areas not shaded indicate no (a-c) GR-S or (d-f) SED reports.

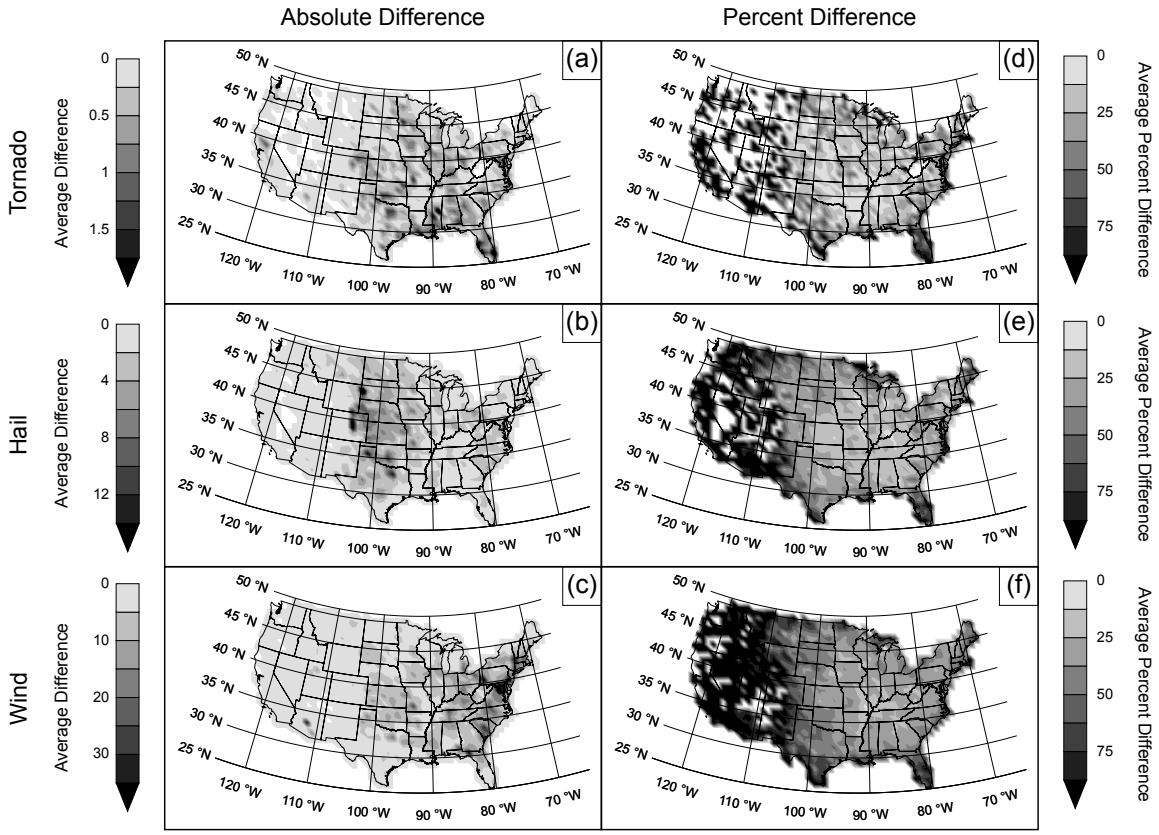

Figure 3.3: Contour plots of average annual (a-c) absolute difference and (d-f) percent difference between GR-S and SED reports, for (a,d) tornado, (b,e) severe hail, and (c,f) severe wind reports from 2010-2019, inclusive. Data are gridded on an approximately 80 $\times$ 80 km grid. Areas not shaded indicate either no reports (both SED and GR-S) or an equal number of SED and GR-S reports.

Confident that the GR-S database is representative of the seasonality and geographic distribution of SED reports, we can use GR-S data to examine other bulk aspects of storm severity. Identifying both the storm mode and updraft type (whether mesocyclonic or non-mesocyclonic) associated with each tracked storm and its matched reports can provide valuable insight into the types of storms that produce various severe phenomena. Figure 3.4 shows the average annual number of tornado, hail, and wind reports per month for 2010-2019, broken into sub-significant and significant severe reports. Overlaid are lines showing the percent of reports per month that were matched with storms classified as single cell, multicell, or MCS storms, and whether or not the storm had a mesocyclonic updraft. Perhaps the most surprising result from this analysis is how often tornadoes are associated with MCS-classified cells throughout the year (Fig. 3.4a). However, the mesocyclonic classfication reveals that many of the cells classified as MCS-type are dynamically consistent with supercell storms rather than the typical non-mesocyclonic cells often found in an MCS. We speculate that this may be driven by the reliance on a relatively low Z_{H} threshold to define storm contours during storm mode classification ($Z_{\mathrm{H}} = 30$ dBZ), which may encompass the precipitation shield of neighboring–and otherwise mostly discrete–storms (supercell or otherwise) and classify those storms as part of an MCS. The lower Z_{H} threshold to define storm contours is important to appropriately resolve strong storms that occur outside of the traditional severe weather season; the potential misclassification of discrete storms sharing a precipitation shield as an MCS or multicell cluster is a known limitation of the dataset and should be taken into account when examining MCSs or multicell clusters within GR-S to avoid contamination of the data by misclassified storms. The contribution of non-mesocyclonic (i.e., more traditional) MCS cells to tornado reports reaches a minimum in the spring and summer, when both the number of tornadoes peaks and the classical U.S. tornado season occurs. In contrast, the contribution of mesocyclonic storms to tornado reports, regardless of storm mode classification, peaks during this time.

Examining hail events (Fig. 3.4b), for much of the spring through fall, single cell storms account for the largest fraction of reports by storm mode. However, while tornado and wind reports are fairly dominated by one storm mode (MCSs are associated with a majority of reports in 11 of 12 months for each tornadoes and wind), no one storm mode stands out as a consistent majority contributor to all hail reports. Single cell storms account for slightly more than 50% of reports in July and August; multicellular storms are never associated with the majority (or even a relative majority) of reports per month; and MCS storms make up a majority (50-64%) of reports only in the winter months, where total matched reports are lowest. However, examining mesocyclonic vs. non-mesocyclonic storms, mesocyclonic storms account for a majority of hail reports year-round. MCS storms contribute the most to wind reports year-round when compared to other storm modes (Fig. 3.4c), with a peak in single cell and multicell contribution in the late summer. This is when "severe weakly forced thunderstorms" (Miller and Mote 2017) are most common in the CONUS, which can cause downdraft-driven severe wind gusts (e.g., microbursts).

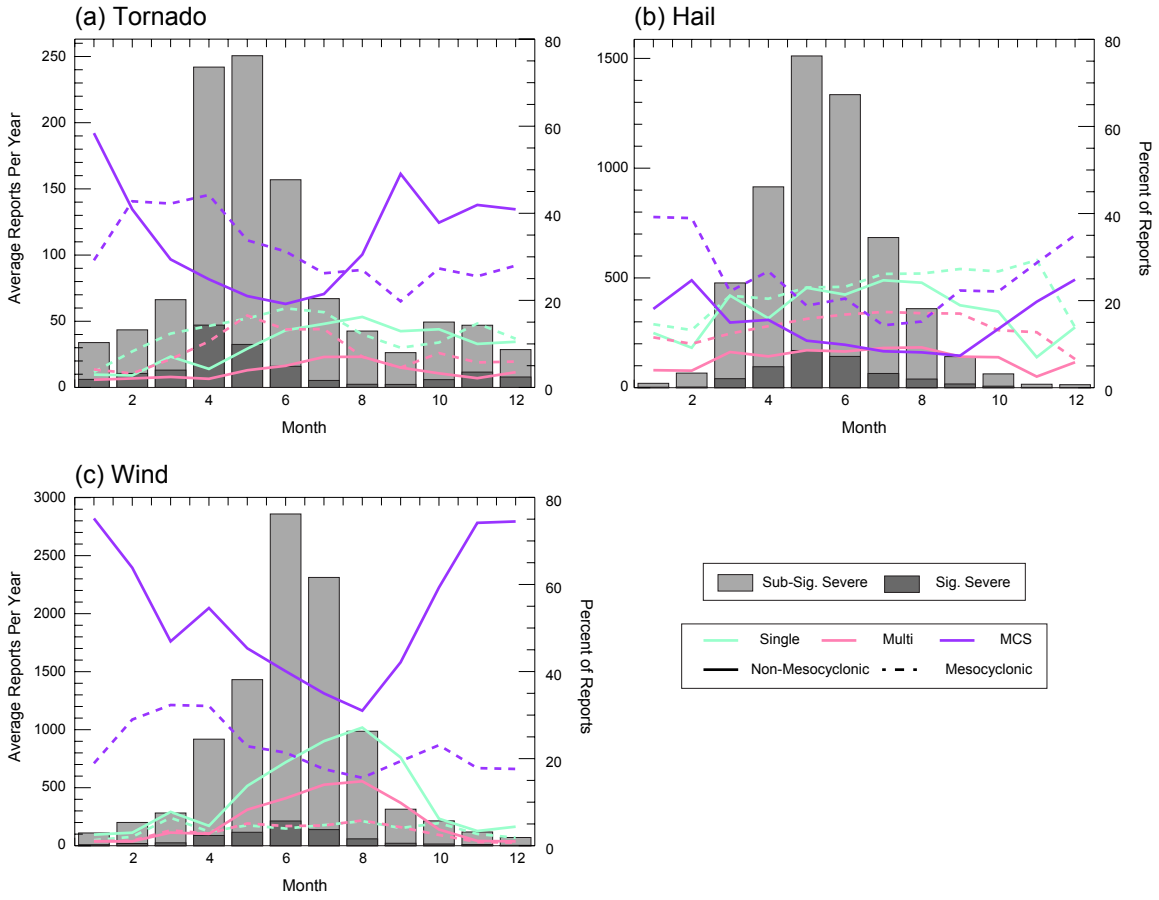

Figure 3.4: Average annual number of GR-S matched (a) tornado, (b) severe hail, and (c) severe wind reports by month for 2010-2019. Lines show the percent of total reports that were matched to a storm of a given storm mode (single cell, multicell, or MCS; in green, pink, and purple, respectively) and with a given updraft type (non-mesocyclonic or mesocyclonic; in solid and dashed lines, respectively). Bar charts are split into lighter and darker gray, which show sub-significant severe and significant severe reports, respectively. The vertical extent of these two bars combined is the total number of all reports, and the sum of all lines in a given month is 100%.

Similar to the annual cycle analysis, Fig. 3.5 reveals the average diurnal cycle of each hazard, relative to the reports' local solar noon. Solar noon is the time the sun aligns with a location's meridian, and using time relative to solar noon (as opposed to UTC time) eliminates the effect of time zones, providing a uniform representation of local time. Each severe report type has a pronounced diurnal cycle, with a peak in report frequency between approximately 2 and 8 hours after solar noon. During local nighttime, severe weather is associated most with MCS storm cells. As tornado reports increase in frequency after solar noon, the overall fraction of reports associated with MCSs drops (Fig. 3.5a). From 2 to 12 hours after solar noon, the majority of tornado reports for each of the three storm modes are produced from mesocyclonic storms. The hail data (Fig. 3.5b) show a more pronounced diurnal cycle when compared to tornadic and wind reports, with hail reports highly concentrated around their daily peak at 4-5 hours after solar noon. Single cell storms are the main contributor to hail reports in the first 7 hours after solar noon, with the contributions of multicellular and MCS storms nearly equal during those hours ($\sim$25%). Overall, mesocyclonic storms account for the majority of hail reports during 23 of the 24 total hours. As was true for the annual cycle, diurnal data show wind reports overwhelmingly associated with MCS storms throughout the majority of the day (Fig. 3.5c). MCSs are known to be prolific producers of damaging straight-line winds, so it is no surprise that these storms produce the most wind reports (a minimum of 48% of all wind reports each hour). Non-mesocyclonic single cell and multicell severe wind-producing storms have a pronounced peak in the hours after solar noon, which, as previously mentioned, is likely attributable to downdraft-driven wind gusts from severe weakly forced thunderstorms that are common in the late summer and early afternoon. Notably, mesocyclonic storms never account for a majority of severe wind reports throughout the day ($< 39\%$ of reports per hour), in contrast to how often they contribute to severe hail and tornado reports during the peak tornado- and hail-producing hours. For all analyses of the diurnal cycle of reports, note that the

total number of reports may be lower during the overnight hours given that most people are asleep (e.g., Wendt and Jirak 2021).

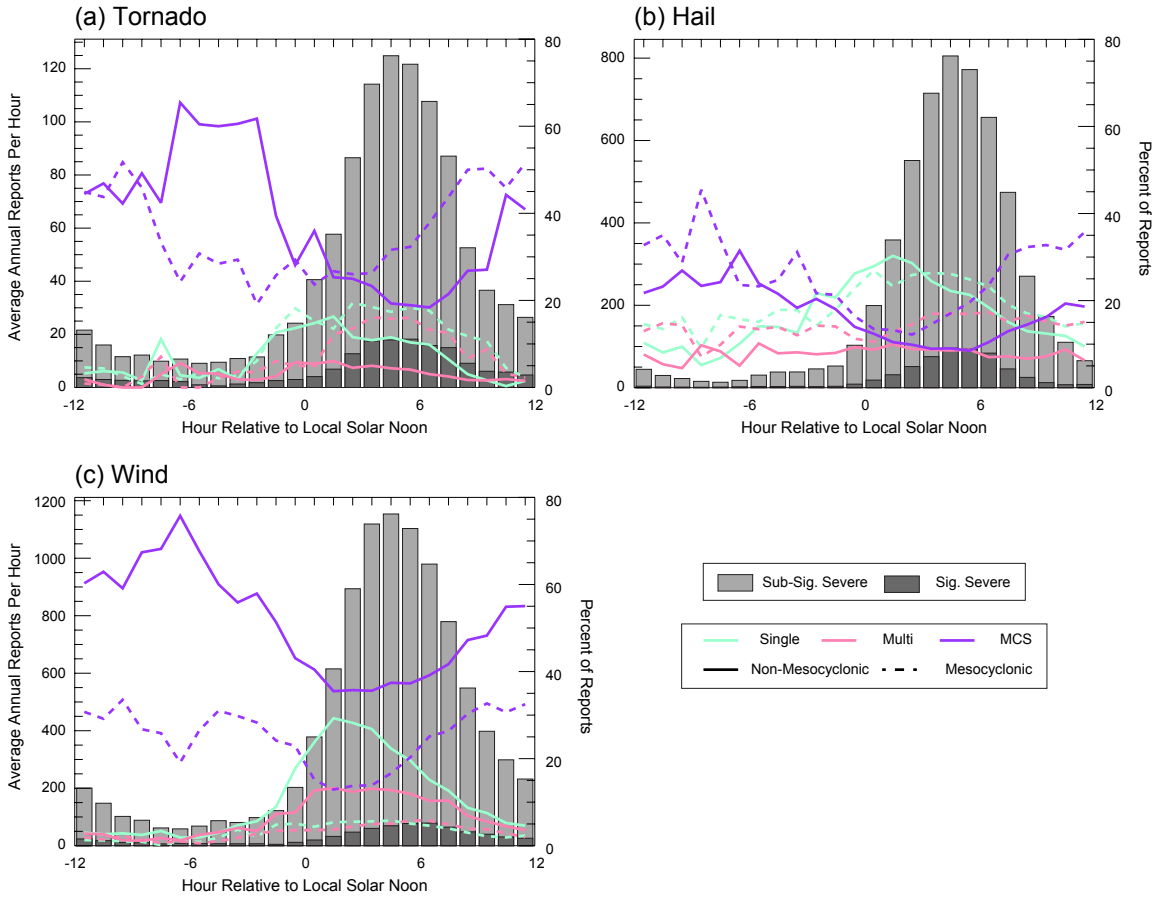

Figure 3.5: Average annual number of GR-S matched (a) tornado, (b) severe hail, and (c) severe wind reports by hour relative to local solar noon for 2010-2019. Lines show the percent of total reports matched to a storm of a given storm mode (single cell, multicell, or MCS; in green, pink, and purple, respectively) and with a given updraft type (non-mesocyclonic or mesocyclonic; in solid and dashed lines, respectively). Bar charts are split into lighter and darker gray, which show sub-significant severe and significant severe reports, respectively. The vertical extent of these two bars combined is the total number of all reports, and the sum of all lines in a given hour is 100%.

Finally, we can examine how storms of various mesocyclonic or non-mesocyclonic classifications and modes contribute to reports of varying magnitude. Figure 3.6 shows histograms of the magnitude of each report type and the fractional contribution of storms of a given storm mode and mesocyclonic/non-mesocyclonic classification. The data show that the vast majority of reports ($> 90\%$) are sub-significant severe (i.e., below EF-2, 2", or 65 kts for tornado, hail, and wind reports, respectively). As EF rating increases, the relative contribution of mesocyclonic storms also increases, to the point where they are responsible for 90.5% and 100% of all EF-4 (84 total) and EF-5 (12 total) tornado reports in the database, respectively. In fact, for any EF rating, mesocyclonic storms account for the majority of tornadoes. GR-S data also show that EF-0, EF-1, and EF-2 tornadoes all predominantly come from cells embedded within MCSs (53.7, 73.4, and 65.6% of tornadoes, respectively). Focusing on significant tornadoes, the percent of tornadoes linked to MCS-classified storms decreases from 65.6% to 33.3% as the percent linked to single cell storms increases to a maximum of 58.3% for EF-5 tornadoes. Mesocyclonic storms are also the main contributor to hail reports, and hailstones are more likely to be associated with mesocyclonic storms as hail size increases. As was evident in Figs. 3.4 and 3.5, no one storm mode clearly dominates hail production. Interestingly, single cell storms account for a majority of hailstones in the lowest 3 bins, but contributions to 4"+ hailstones are relatively equal across storm modes. This comes with the caveat that the largest hailstone bin contains 0.7% of the total reports in the smallest hailstone bin. Wind reports become increasingly associated with mesocyclonic storms as wind speed increases, with the exception of the strongest winds in the dataset. These winds (95+ kts) are predominately from non-mesocyclonic storms (52.7%), in stark contrast with the strongest tornadoes and largest hail, which are overwhelmingly associated with mesocyclonic storms. Wind reports are largely dominated by storms embedded within MCSs, with 62.9-81.3% of reports in each bin attributed to MCSs. As touched on in the discussion of hail-producing storms, it is important to recognize that as EF rating, hail size, and wind speed increase, the sample size of

reports decreases. Therefore, interpretation of mode and mesocyclonic/non-mesocyclonic breakdown must be done carefully, especially where sample sizes drop below a few hundred reports.

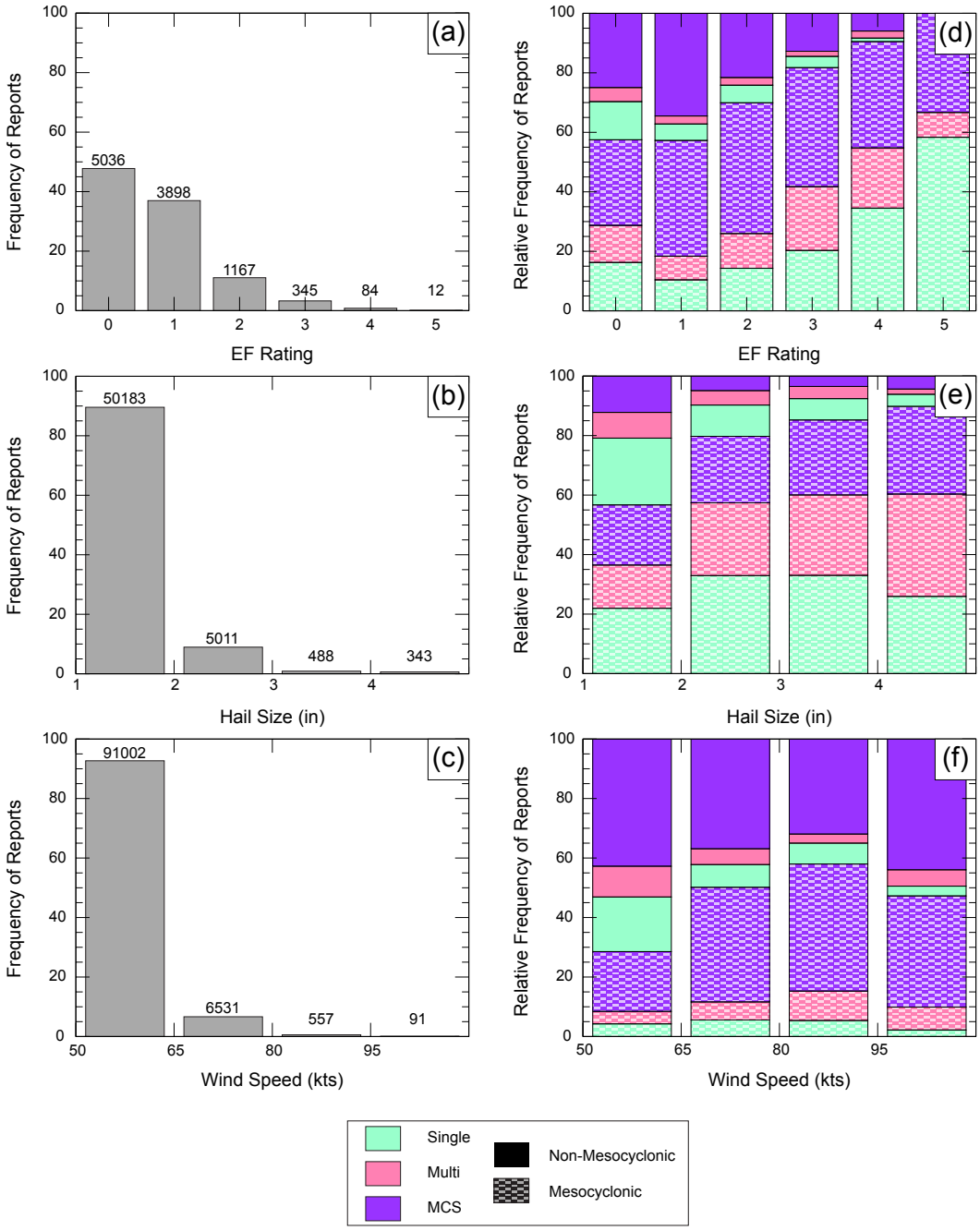

Figure 3.6: Breakdown of (left; a-c) total frequency of reports by (a) EF rating, (b) hail size, and (c) wind speed, as well as (right; d-f) the percent of reports associated with a given combination of storm mode and supercell classification by (d) EF rating, (e) hail size, and (f) wind speed. On (a-c), the total number of reports in each bin are listed on top of the individual bars.

3.2 Comparison to past studies

Given that GR-S storm-matched reports were demonstrated to be a representative sample of SED reports from 2010-2019, we can also use GR-S data to revisit analyses and conclusions from prior papers to assess reproducibility. Herein, we focus on two studies, Trapp et al. (2005) and Ashley et al. (2019), which examined the prevalence of various storm modes and their propensity to produce severe weather. Table 3.2 lists basic information about data sources and methods employed in the papers, as well as a summary of those used in the present study.

For comparisons between GR-S data and prior results focusing on MCS or QLCS storms, we will compare the papers' findings to only our non-mesocyclonic MCS cells. Comparing strictly our non-mesocyclonic MCS data to other studies' full MCS datasets resulted in greater consistency, potentially pointing to mesocyclonic MCS cells being more dynamically consistent with single cell mesocyclonic storms than with non-mesocyclonic MCS storms. This result is relevant to any future work using GR-S MCS data with the storm mode classification employed here. Also important to note is the delineation between an MCS and a QLCS. As discussed in Schumacher and Rasmussen (2020), a QLCS is a subset of the MCS archetype. While MCSs are typically defined as convective complexes with a maximum dimension $\geq$100 km, a QLCS is an MCS further characterized by an aspect ratio around 3:1, meaning that the system has one long and one short dimension. Given that the terms MCS and QLCS are often conflated, it is important to keep in mind the true nature of MCS-classified cells herein and how they may or may not be a part of a QLCS-type convective complex.

GR-S data are first compared to select conclusions from Trapp et al. (2005). The study's main goal was to "estimate the percentage of U.S. tornadoes that are spawned annually by squall lines and bow echoes, or quasi-linear convective systems (QLCSs)" using subjectively classified radar echoes over a three-year period (1998-2000, inclusive). Classification was done for QLCS and individual cells near the time of tornadogenesis per Table

3.2. They delineated between QLCS and cell type echoes based on "dynamics unique to these phenomena" and mentioned that, while tornadoes can form by mesocyclonic and non-mesocyclonic means, any distinction between cells producing tornadoes via these two different mechanisms was not investigated therein. Their final dataset included 3828 tornadoes.

Figure 3.7 shows reproductions of Trapp et al. (2005) Figs. 3b (Fig. 3.7a), 6 (Fig. 3.7b), and 8a (Fig. 3.7c) using GR-S data. Figure 3.7a shows the breakdown of the number of tornado reports by EF rating on a logarithmic scale for both single cell and MCS storms. Crucially, both MCS mesocyclonic and non-mesocyclonic lines are shifted such that they have an equal number of EF-2 reports per storm type; Fig. 3.7a therefore emphasizes the relative distributions of tornado intensity by storm type rather than absolute values. Trapp et al. (2005) found that there "appear to be disproportionately more F1 tornadoes from QLCSs, and more F3–F4 tornadoes from cells." Figure 3.7a shows this as well, where the non-mesocyclonic MCS cell line is above the single cell line for EF-1 tornadoes and below the cell line for EF 3-4 tornadoes. Here the mesocyclonic MCS curve more closely matches the single cell curve, pointing again to their dynamical similarities. On the other end of the spectrum, data from Trapp et al. (2005) (GR-S) show no F5 QLCS (EF-5 non-mesocyclonic MCS) tornadoes given they are quite rare, and also note that F5 (EF-5) tornadoes only comprise 0.2% (0.11%) of the total dataset. Trapp et al. (2005) also noted that their QLCS curve was fairly log-linear except for F0 tornadoes, potentially attributable to underreporting of the weakest tornadoes. This same linear shape, along with relatively low EF-0 tornado counts, is visible in the GR-S non-mesocyclonic MCS curve.

Figure 3.7b shows the cumulative distribution of all tornado reports broken down by month and storm type. Trapp et al. (2005) found that 32% of all QLCS tornado reports occurred within the first three months of the year, compared to just 14% of single cell reports. The lower relative fraction of cell reports compared to QLCS reports is mirrored with the GR-S data, with tornadoes in January-March making up 20% of non-mesocyclonic

MCS tornadoes and just 8% of single cell tornadoes. Therefore, with both datasets, a higher proportion of annual MCS tornadoes occurred in the first three months of the year compared to the proportion of annual single cell tornadoes. Finally, Fig. 3.7c shows the diurnal cycle of tornado reports, using time relative to local solar noon. Trapp et al. (2005) found that cell reports peaked close to 18 local standard time (LST), with a similar albeit smaller peak in QLCS data near 18 LST. GR-S data show similar trends, with peaks in single cell and non-mesocyclonic MCS data between 3-5 hours after solar noon, and a higher peak for single cell than non-mesocyclonic MCS data (i.e., a more amplified diurnal cycle). Mesocyclonic MCS cells are again more consistent with single cell storms, with a $\sim$2 hour offset in their diurnal cycle compared to the single cell data.

Storm Mode Classification Techniques			
Field	Trapp et al. (2005)	Ashley et al. (2019)	This Study
Years:	1998–2000	1996–2017	2010–2019
Technique:	Hand Analysis	Machine Learning	Objective Analysis
Z_H Data:	Composite column-maximum images from NCDC (NCEI), other sources	NOWrad composite reflectivity data (Grassotti et al. 2003)	GR-S column-maximum data
Cell Classification:	Relatively isolated, circular or elliptical in shape, with $Z_{Hmax} \geq \sim$50 dBZ	—-	1 track in 30-dBZ Z_{Hmax} contour, or 2 tracks within 30-dBZ Z_{Hmax} contour $<$ 3000 km^2
MCS Classification:	—-	Region of $Z_{Hmax} \geq$ 40 dBZ persisting for at least 3 hours, with contiguous to semi-contiguous 40-dBZ contour maximum dimension $\geq$ 100 km	2 tracks within 30-dBZ Z_{Hmax} contour $\geq$ 3000 km^2 or 3+ tracks within 30-dBZ Z_{Hmax} contour. Maximum dimension $\geq$ 100 km
QLCS Classification:	Quasi-linear, $Z_{Hmax} \geq$ 40 dBZ region with maximum dimension $>$ 100 km	MCS with convective region aspect ratio $\geq$ 3	—-

Table 3.2: Comparison of techniques used to classify storm mode in past literature.

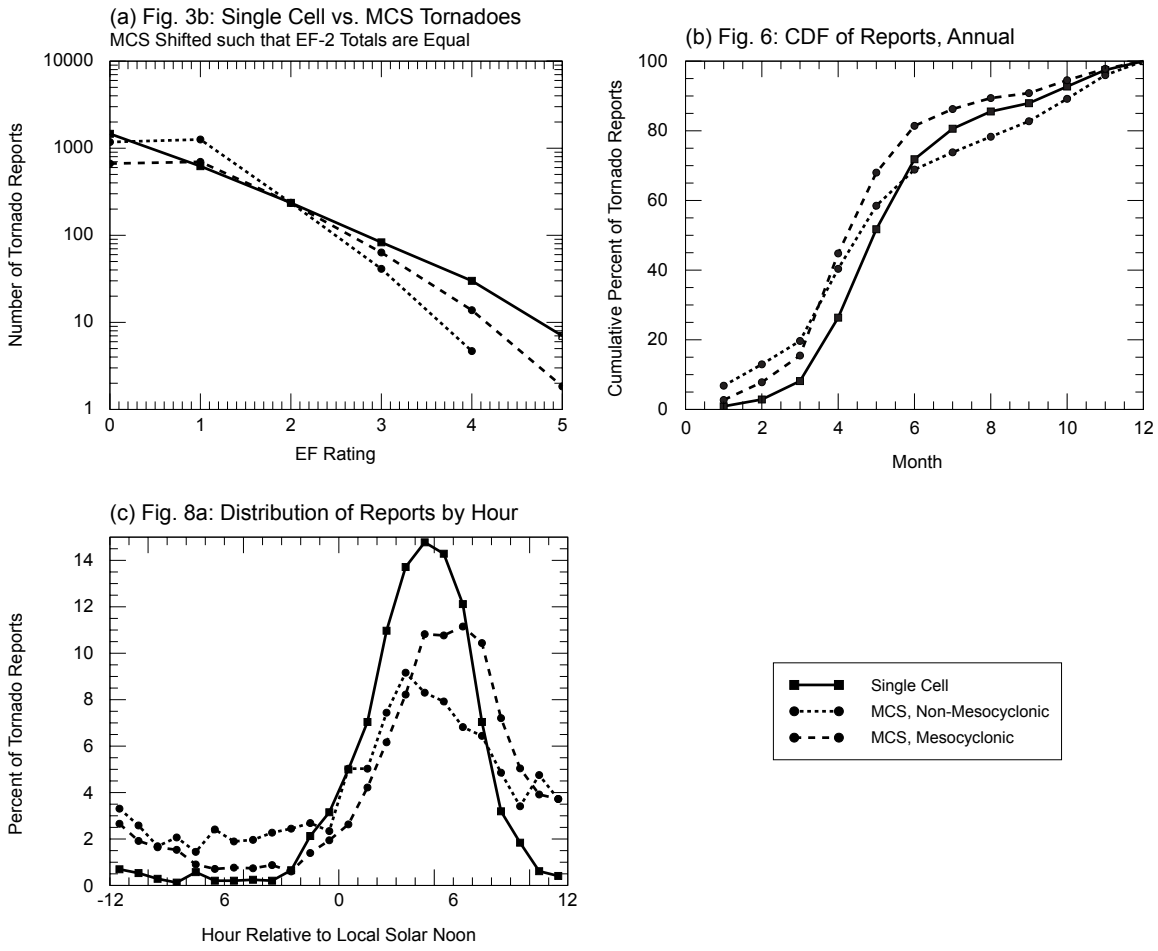

Figure 3.7: Plots reproducing (a) Fig. 3b, (b) Fig. 6, and (c) Fig. 8a in Trapp et al. (2005) using GR-S data. Lines are broken into single cell storms, non-mesocyclonic MCS storms, and mesocyclonic MCS storms. On panel (c), data within each one-hour bin are plotted at the 30 minute mark of that hour.

GR-S data are additionally compared to the findings of Ashley et al. (2019), who used machine learning methods to classify storm mode using a 22-year radar dataset. For training the model, QLCS storms were labeled by hand, and the model was trained on labeled QLCS and non-QLCS events. They defined an MCS per Table 3.2, with that definition motivated primarily by the work of Parker and Johnson (2000). A QLCS is defined as "an MCS that has instantaneous convective ($\geq$40 dBZ) regions that are longer than 100 km and must be at least 3 times as long as they are wide." The major differences between their definition of an MCS and the definition used herein is the 30- vs. 40-dBZ threshold for defining radar echoes for classification, and no temporal threshold vs. a 3 hour temporal threshold for GR-S and Ashley et al. (2019), respectively. Their paper focuses on the spatiotemporal distribution of both QLCSs and QLCS-matched tornado reports, and only the latter will be analyzed herein.

Table 3.3 shows the percent contribution of severe reports attributable to QLCSs in Ashley et al. (2019) juxtaposed with storm-matched GR-S reports attributable to non-mesocyclonic MCS cells. In each category, the percent of reports attributed to MCSs is fairly similar when comparing GR-S data to the results in Ashley et al. (2019). The greatest difference is with attribution of significant severe wind reports (28% of storms in Ashley et al. (2019) vs. $\sim$42% in GR-S). Although trends in these data are similar, differences are no doubt the result of a myriad of differences in methods throughout the data analysis process. Both datasets show a high percentage of wind reports and a low percentage of hail reports attributed to QLCS/non-mesocyclonic MCS storms. Beyond examining total reports attributed to QLCSs, their Fig. 11 shows a breakdown of all severe reports by month and hour, with percent attributed to QLCSs overlaid. Similar to Fig. 3.4 herein, they found tornado and hail reports peak in the late spring and wind reports peak in the early summer. They also found that QLCS contribution to total reports was maximized during the winter months and minimized in late summer/early fall, which was similar to the summer/early

Percent of Severe and Sig. Severe Reports Attributed to MCSs/QLCSs		
Report Type and Severity	GridRad-Severe MCS	Ashley et al. (2019) QLCS
Tornado, Severe:	27.54%	21%
Tornado, Sig. Severe:	18.72%	26%
Hail, Severe:	11.43%	10%
Hail, Sig. Severe:	4.74%	7%
Wind, Severe:	42.31%	28%
Wind, Sig. Severe:	36.61%	34%

Table 3.3: Comparison of percent of reports attributable to MCSs for different report types and severity. Data shown are from Ashley et al. (2019) and GR-S data.

fall minima and wintertime maxima seen in the non-mesocyclonic MCS GR-S data. Examining their hourly data, they found a minimum in QLCS contribution during times of peak reporting ($\sim$18-03 UTC), which, when examining GR-S data binned by local time (in UTC, not shown), non-mesocyclonic MCS contributions are minimized from 20-04 UTC for tornadoes and from 17-03 UTC for hail and wind reports. The average percent contribution of QLCSs to hail reports in their study was lower than that for wind and tornado reports, which is also reflected in the GR-S data. Overall, the similarities between GR-S conclusions and those of Trapp et al. (2005) and Ashley et al. (2019) point to the efficacy of the GR-S techniques as a whole and the storm mode classification algorithms used herein.

Chapter 4

PMM analysis of tornadic vs. nontornadic non-supercellular MCS storms[1]

The creation of this large GR-S dataset allows for the opportunity to investigate very specific types of storms that ordinarily would have a small sample size, but given the 1.3 million storms in GR-S, even the most specific type of storm still has a fairly large sample size. Given the relative difficulty of nowcasting the tornadic potential of a non-supercellular MCS storm, GR-S radar data for tornadic and nontornadic non-supercellular MCS storms is interrogated to determine any potential differences between the two groups' appearances on radar. This can potentially aid forecasters when trying to make a warning decision and isolate the variables that show the greatest differences between the two groups for use in additional research applications.

4.1 Selecting tornadic and null populations

Tracked storms identified as non-supercellular MCS cells for $\geq 75\%$ of their lifetime are isolated for analysis. This 75% threshold aims to focus attention on cells that are predominantly classified as MCS cells while also allowing for some variation in mode classification during their lifecycles. These non-supercell MCS storm cells are further split into tornadic and nontornadic populations based on whether they were linked with a tornado report at any point in their lifetime. Sensitivity tests to the MCS lifetime threshold spanning 50-100% were carried out and resulted in fairly similar bulk population characteristics. Specifically, regardless of the threshold chosen, the degree of separation between tornadic and nontornadic distributions of bulk characteristics such as azimuthal shear, radial divergence, and

[1]The work within this chapter has been published in the peer-reviewed literature as Murphy and Homeyer (2023).

spectrum width were fairly similar. In particular, Kolmogorov-Smirnov (K-S) significance test results–a metric of the difference between two distributions–differ by a median of $\sim$8% when examining distributions of lifetime maximum, median, and minimum values of radar metrics for different MCS lifetime thresholds. This means that the separation between these distributions were fairly similar regardless of whether the minima, medians, or maxima were evaluated, and regardless of how long a cell needed to be classified as an MCS to be included. Ultimately, a less restrictive threshold was ultimately preferred to maximize sample size.

Identifying differences between tornadic and nontornadic storms, especially in the time before first tornadogenesis, is incredibly important for warning applications. Comparing these two populations without any further processing would be informative; however, the main challenge for improving warning metrics for non-supercellular MCS storms comes from identifying nontornadic storms that otherwise look promising for tornadogenesis, and comparing their characteristics to those of tornadic storms. Using radar data, we can further identify unique storm characteristics that may separate these populations. Radar characteristics used to isolate nontornadic storms that have similar characteristics to tornadic storms–i.e., the null population–include azimuthal shear, radial divergence, and spectrum width over various layers, as well as vertically integrated liquid density (VIL density, a metric for severe hail potential; Greene and Clark 1972; Amburn and Wolf 1997) and low-level hail differential reflectivity (H_{DR}, a polarimetric metric for severe hail potential; Aydin et al. 1986). Layers include low-level (LL; $0 \leq$ z $<$ 4 km AGL), midlevel (ML; $4 \leq$ z ≤ 7 km AGL), and column maximum values. For each storm, the median value of each variable across the entire storm's lifetime is retained for comparison. Figure 4.1 shows the distributions of these variables for tornadic and nontornadic storms. Median data are used instead of storm lifetime maximum or minimum values because results of K-S significance tests to assess differences in the distributions were typically higher for storm lifetime median

values. It is possible that this result arises due to mitigation of the influence of spurious features and tornadic circulations by using a median filter.

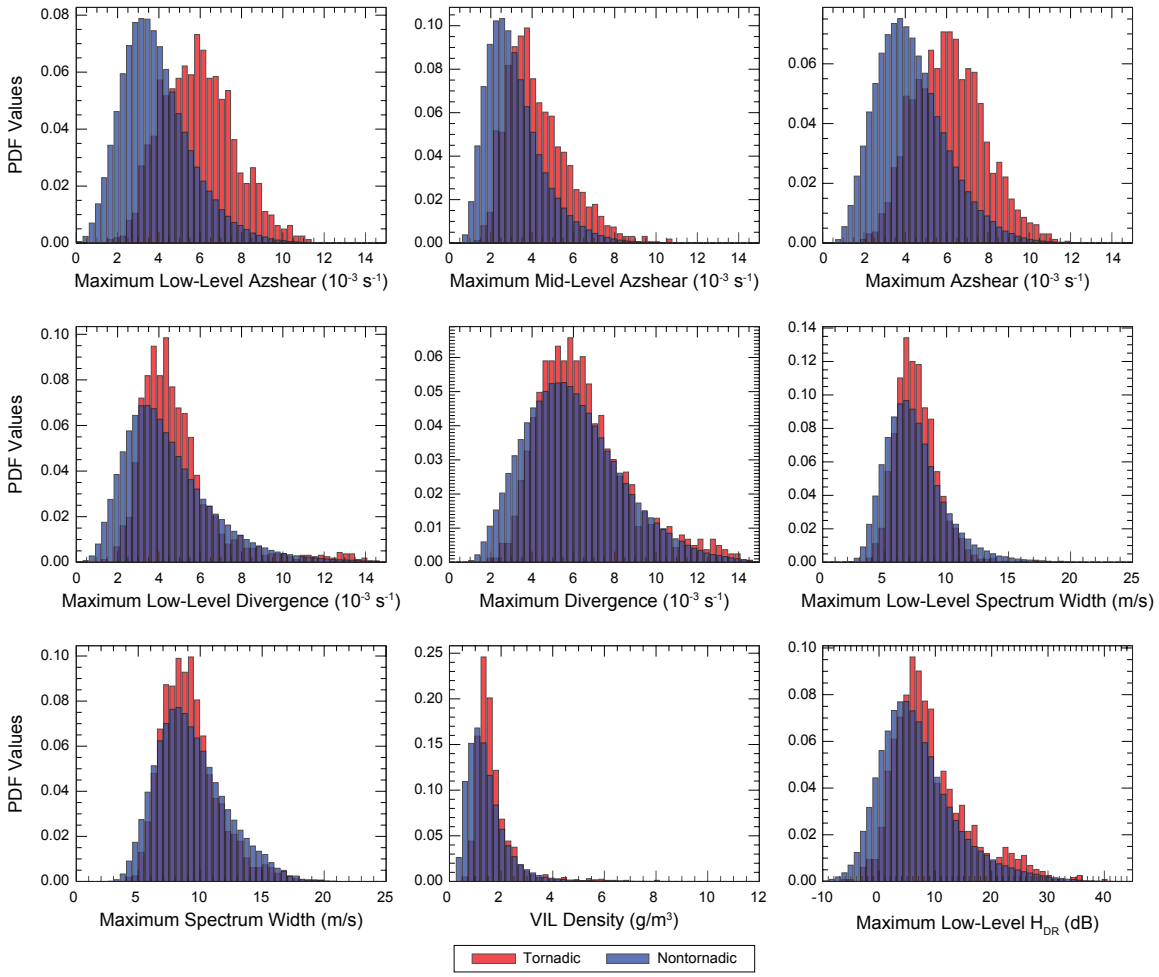

Figure 4.1: Distributions of kinematic variables for non-supercellular tornadic and nontornadic MCS storms within the GR-S dataset from 2010-2019, inclusive. Data are the median values of the listed radar variable across each storm's lifetime.

The degree of separation between the nontornadic and tornadic non-supercellular MCS cell populations for each variable is quantified to determine which best delineate between storm type. Null storms can then be selected from the smaller population of nontornadic storms that overlap the distribution of tornadic cell values for variables with a high degree of separation, since such storms would be most likely to be (incorrectly) warned as a potential tornadic storm. To quantify the significance of separation between the tornadic and nontornadic distributions, K-S and two-sample T-tests were performed. T-tests show that

all variables showed a significant difference ($p < 0.05$) between the tornadic and nontornadic distributions. K-S tests, with values calculated from normalized CDFs, showed that maximum ML azimuthal shear (0.07), maximum LL azimuthal shear (0.07), maximum column-max azimuthal shear (0.06), and maximum LL divergence (0.04) had the largest K-S values of all variables and therefore showed the largest separation between tornadic and nontornadic storms. Maximum LL H_{DR} (0.07) also had one of the largest K-S test values, but given that it could not be calculated for storms from 2010-2012, it was not considered as a criterion for null case selection.

Out of the identified radar-derived variables with high K-S test values, null storms are selected using LL azimuthal shear values, since high LL azimuthal shear values are often interpreted as anecdotal radar evidence of a mesovortex capable of producing a tornado. Using the histogram of tornadic LL azimuthal shear values in Fig. 4.1, null storms are randomly sampled from the nontornadic distribution at the observed frequency of tornadic storms. The resulting null sample exhibits an equivalent LL azimuthal shear distribution to the tornadic sample and enables confident comparison of storms that differ nearly solely by their tornadic production. Figure 4.2 shows the distribution of tornadic and null storms for the nine variables shown in Fig. 4.1. Two of the 1627 tornadic storms did not have LL azimuthal shear data, so the null sample is 1625 storms. Sensitivity testing demonstrates that null distributions in Fig. 4.2 do not change meaningfully with varying random samples (not shown).

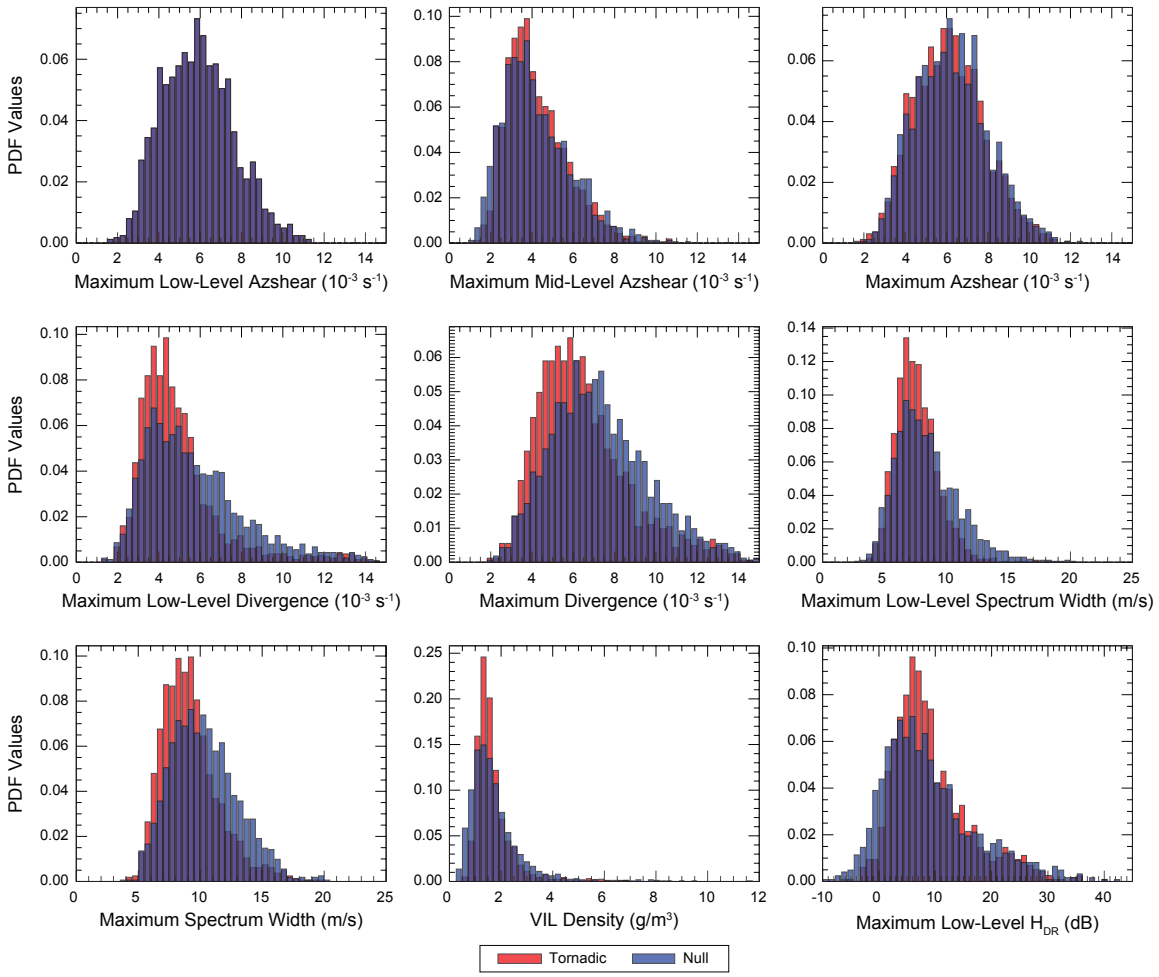

Figure 4.2: As in Fig. 4.1, but comparing tornadic storms to the sampled null storms.

4.2 Probability-matched mean technique

After identifying tornadic and null storms, PMMs of radar data for each population are created. Similar to the work in Homeyer et al. (2020), PMMs show the average spatial distribution of a given radar variable with values scaled to avoid non-physical smoothing due to the averaging process. These data differ in that the focus herein is on non-supercellular MCS storms, compared to Homeyer et al. (2020)'s analyses of supercell storms. Examining storm-average appearances in this way highlights the signatures within tornadic or non-tornadic storms that show up most consistently ("repeatable and substantial differences";

Homeyer et al. 2020), emphasizing the most robust distinctive features within tornadic and nontornadic storms and screening out less consistent storm-to-storm variabilities. PMMs are created at 0.5, 1.5, 3, 5, and 10 km AGL for Z_H, azimuthal shear, radial divergence, and σ_V for all cases, and Z_{DR}, K_{DP}, and ρ_{HV} for storms with polarimetric data (2013 and later). Tornadic storms are analyzed at 0 and 20 minutes prior to first tornadogenesis, and null storms are analyzed at times of peak 30-dBZ echo top height and peak LL rotation. These two times within the null storms' lifetimes are selected to focus on the time of peak storm intensity (and presumed strongest updraft) and strongest low-level circulations, respectively. Both times aim to isolate the most likely time for tornadogenesis via stretching of low-level vorticity (Markowski and Richardson 2009). Similar to Homeyer et al. (2020), extreme values from the individual storm observations are trimmed before computing the PMMs. Extrema in this study are considered to be values less than the 0.1st and greater than the 99.9th percentile of all observations contributing to a PMM (as in Homeyer et al. 2023). For single-polarization analyses, 1625 null and 1627 tornadic storms are analyzed, respectively; for dual polarization analyses, these numbers are reduced to 1231 and 1166 storms (an approximate reduction of $\sim$25%). A total of 840 and 621 of the 1627 tornadic storms could be tracked at 20 minute lead times for single- and dual-polarization analyses, respectively. By creating and comparing PMMs of tornadic and null storms, further differences may be revealed and highlighted between these two populations.

4.3 Results

4.3.1 Single-polarization

Single-polarization radar variables lend insight into the difference between null and tornadic storms at and before their peak intensity. Fig 4.3 shows storm-centered, constant-altitude PMMs of Z_H for null and tornadic storms. Individual storm observations are centered on each storm's maximum 30-dBZ echo top location and rotated such that storm motion points to the right before PMMs are calculated. When discussing features of any

individual PMM plot, signatures and their locations are all described relative to storm center and storm motion, to ensure clarity. The primary differences between the null and tornadic data in Z_H are with the mean storm shape–null storms have more of a diffuse, circular shape when compared to tornadic storms (in both times analyzed), which have higher reflectivity more tightly aligned on a lower-left to upper-right axis (relative to storm motion). This may indicate a more consistent MCS line orientation relative to storm motion for tornadic storms compared to null storms. Panels of tornadic storm data at tornadogenesis also show contoured frequency of tornado reports, which are offset slightly right of storm motion and up-motion relative to storm center (i.e., the echo top maximum).

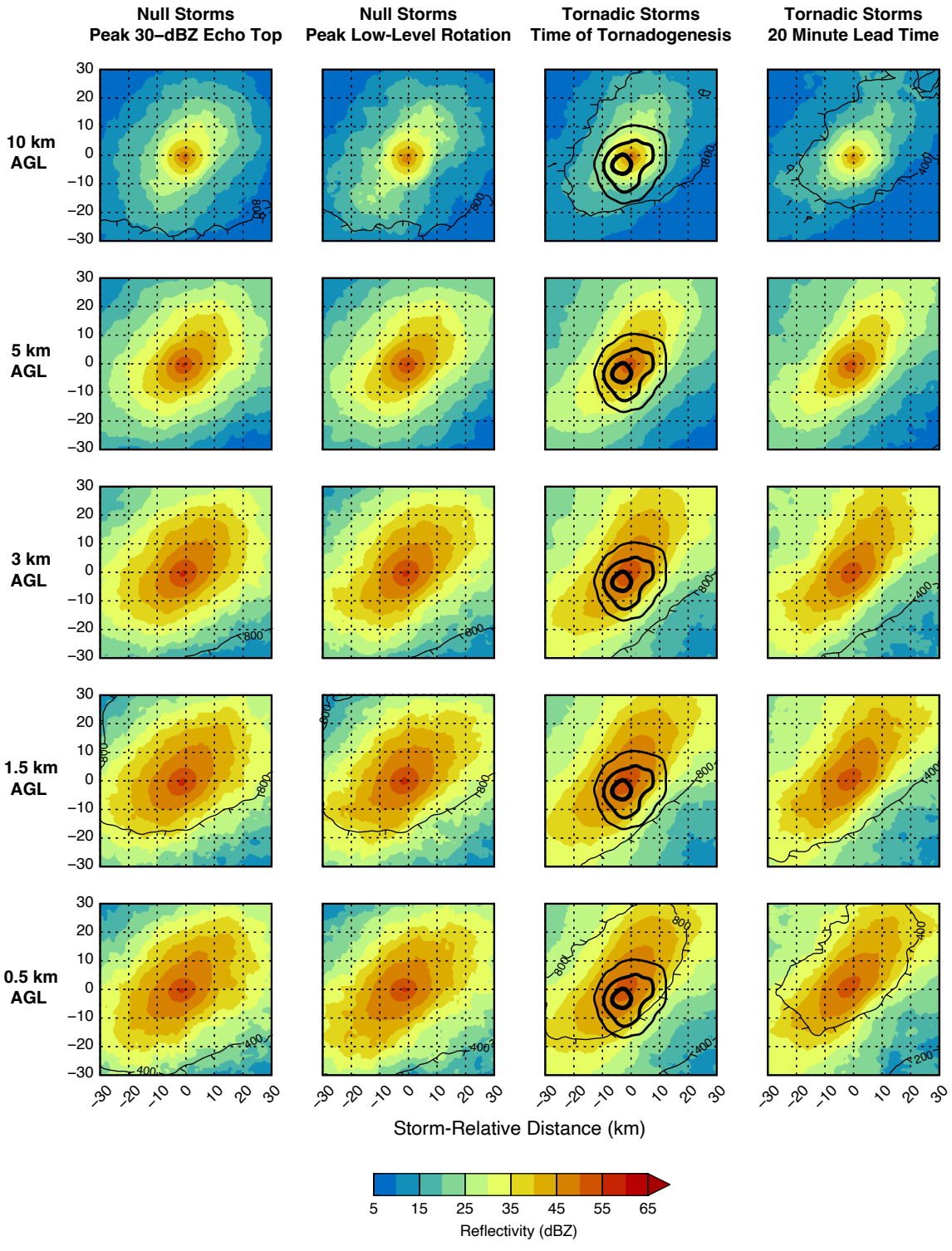
Null Storms
Peak 30–dBZ Echo Top
Null Storms
Peak Low-Level Rotation
Tornadic Storms
Time of Tornadogenesis
Tornadic Storms
20 Minute Lead Time
10 km AGL
5 km AGL
3 km AGL
1.5 km AGL
0.5 km AGL
Storm-Relative Distance (km)
5 15 25 35 45 55 65
Reflectivity (dBZ)

Figure 4.3: PMM constant altitude plots of Z_H for null and tornadic storms. Null data are shown for times of peak 30-dBZ echo top height and peak low-level rotation, and tornadic data are shown at tornadogenesis and 20-minute lead time. Data are shown at 0.5, 1.5, 3, 5, and 10 km AGL. Thin labeled contours represent the number of storms included in each mean, with dashes pointing towards lower values. Thicker contours represent total tornado reports (only in PMMs at time of tornadogenesis), with increasing contour thickness corresponding to higher frequency of reports. Contours enclose locations whose total report count is at least 30, 60, or 90% of the maximum density of reports (after Gaussian smoothing). Storm motion points to the right of each plot.

Investigations of azimuthal shear, radial divergence, and σ_V PMMs show how the location of common storm-relative features may differentiate between null and tornadic storms. Figure 4.4 shows storm-centered, constant-altitude PMMs for azimuthal shear. A key feature present in all PMMs is a mesovortex (quasi-elliptical, highly positive values) right of storm motion and displaced up to 10 km from storm center. For tornadic storms, the location of the low-level mesovortex is in close proximity with the most frequent location of tornadogenesis and overall collocated with the outermost tornadogenesis location contour. Distinct differences between azimuthal shear maxima and storm center are visible between the null and tornadic data. Throughout the majority of the storms' depths for tornadic storms at first tornadogenesis and 20-minutes prior, azimuthal shear maxima occur in a broadly consistent storm-relative location. Differences between tornadic and nontornadic cells are most evident at low-levels, where low-level azimuthal shear maxima are less concentrated/coherent and displaced further right of storm center and down-motion in nontornadic cells (i.e., less vertically aligned with upper-level mesovortex locations at 5-10 km). From 3-10 km, both types of storms show a dipole of positive/negative azimuthal shear, rotating roughly 90 degrees clockwise in orientation through the column. Corresponding radial divergence PMMs in Fig. 4.5 show convergence (negative divergence) roughly

collocated with tornadogenesis locations at 0.5 and 1.5 km, and convergence at the same altitudes within the null storm PMMs is displaced right and up-motion relative to storm center. In both cases, low-level convergence signatures are roughly collocated with positive low-level azimuthal shear (Fig. 4.4). Looking at maximum convergence at 3 and 5 km and maximum divergence at 10 km, these signatures–indicative of the storm updraft's location–are close to storm center for both tornadic and null storms, as expected. This means that low-level vertical motion is better coupled with the mid- to upper-level storm updraft in tornadic cells than it is for the null cells. Finally, Fig. 4.6 shows a near collocation of enhanced σ_V values with storm center throughout the depth of tornadic storms, whereas such maxima are displaced from storm center (and mid- to upper-level maxima) and less coherent at 0.5 km AGL for null storms. Magnitudes of low-level σ_V, however, are notably higher in nontornadic storms. Enhanced σ_V can be driven by several factors, including wind shear, rotation, and measurement error (e.g., Doviak and Zrnić 1993). Assessing the PMM fields in tandem, tornadic storms are marked by a collocation of low-level vertical vorticity, convergence, and enhanced σ_V with the mid- to upper-level updraft location. Conversely, these low-level features in null storms are more diffuse, less coupled with each other, and displaced from the mid- to upper-level updraft, impeding any stretching of the low-level rotation in the vertical and potential resultant tornadogenesis. It is important to note that the enhanced low-level azimuthal shear signature in tornadic storms stands out despite controlling for low-level azimuthal shear differences between tornadic and null storms via the null storm selection criteria, highlighting the importance of taking into account both shear magnitude and location in assessing potentially tornadic non-supercellular MCS storms.

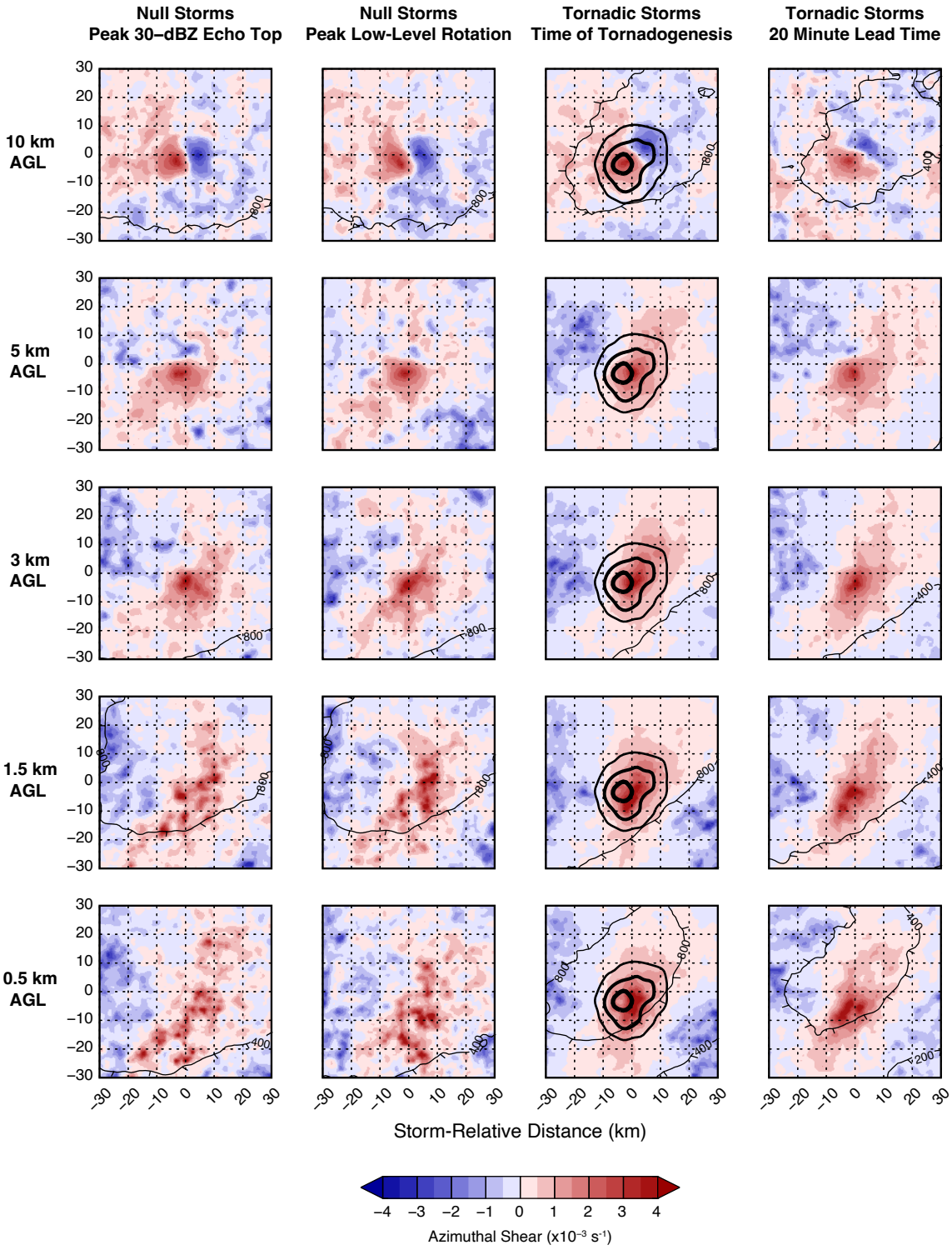

Figure 4.4: As in Fig. 4.3, but for azimuthal shear.

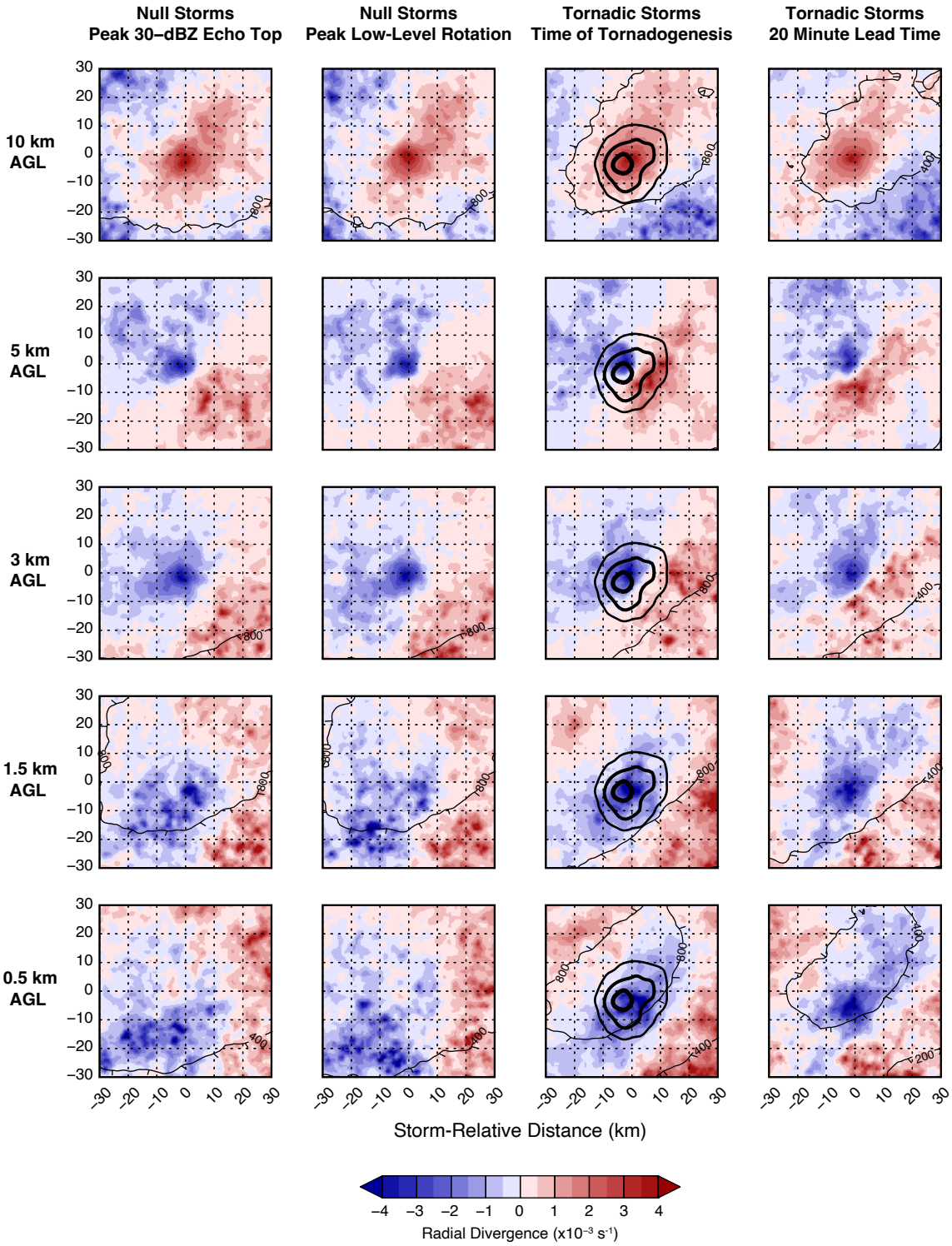

Figure 4.5: As in Fig. 4.3, but for radial divergence.

Vertical PMM cross sections through tornadic and null storms offer additional insight into the storms' mean vertical structure beyond what can be inferred from constant-altitude PMMs. Figures 4.7 and 4.8 show cross sections through null and tornadic storms at the same times as Figs. 4.3-4.6, cutting through the storms both parallel to (Fig. 4.7) and perpendicular to (Fig. 4.8) storm motion. These cross sections are through the approximate frequency maximum in tornado reports (see Figs. 4.3-4.6) at (-3, -3) km relative to storm center. It is important to note that neither of these cross sections bisect storm center (and the approximate location of the mid- to upper-level updraft), so examinations here are focused on tornadogenesis-centered storm structure (or the equivalent storm-relative location, in the case of null storms) rather than storm-centered structure as was the case in Figs. 4.3-4.6.

A few storm characteristics are more evident from these vertical sections. Focusing on the parallel cross-section in Fig. 4.7, a vertically aligned column of positive azimuthal shear indicating the mesovortex location is nearly coincident with the most frequent tornadogenesis location in tornadic storms, whereas in null storms, the low-level mesovortex is displaced down-motion of the preferred tornadogenesis location and is no longer vertically aligned with the upper-level mesovortex location (which presents similar to that in tornadic cells). Divergence plots also show higher convergence at the lowest observed levels for tornadic cells, with low-level convergence in null cells approaching zero. From the cross sections perpendicular to storm motion (Fig. 4.8), both tornadic and null cases show enhanced azimuthal shear at low levels over the preferred tornadogenesis location, although the maxima for tornadic storms are more broadly distributed about the tornadogenesis location. Similar to Fig. 4.7, divergence plots show stronger convergence at lowest levels for tornadic data than for null data, and σ_V behavior mirrors that of Fig. 4.7. Overall, it appears that the most critical characteristics of a tornadic cell is the vertical alignment of the mesovortex coupled with coincident enhanced low-level convergence (i.e., a strong low-level updraft coupled with the mid- and upper-level storm updraft). This association of

enhanced rotation on radar with potential QLCS tornadogenesis has been noted before, including as a "confidence builder" in the Three Ingredients Method for identifying potential QLCS mesovortexgenesis and tornadogenesis (e.g., Gibbs 2021).

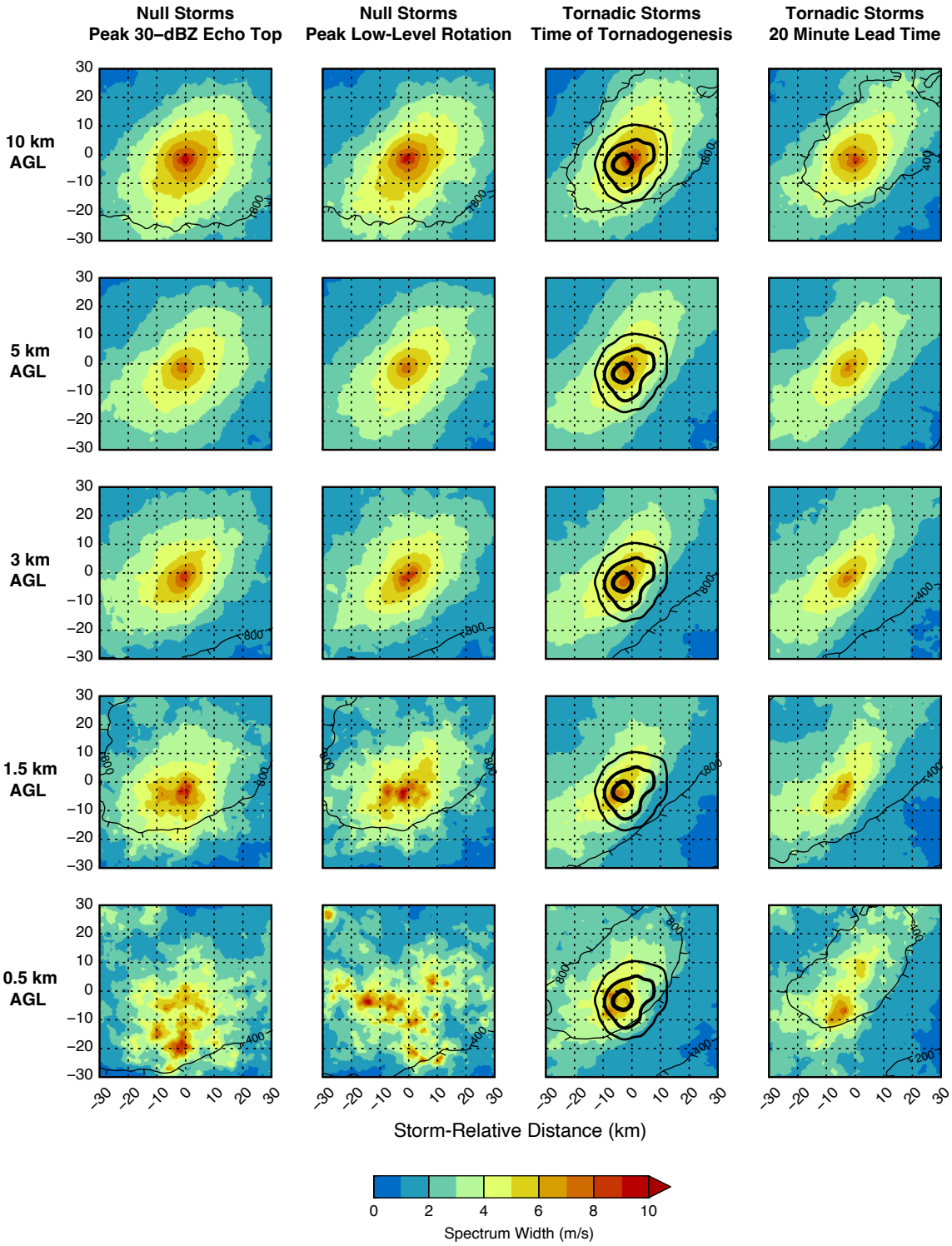

Figure 4.6: As in Fig. 4.3, but for σ_V.

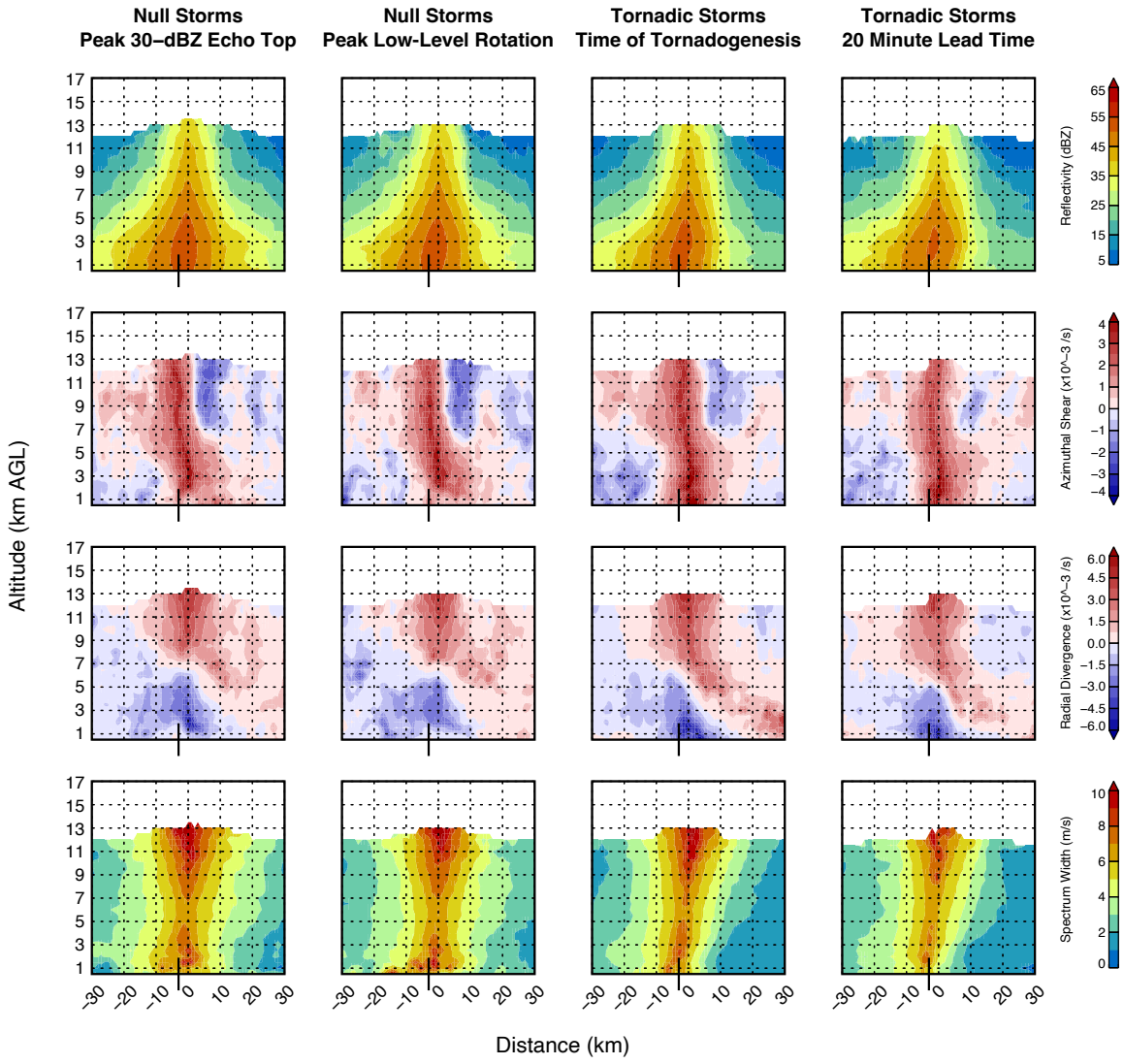

Figure 4.7: Cross sections parallel to storm motion of PMM (rows) Z_H, azimuthal shear, radial divergence, and σ_V for (columns) null and tornadic storms. Null data are shown for times of peak 30-dBZ echo top height and peak low-level rotation, and tornadic data are shown at tornadogenesis and 20-minute lead time. Cross sections bisect the approximate maximum of tornado reports (see Fig. 4.4) with a center at (0, -3) km relative to storm updraft. Bold tick marks show the approximate location of the maximum in tornado report frequency.

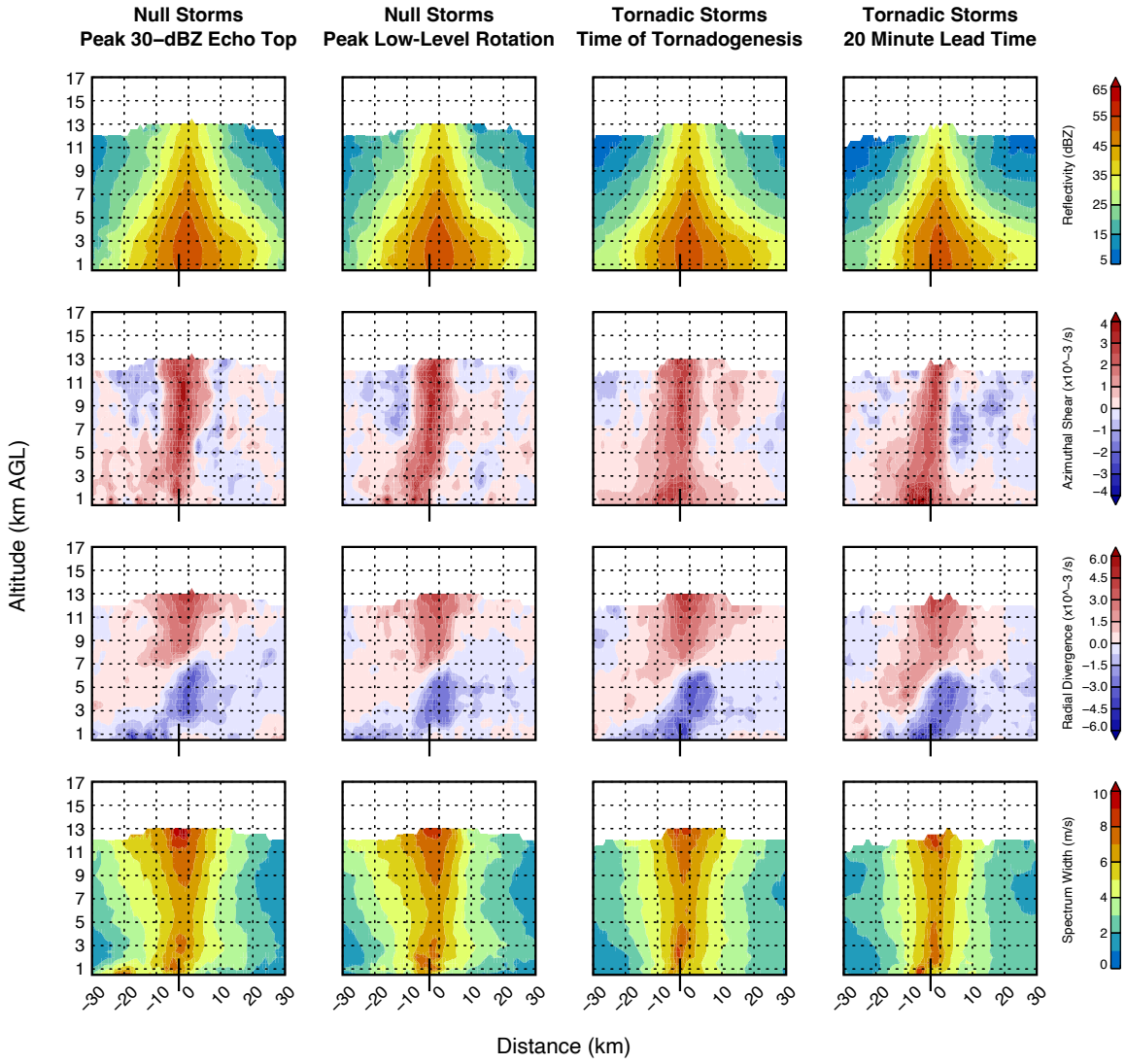

Figure 4.8: As in Fig. 4.7, but taken perpendicular to storm motion. Cross sections bisect the approximate maximum of tornado reports (see Fig. 4.4) with a center at (-3, 0) km relative to storm updraft.

4.3.2 Dual-polarization

After the nationwide upgrade of all WSR-88D radars to polarimetric capabilities by 2013, PMMs of polarimetric radar data can also be examined to identify potential distinguishing microphysical characteristics. Figures 4.9, 4.10, and 4.11 show constant-altitude PMMs

for Z_{DR}, K_{DP}, and ρ_{HV}, respectively. Examining these figures together, there are two main differences that are evident between null and tornadic storms. The first difference is the orientation of storm motion relative to the broader MCS convective line orientation (as seen in PMMs of Z_H in Fig. 4.3). Namely, Figs. 4.9 and 4.10 show a preferential line orientation at $\sim$45° left of storm motion at low levels for tornadic storms, and a more diffuse orientation for null storms. Second, all figures show slight differences between magnitudes of low-level Z_{DR} and K_{DP} maxima. For Z_{DR}, the maxima for tornadic and null storms (as shown by plotted contours) are the same, although contours of null storm maxima cover a broader area. For K_{DP}, maxima are also the same from 0.5-3 km across all storm types and times, but this time the maximum contours at first tornadogenesis cover a larger area (although less pronounced than the relative sizes of the Z_{DR} maximum contours). This may suggest a slight tendency towards a lower concentration of larger drops in null storms and a higher concentration of more moderately-sized drops near time of tornadogenesis; however, Z_{DR} and K_{DP} values for null and tornadic storms are overall quite comparable. For ρ_{HV}, minimum values are comparable between PMMs and show no consistent differences between storm type.

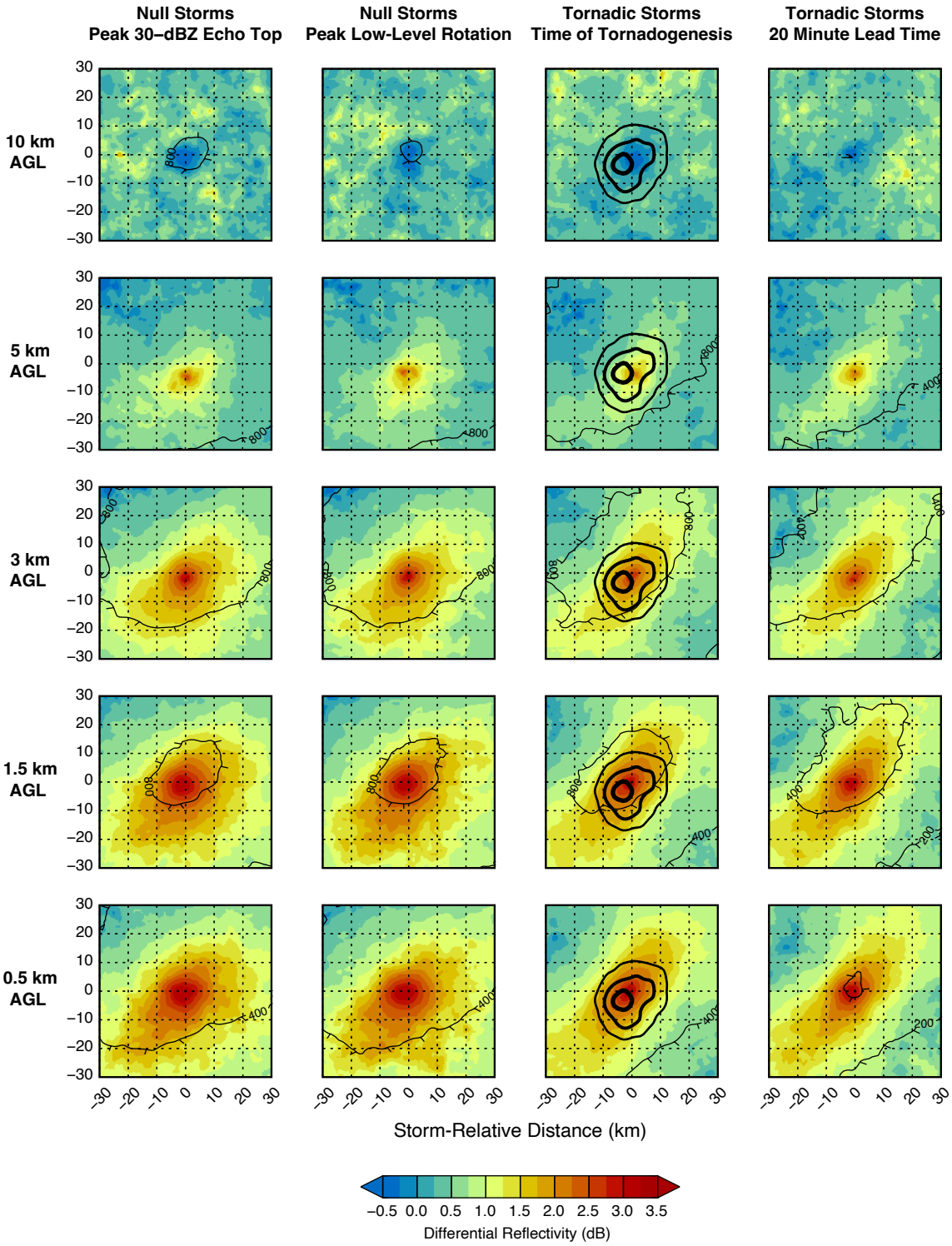

Figure 4.9: As in Fig. 4.3, but for Z_{DR}.

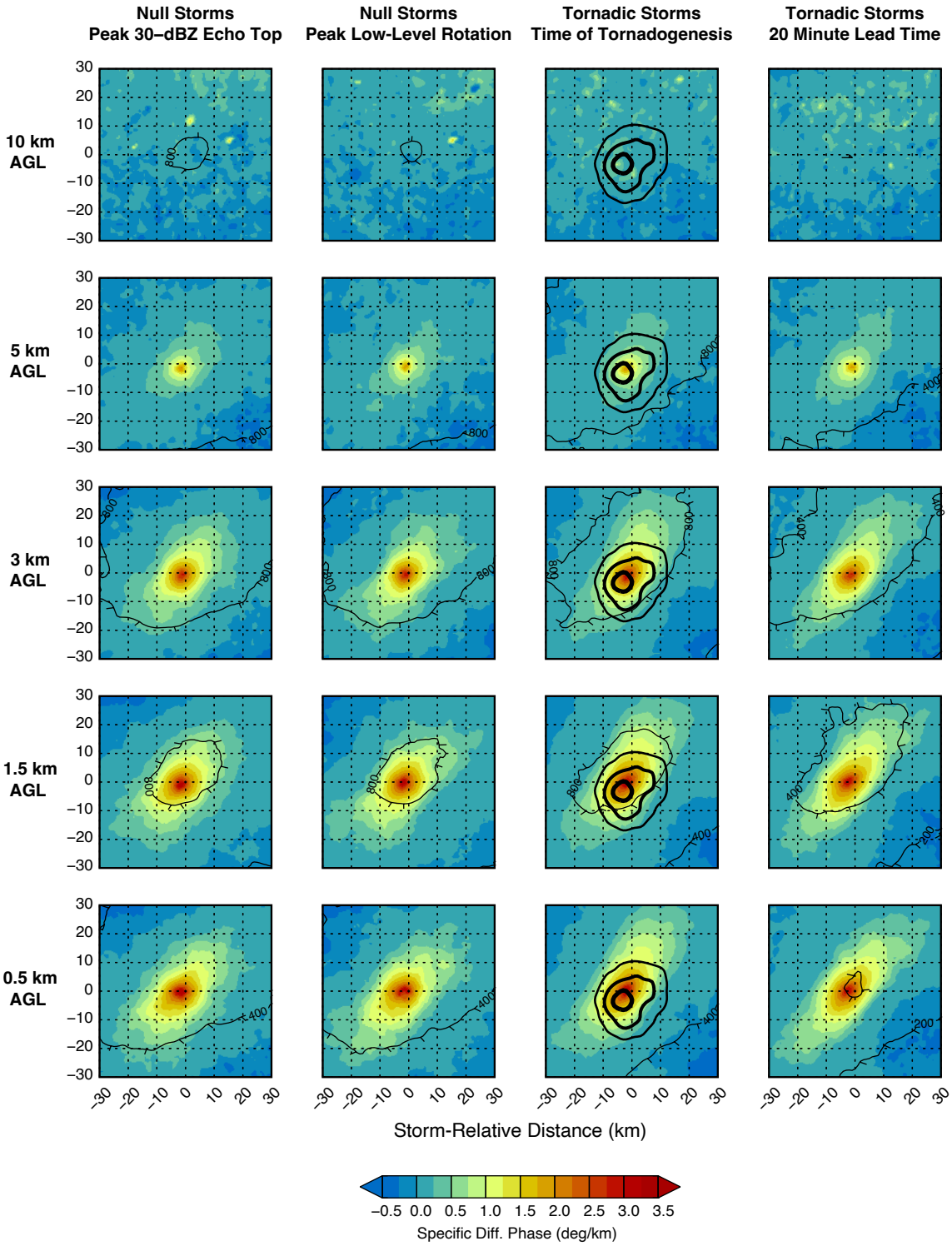

Figure 4.10: As in Fig. 4.3, but for K_{DP}.

Figures 4.12 and 4.13 show vertical cross sections of the polarimetric variables' PMMs through tornadic and null storms parallel and perpendicular to storm motion, respectively. Similar to Figs. 4.7 and 4.8, cross sections are through the preferred tornadogenesis location. Consistent with the constant-altitude plots, there are slight differences, but tornadic and null values are again mostly indistinguishable. Microphysically, higher Z_{DR} in null storms suggests larger drops near the surface compared to smaller drops in tornadic storms. The melting layer (as deduced from local minima in ρ_{HV} away from the composite storms' convective cores) is comparable at 3-4 km AGL in both storm populations, so any differences owing to environmental variability are likely to have limited impact on interpretation herein. Qualitatively, it appears that the more striking differences between the null and tornadic storms are seen in the single-polarization analyses, and polarimetric data provide limited additional discriminating information between non-supercellular tornadic and non-tornadic storms within MCSs.

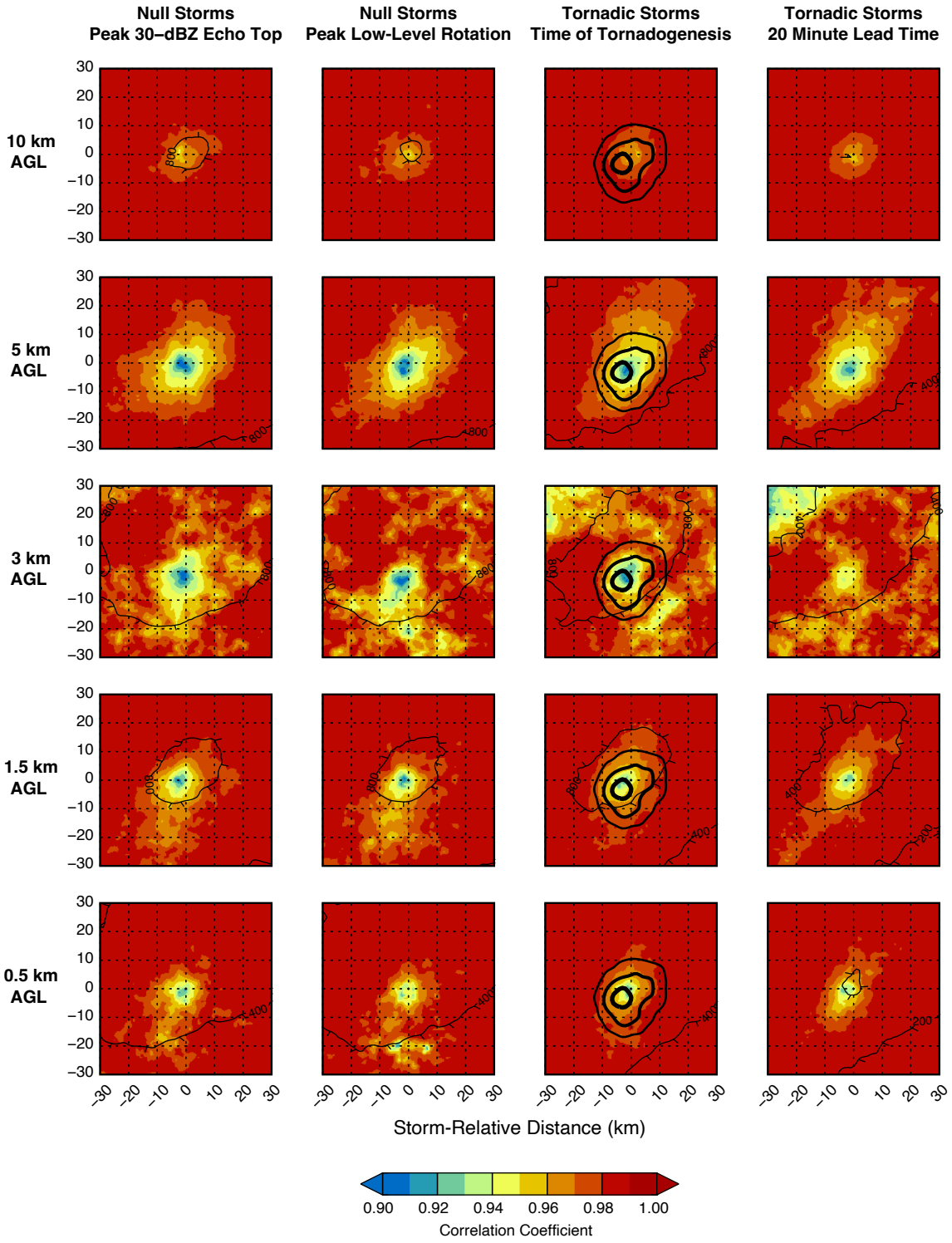

Figure 4.11: As in Fig. 4.3, but for ρ_{HV}.

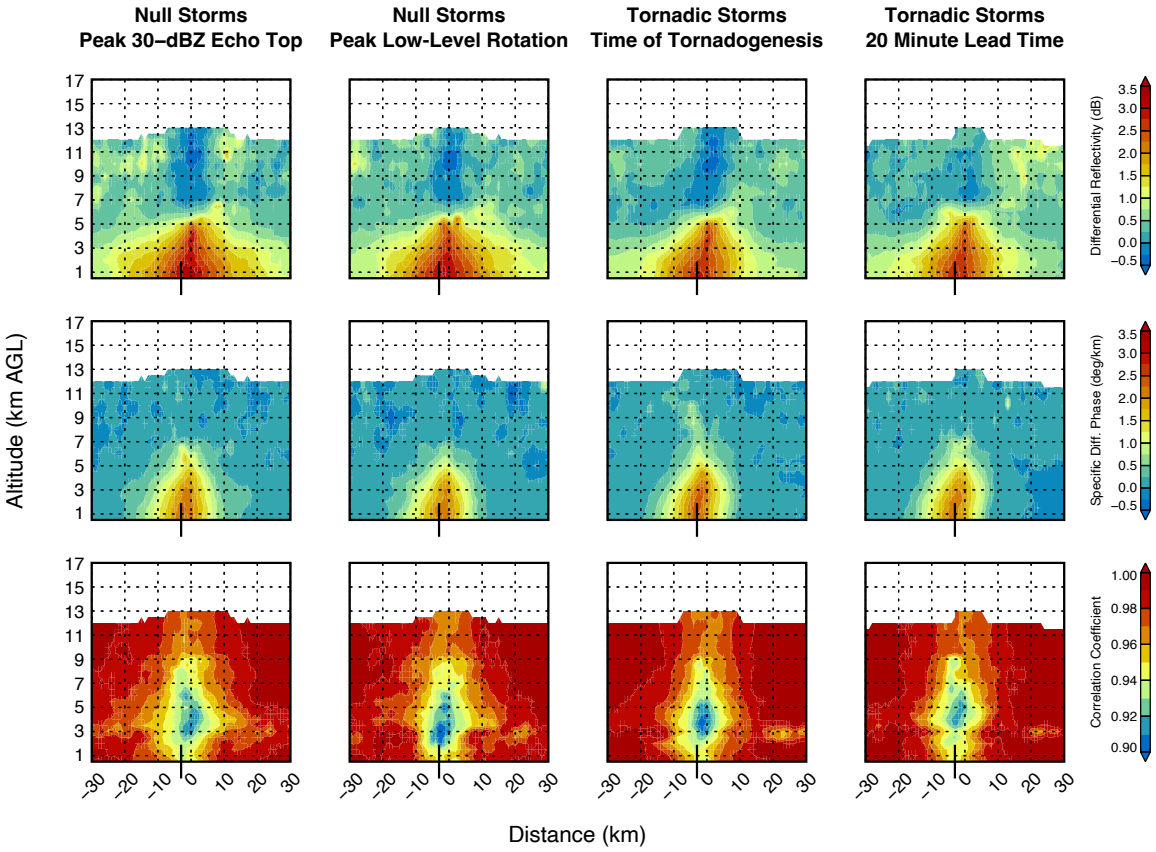

Figure 4.12: As in Fig. 4.7 but examining (rows) Z_{DR}, K_{DP}, and ρ_{HV}.

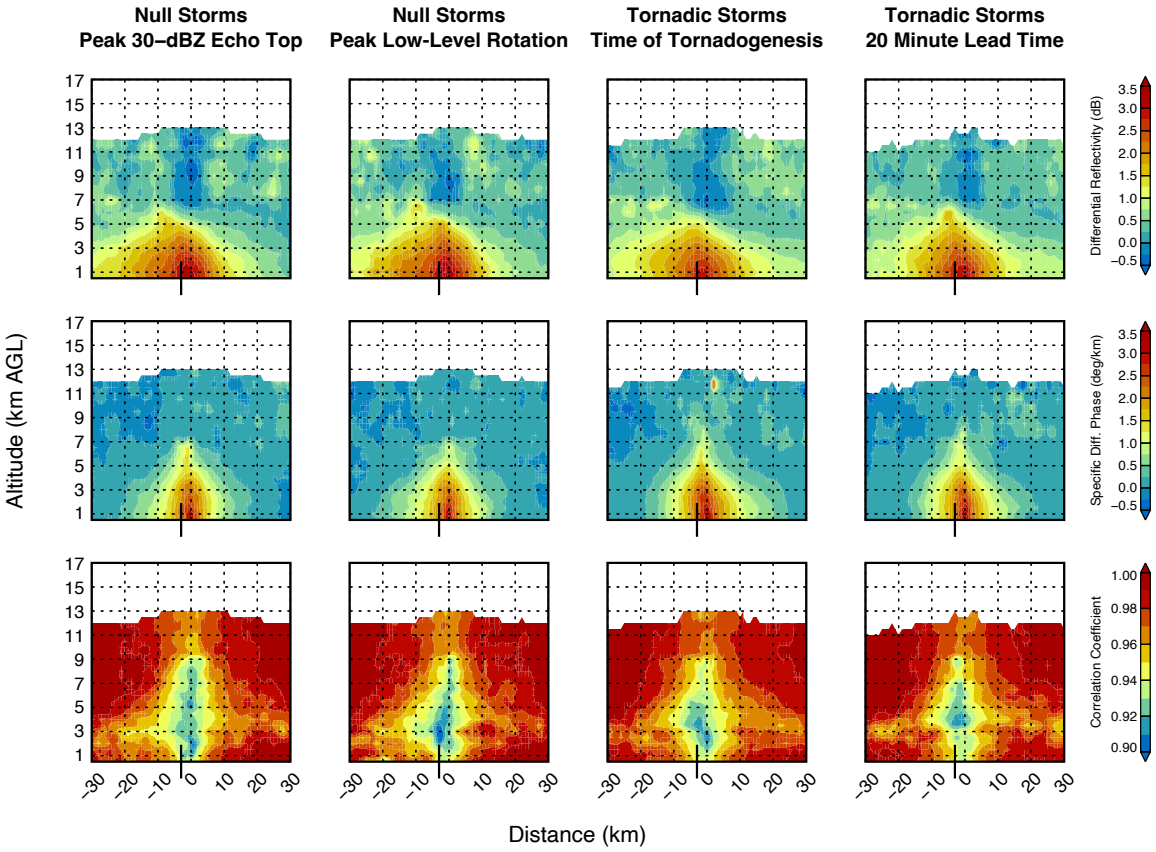

Figure 4.13: As in Fig. 4.8 but examining (rows) Z_{DR}, K_{DP}, and ρ_{HV}.

Chapter 5

Machine learning methods for improved non-supercellular MCS tornado prediction

5.1 Data and Methodology

5.1.1 Model architecture

Now that signatures differentiating tornadic and nontornadic MCS non-supercellular storms have been identified, these data can be used as inputs to a machine learning model to determine whether or not machine learning techniques can determine if a storm will become tornadic, and with what accuracy. To do tornado prediction using radar data, a convolutional neural network (CNN) is used. It is important to note that, with machine learning, it is best to use the simplest possible model in order to increase interpretability of results (Chase et al. 2023). In the case of radar data, not only are the values of the data points important, but the spatial distribution of those values are often also very informative (e.g., the preferential axis of Z_{H} for tornadic storms in Section 4.3.1). For that reason, Chase et al. (2023) suggest in their guide on machine learning in meteorology to default to a CNN architecture for spatial datasets when "you are unsure of what features to extract." Given the lack of an archetypical structure/typical tornadogenesis location for MCS non-supercellular storms (compared to a discrete supercell, where the main feature of interest would be the hook echo), a CNN is an appropriate choice for using ML to identify MCS non-supercellular tornadoes. The CNN used herein is based on open-source code provided by Chase et al. (2023) to facilitate use of machine learning techniques in meteorological applications.

Training and validation data are from 2010-2019 data in the GR-S dataset, and are split into 4035 storms for training and 999 storms for validation. These storms were randomly selected, meaning that storms in close spatiotemporal proximity during the same event may have been split between the training and validation datasets. This means that storms that may have dynamically influenced each other could be present in both training and validation datasets, and the datasets themselves may be somewhat dependent on each other. However, the focus herein remains on the radar presentation of these storms and does not include environmental data. At the very least, this is a potential limitation of the work that must be considered. Data augmentation techniques as in Lagerquist et al. (2020) were applied to the training dataset, given that they attributed a great deal of their own study's success to data augmentation. Data augmentation (including rotations, shifts, and additions of noise) increased the training dataset by a factor of 18 and resulted in 72630 total training examples. Testing data were sourced from 2020-2021 GR-S data, to remain independent of the training and validation datasets. Given the results of Section 4.3, only single-polarization variables (Z_{H}, azimuthal shear, radial divergence, and σ_V) were fed into the model. For each of these variables, the specific ingest data are the storm-centered volumes of individual tornadic and non-tornadic storms averaged over low- and mid-levels (0.5-2.5 km and 4.0-6.0 km AGL, respectively) such that 8 different fields (4 variables at 2 layer-average levels) are ingested. Knowing that the major differentiating factors between tornadic and nontornadic storms in Chapter 4 were close to the tracked storm center, the inner 15 x 15 km grid for each storm is ingested rather than the full 30 x 30 km grid. Within that inner grid, those data currently at 0.5 km horizontal resolution are interpolated to a 32 x 32 grid to facilitate pooling layers after convolution (which reduce the horizontal dimensionality for 2D convolutions and horizontal and vertical dimensionality for 3D convolutions by a factor of 2). Training data are then normalized as in Lagerquist et al.

(2020) and data augmentation applied. The final training and validation datasets containing tornadic and nontornadic examples are then randomly shuffled before training begins, to avoid any exclusively tornadic or nontornadic batches during training and/or validation.

CNNs contain a number of human-tuned hyperparameter choices, such as number of convolutional layers, activation function, and dropout rate. As the number of hyperparameters grows, with multiple options for each hyperparameter category, finding the optimal combination of hyperparameters to improve model performance becomes incredibly time-intensive. A hyperparameter selection code developed by Chase et al. (2023) is used herein to select ideal hyperparameters, with hyperparameter options outlined in their Fig. A2. The hyperparameters tested, as well as those chosen, are shown in Table 5.1.

5.1.2 Quantifying model success

When evaluating model performance, a number of metrics are commonly used in meteorology. Roebber (2009) describe these in terms of a 2x2 square, where each of the 4 squares are defined by whether or not a storm was tornadic and whether or not a storm was warned (or, in this case, identified as tornadic by the ML model). A measure of the model's ability to identify tornadic storms, POD is the ratio of tornadic storms identified as tornadic to the total number of tornadic storms. Conversely, false alarm ratio (FAR) is the ratio of nontornadic storms that were warned as tornadic to all storms warned as tornadic, and is a metric of how often a model falsely predicts that a storm is tornadic. In practice, a forecaster strives for a high POD and low FAR to ensure that the public is warned for as many tornadic storms as possible while also minimizing the number of unnecessary warnings. Balancing the two is often difficult, given that if a forecaster warns a higher proportion of strong storms to avoid missing a potentially tornadic storm, that may increase the total POD but also increase the FAR. On the other hand, if a forecaster only warns the strongest storms to avoid issuing false alarms, they may have a low FAR but also a low POD. Brooks and Correia (2018) show that, from 1986-2016, POD at positive lead

times increased from ∼0.25 to a maximum of ∼0.7 in 2003 and 2011, before falling to near 0.5 in 2012 and beyond "when the default warning duration decreased, and there is an apparent increased emphasis on reducing false alarms" (Brooks and Correia 2018). In their study "Compared to What? Establishing Environmental Baselines for Tornado Warning Skill", Anderson-Frey and Brooks (2021) note that the actual POD for QLCS storms is approximately 48%. This is close to the POD of 49% for QLCS storms found in Brotzge et al. (2013). Anderson-Frey and Brooks (2021) also note that, in general, the FAR for all warnings is approximately 75%. These metrics establish a baseline against which the results of the CNNs trained, validated, and tested using GR-S data can be contextualized. Namely, a POD over 49%, FAR under 75%, and CSI over 0.2 (based on those POD and FAR values; see Roebber 2009, Equation (5) for detailed calculation) are each metrics of superior performance compared to the expected baseline for QLCS tornado warnings.

5.2 Results

5.2.1 CNN performance

5.2.1.1 Model 1

First, a machine learning model was built using the same tornadic vs. null storms from Section 4.1: 839 tornadic storms at both 10- and 20-minute lead times, and 3356 nontornadic storms selected such that their bulk low-level azimuthal shear properties match those of the tornadic storms. As shown in Section 4.3, these tornadic and strong nontornadic storms differ mainly in the collocation of low-level vorticity and convergence with the mid- to upper-level updraft (or lack thereof). By training a model specifically on the tornadic and nontornadic cases that both look promising for tornadogenesis from a bulk low-level azimuthal shear perspective, the hope is that the model would learn the nuances between these more challenging cases and, despite the lack of weaker nontornadic storms in the training and validation datasets, any weak storms included in final testing data would

still be overwhelmingly categorized as nontornadic based on the nuances learned comparing the tornadic and strong nontornadic storms. As mentioned in Section 5.1.1, these 5034 total storms are then divided randomly into a 4035 storm training and 999 storm validation dataset. This model trained and validated on tornadic and strong nontornadic cases is referred to as Model 1.

Information in GR-S Event Storm Track Files			
Hyperparameter	Available Options	Model 1	Model 2
2D Convolutional Layers	1, 2	2	1
Convolutional Kernel Size	3, 5, 7	3	5
Convolutional Layer Activation Function	Relu, Sigmoid, Tanh	Relu	Relu
Number of Kernels	4, 8, 16, 32	4	16
Dense Layers	1, 4	1	1
Dropout	0.05-0.5	0.49	0.07
Optimizer	Adagrad, Adam, RMSprop, SGD	Adagrad	Adagrad
Number of Neurons	4, 8, 16, 32	8	8
Dense Activation	Relu, Sigmoid, Tanh	Relu	Relu
Batch Normalization	Yes, No	No	No
Batch Size	64, 128, 256, 512, 1024	256	512

Table 5.1: Hyperparameter options for model tuning as well as final hyperparameter choice for models 1 and 2.

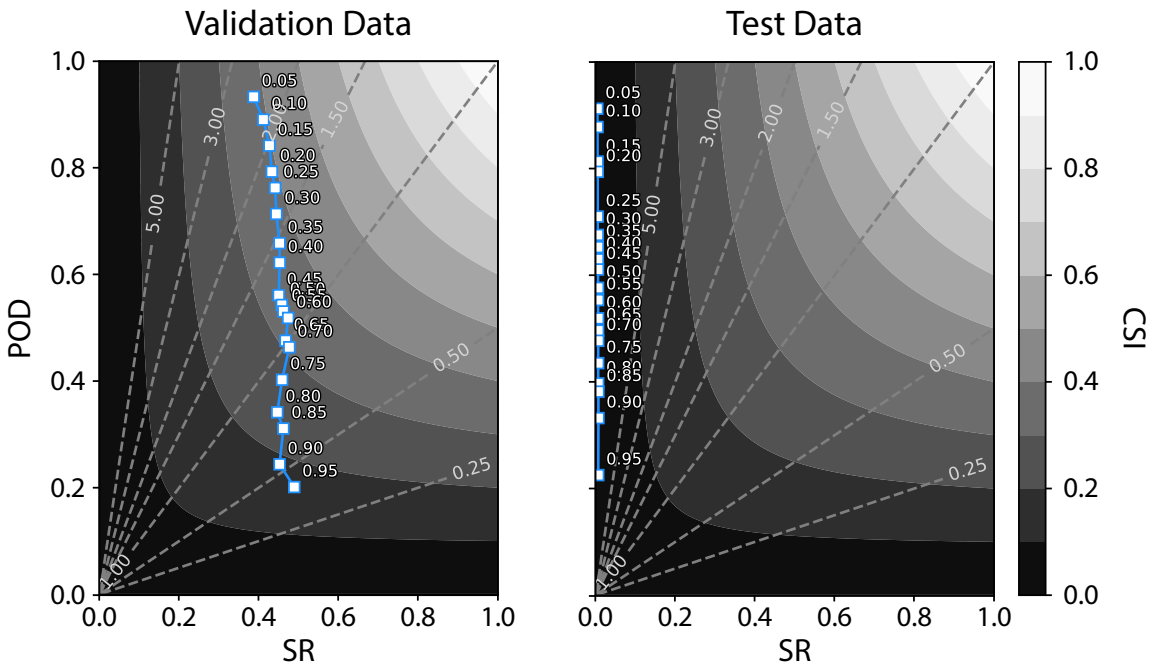

Figure 5.1: Performance diagrams for validation (left) and testing (right) data for Model 1. Testing data include all tornadic and nontornadic non-supercellular MCS storms from 2020-2021.

First, Model 1 was tested on all tornadic and nontornadic non-supercellular MCS storms from 2020-2021, which include a total of 262 tornadic and 70941 nontornadic storms, where tornadic storms are sampled at 20-minute lead times. Unlike how tornadic storms are sampled at multiple times during training and validation, for testing, we sample them only once at a 20-minute lead time, such that any performance metrics are representative of the true ratio of tornadic and nontornadic non-supercellular storms in nature. Figure 5.1 shows the performance diagram for the model trained and validated on the PDF-matched data and tested on all tornadic vs. nontornadic storms. Performance diagrams compare POD on the y-axis to success ratio (SR), equal to 1-FAR, on the x-axis. Shaded is the critical success index (CSI) which represents how successful the model is at accurately

characterising storms as tornadic and non-tornadic. The model predicts the percent probability that the storm is tornadic and then translates that probability into a boolean prediction of tornadic or nontornadic; points at 5% intervals from 5% to 95% are plotted.

Examining both plots in Figure 5.1, the SR for all percent thresholds is fairly constant at $\sim$50%, indicating that approximately half of all storms identified as tornadic by the model are actually nontornadic. Despite the relative insensitivity of SR to percent threshold, POD generally increases with decreasing threshold, to a maximum of $\sim$90% at a 5% tornadic probability threshold. These metrics–an FAR of 57%, POD of 84%, and a maximum CSI of nearly 0.4 at a 15% threshold for classification of a storm as tornadic–exceed all baseline metrics for QLCS tornado warnings.

However, the greatest measure of performance lies with testing of the model on an independent dataset, and those metrics are markedly less impressive. In fact, FAR is nearly 100% for all probability thresholds, indicating that the model is significantly overpredicting tornadoes. Although picking a low probability threshold results in an increasingly high POD nearing 90%, the maximum CSI value is 0.007, meaning that skill is comparably low and this model does not have skill beyond the baseline at improving tornado warning metrics for all QLCS storms.

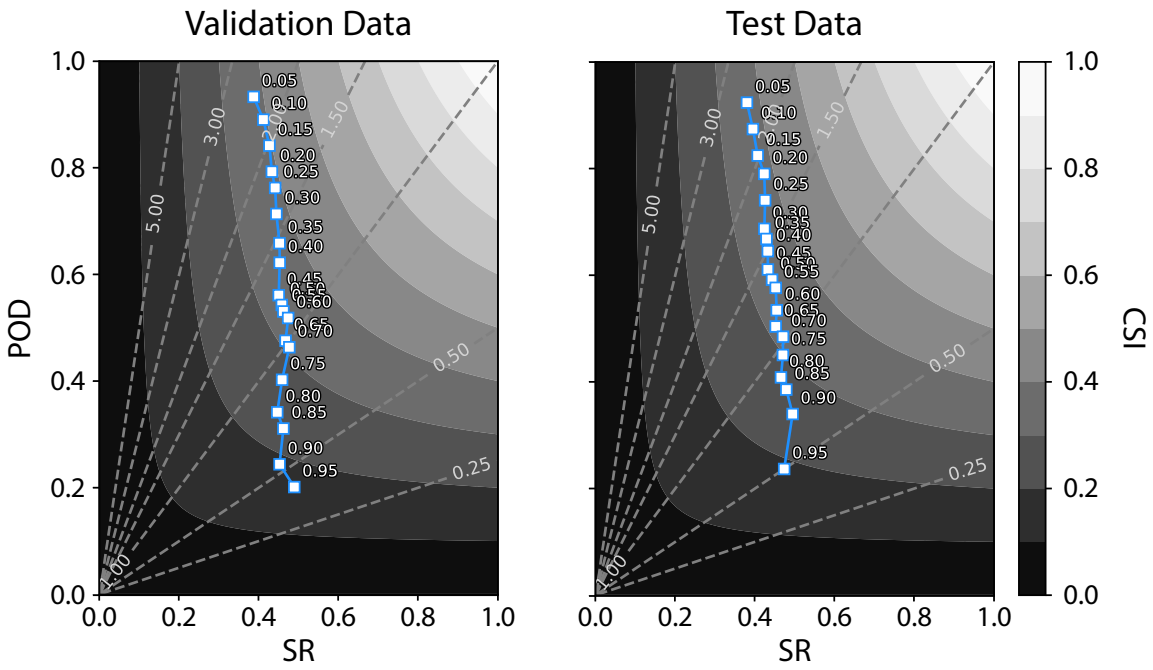

Figure 5.2: As in Fig. 5.1, but tested on all tornadic and only strong nontornadic non-supercelluar MCS storms from 2020-2021.

Recall that the testing dataset in Figure 5.1 included all non-supercellular MCS non-tornadic storms, whereas the training and validation sets included just the strongest non-tornadic cases. The assumption that the model would learn a substantial amount from the more difficult cases and be able to apply that knowledge to the weaker nontornadic cases produced little in meaningful results. Curious how the model would perform if tested on PDF-matched data similar to the training and validation datasets, testing was run again using a testing set including only tornadic and strong nontornadic cases. Similar to the PDF-matched training and validation datasets, these data also exist at a 2:1 nontornadic:tornadic ratio, with 262 tornadic storms from 2020-2021 at 20-minute lead times (as in the prior testing set) and an analogous nontornadic set of 524 storms selected to match the tornadic distribution of low-level azimuthal shear (e.g., see Section 4.1). Figure 5.2 shows the performance diagram for the model trained, validated, and tested on PDF-matched data. Validation results are identical given the identical training and validation datasets, but testing

results show marked improvements over those in Fig. 5.1. CSI values peak at 0.38 at the 20% probability threshold, where POD is 79% and FAR is 58%. These results show that, while the model trained and validated on tornadic and strong nontornadic cases does not have skill above the baseline at warning for non-supercellular MCS tornadoes in general, when isolating strong nontornadic storms in a 2:1 ratio to tornadic storms, the model has skill beyond the baseline at correctly discriminating between tornadic and nontornadic storms.

However, the results of this model and its *true* applicability are more complex than they first appear. The testing results when the model is trained, validated, and tested on PDF-matched data assumes a 2:1 ratio of nontornadic to tornadic storms, as was used in training and validation. However, the true ratio is closer to 270:1. That ratio is less extreme for nontornadic storms whose low-level azimtuhal shear lies within the tornadic distribution, but not much–this ratio is 96:1 and 176:1 when selecting only nontornadic storms with low-level azimuthal shear values within 1 and 2 standard deviations of the tornadic mean, respectively. Therefore, these attractive performance metrics may be misleading given how the ratio is skewed in the testing set compared to the true ratio of these storms in nature. Table 5.2 shows how these metrics would change given the same proportion of tornadic and nontornadic storms warned by the model if the full testing dataset was used. For example, the FAR, which is the ratio of false positives to all positives, is 0.58. Knowing the POD is 0.79, meaning that 207 of the 262 tornadic storms were warned (i.e., 207 true positives), that means that 286 of the 524 nontornadic storms were warned to acheive a FAR of 0.58. If we pretend that all nontornadic storms in the dataset are strong, to test the theory that the model struggles preferentially with weaker storms, then we can extrapolate the results we see here to the full 70941 storm dataset to determine model performance on the full set of (theoretically "optimally" strong) nontornadic storms. Taking that ratio of warned nontornadic to all nontornadic storms and scaling it up to the full testing dataset size, that means 38720 of the 70941 nontornadic storms were warned. In that scenario, the analogous

FAR for the whole dataset is 99.5%, also near the 100% FAR that Model 1 achieved with a full testing dataset. POD remains unchanged since the number of tornadic storms does not increase, but CSI changes drastically; again using Equation (5) from Roebber (2009), CSI with a full testing set would be 0.005, almost the exact same value from Model 1's results when evaluated on the full testing dataset.

Another way to examine these results where the model is tested only on strong storms, besides assuming that all nontornadic storms have high enough low-level rotation to be classified as strong (which, they do not) is to assume that the forecaster is able to screen out the weaker storms themselves before referring to the model to help with these more difficult cases. In this case, we assume that forecasters use this model only for storms with low-level azimuthal shear greater than 1-2 standard deviations below the mean tornadic value. These performance metrics improve, but not past the baseline already met by conventional forecasting techniques. Using the same methods to extrapolate the 2:1 testing results to nontornadic datasets only including storms that exceed 2 and 1 standard deviations below the mean of the tornadic storms' values, FAR decreases to 99.2% and 98.5%, and CSI increases to 0.008 and 0.01, respectively (Table 5.2). These metrics still pale in comparison to the baseline metrics of a POD $> 49\%$, FAR $< 75\%$, and CSI > 0.2 Therefore, despite Model 1's encouraging performance when using a PDF-matched dataset, it is the sample size of the dataset that appears to be the main driver of the favorable–but ultimately misleading–results.

5.2.1.2 Model 2

Given that Model 1 did not get near baseline metrics except for when the testing dataset size was skewed heavily towards tornadic storms, a new method of selecting the training and validation data was created. Specifically, instead of selecting nontornadic storms to emulate the general low-level azimuthal shear behavior of tornadic storms, the null dataset

Model 2, PDF-Matched Metrics Scaled to True Dataset					
Metric	PDF-Matched	Full Dataset	2-σ	1-σ	Baseline
FAR	58%	99.3%	99.1%	98.5%	$< 75\%$
POD	79%	79%	79%	79%	$> 49\%$
CSI	0.38	0.007	0.008	0.01	> 0.2

Table 5.2: Extension of Model 1 PDF-matched results to the full testing dataset, only nontornadic storms with low-level azimuthal shear above 2-σ below the tornadic mean, and only nontornadic storms with low-level azimuthal shear above 1-σ above the tornadic mean. Comparisons assume an equal proportion of tornadic and nontornadic storms are warned in both scenarios. Baseline metrics are also listed for comparison.

was randomly selected from all storms from 2010-2019 such that the null storms outnumber the tornadic storms 4:1. That way, when training and validating a model on nontornadic storms as well as tornadic storms at 10- and 20-minute lead time, the ratio of null to tornadic storms is still 2:1 (as in Section 4.1). While the former efforts focused on learning the differences between tornadic and strong nontornadic storms and assuming that those lessons would translate to effective identification of weaker nontornadic storms, this model explicitly includes nontornadic storms at a full range of observed intensities to remove that assumption. Testing data include the full suite of non-supercellular MCS tornadic and nontornadic data from 2020-2021, as in Model 1.

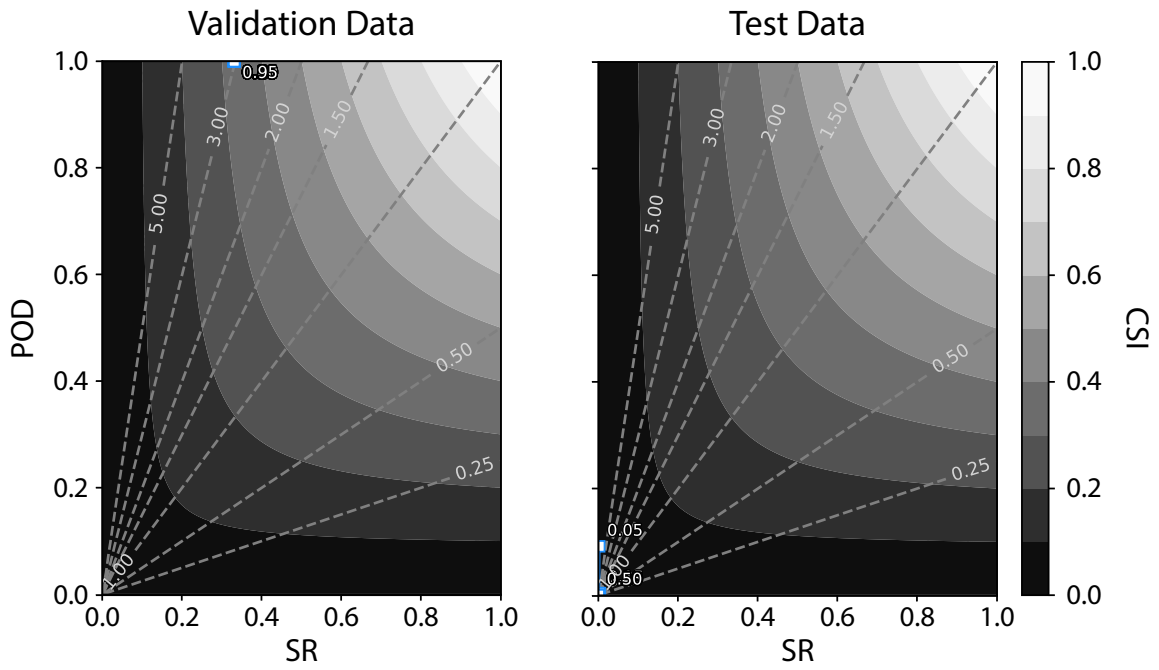

Figure 5.3: As in Fig. 5.1, but with the model trained and validated on a sample of all tornadic and all nontornadic non-supercelluar MCS storms from 2010-2019.

Figure 5.3 shows the performance diagram of the model trained, validated, and tested on samples representative of all tornadic and nontornadic non-supercellular MCS storms. It is evident from the extremely low POD, high FAR, and near-zero CSI on the test data performance diagram that this model (as it is currently constructed) has virtually no predictive skill and, at the very least, skill far lower than the baseline. This means that no model evaluated herein showed skill beyond the baseline at predicting tornadic potential of non-supercellular MCS storms. Further discussion of these models, and potential forward steps towards using ML to improve non-supercellular MCS tornado warnings, is presented in Chapter 7.

Chapter 6

Societal impact of non-supercellular MCS tornadic storms

Underlying any scientific discussion of non-supercellular MCS tornadic storms and the tornadoes they produce are the communities they impact. Given their lower POD and much higher propensity to be warned at negative lead times when compared to supercellular tornadoes (Brotzge et al. 2013), tornadoes from MCS/QLCS storms often impact communities with much less advance notice (on an individual storm scale) if any notice at all. That notice can have considerable impact on outcomes; Simmons and Sutter (2008) show a 41% decrease in fatalities and 47% decrease in injuries for tornadoes that are warned in advance (between 6-10 minutes and 11-15 minutes, respectively) compared to those that are not warned in advance. They specifically note that "warnings significantly reduce fatalities for short-track and weaker tornadoes," and since tornadoes produced by non-supercellular storms are typically sub-significant (e.g., Section 3.1) and short-lived, a direct link between low lead times for MCS tornadoes and increased societal impact (via deaths and injuries), although complicated by confounding factors including warning communication and societal response, can be reasonably inferred. Herein, we aim to examine the underlying demographic makeup and social vulnerability of communities impacted by non-supercellular MCS tornadoes, how those demographic characteristics compare to CONUS-mean characteristics, and what this could tell us about any potential disproportionate impacts of MCS tornadoes and the current brevity (or absence) of warnings for these storms on various demographic groups or communities with varying social vulnerabilities.

6.1 Data and methodology

This investigation also relies heavily on the GR-S dataset (see Chapter 2), where individual tornadic storms can be isolated for analyses. In addition, information about deaths

and injuries associated with these tornadoes is sourced from the SED dataset (NCEI/NOAA 2022). To get information about the demographic breakdown and social vulnerability of the broader population in the vicinity of reports, census data were sourced from the Columbia University Socioeconomic Data and Applications Center (SEDAC; Center For International Earth Science Information Network-CIESIN-Columbia University 2017a,b). These data include demographic information such as age, race, and ethnicity for locations within the CONUS at a $\sim$1 km spatial resolution. The Social Vulnerability Index (SVI; Center For International Earth Science Information Network-CIESIN-Columbia University 2021a,b) is also used to determine the vulnerability of a location. Cutter and Finch (2008) note that social vulnerability "identifies sensitive populations that may be less likely to respond to, cope with, and recover from a natural disaster." SVI values are calculated using four main demographic characteristics: socioeconomic status, household composition and disability, minority status and language, and housing type and transportation (Center For International Earth Science Information Network-CIESIN-Columbia University 2021a). SVI values range from 0 to 1, with 0 being the least vulnerable and 1 being the most vulnerable.

Both demographic and SVI data from 2010 were used as estimates of CONUS-wide population characteristics. Table 6.1 outlines what demographic groups will be considered herein. These include elderly populations and all "Race and Ethnicity" categories available from SEDAC. Table 6.1 also introduces all category abbreviations as in the SEDAC dataset. Investigations herein will focus on population densities in the vicinity of tornadoes versus a given demographic group's CONUS-mean population density, which is the average of all local population densities at each gridpoint for that group.

6.2 Results

Of primary interest is the mean demographic breakdown of locations impacted by non-supercellular MCS cells compared to both the CONUS-average population breakdown and

Demographic Categories	
Category	Description
A6	Ages 65 to 79
A7	Age 80 and Older
WH	White Alone
BL	Black or African American Alone
AM	American Indian and Alaska Native alone
AS	Asian Alone
PI	Native Hawaiian and Other Pacific Islander Alone
ORA	Some Other Race Alone
TWO	Two or More Races
NHI	Not Hispanic
HI	Hispanic
NHW	Non-Hispanic White
NHB	Non-Hispanic Black
SVI	Social Vulnerability Index

Table 6.1: Description of different population subsets analyzed herein and their abbreviations as given within the original SEDAC dataset.

the mean demographic breakdown of regions impacted by discrete supercells. The comparison is made to discrete supercells since these are often of primary interest in studies of severe weather and tornadoes and, perhaps as a result of this enhanced focus, Brotzge et al. (2013) found that "tornado warnings for supercells had a statistically higher probability of detection (POD) and lead time than tornado warnings for nonsupercells" and that discrete supercells or those in clusters were slightly less "difficult" to warn for compared to those in lines. Comparing the demographic breakdown for tornadic storms that have higher POD

and anecdotally more research focus to those that do not may add a quantitative equity viewpoint to future selection of research foci in meteorology.

Similar to Section 4.1, discrete supercells and non-supercellular MCS storms are isolated for analyses if that combination of storm mode and supercell classification both existed at the time of the report and existed for $\geq 75\%$ of their lifetimes. In that way, storms are isolated that exhibited specific mode and dynamic characteristics both for the majority of their lifetimes and at the time of the severe weather. Samples include 2471 tornadic non-supercellular MCS storms and 693 discrete supercellular storms, not all of which produced impacts as recorded within the SED dataset. Future mention of these storms is simplified to MCS and supercellular storms, respectively.

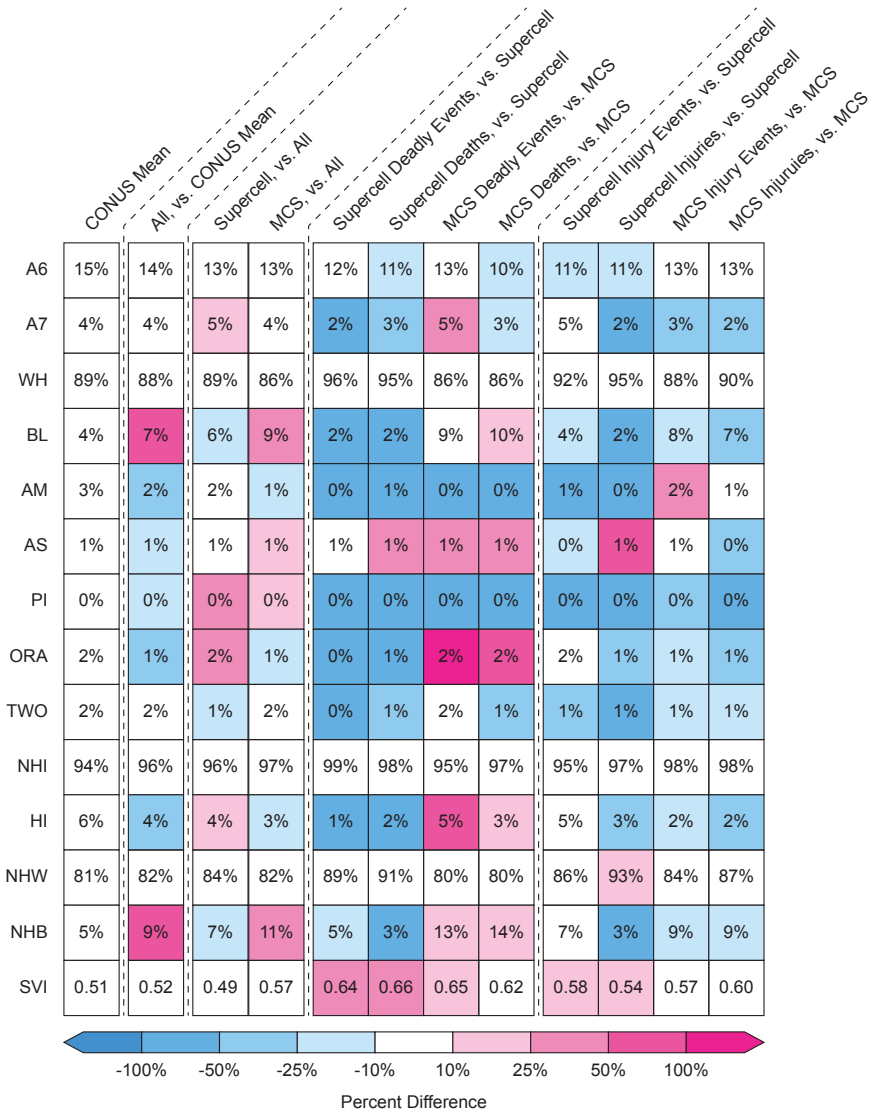

Figure 6.1: Percent contribution of (rows) various demographic groups to (columns) the local population. Data are sorted by CONUS-mean density and average density at locations of all tornadoes, all tornadoes by mode, all deadly tornadoes by mode, and all tornadoes associated with injuries by mode. For both deaths and injuries, the mean demographic contribution is calculated both on a per-event and per-impact (i.e., per-death or per-injury) basis. Color shading represents percent change between the data in that column and the column the data are being compared to, described in each column's title at the top of the plot.

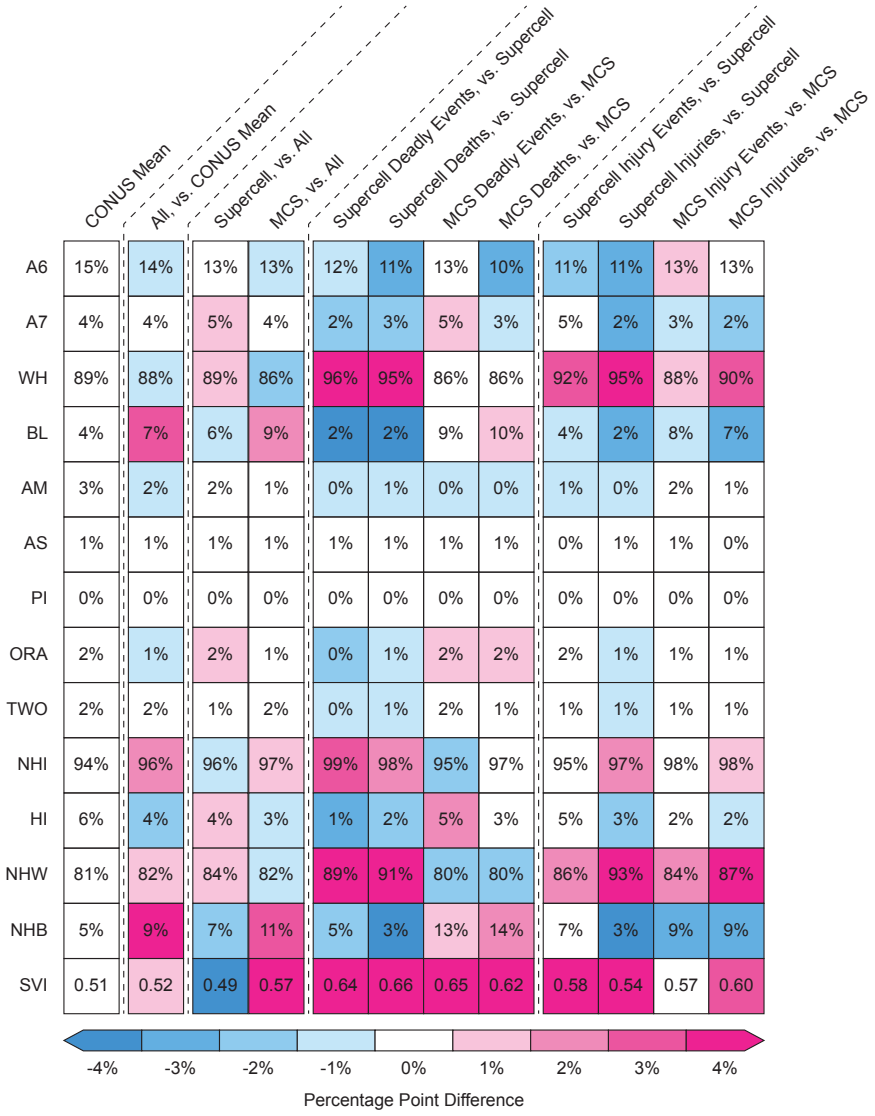

Figure 6.2: As in Fig. 6.1, but shading indicates percentage point change.

Layers of insight can come from examining the mean demographic characteristics at the locations of tornadoes. By incrementally examining areas hit by tornadoes that match more and more criteria (e.g., CONUS mean densities vs. all tornadoes vs. MCS tornadoes vs. deadly MCS tornadoes), observations can be made about how demographics change incrementally as the type of tornado and its impact gets more specific. Figures 6.1 and 6.2 show the underlying percent contribution of demographic groups in Table 6.1, including CONUS-mean densities, and average density at locations of all tornadoes, all tornadoes by

mode, all deadly tornadoes by mode, and all tornadoes associated with injuries by mode. Colors are representative of either 1) percent differences in mean percent contribution (i.e., percent change of the percents; Fig. 6.1) or 2) overall change in mean percent contribution (i.e., absolute difference of the percents; Fig. 6.2) when comparing data in one column to data in another. While the values are the same in both figures, the different approaches to analyzing change between two given categories allows for understanding of how much a population's contribution changes relative to their original contribution and relative to the total population in that area. For demographic groups with low relative population densities, a smaller absolute change would still lead to a larger percent change; for demographic groups with high relative population densities, a relatively small percent change could be a large absolute change.

In Fig. 6.1, percent differences are shaded, which preferentially highlight changes in demographic groups with low relative population densities. Focusing on rows from American Indian and Alaska Native alone (AM) to Two or More Races (TWO), all which have a CONUS mean density of 3% or less, these demographic groups have higher relative densities on average across the CONUS than they do in the vicinity of tornadoes. When focusing on supercellular or MCS tornadoes, relative densities for a few categories are slightly elevated from the baseline of all tornadoes, specifically Native Hawaiian and Other Pacific Islander Alone (PI) and Some Other Race Alone (ORA) which have 25-50% higher relative densities in the vicinity of supercells compared to all tornadoes. Segmenting the data further by deaths or injuries associated with supercellular or MCS tornadoes, there are a mixture of increases and decreases in relative population density compared to all supercell or all MCS tornadoes, with no (albiet rounded) percents shown greater than the CONUS mean.

Particularly interesting are any compounding disproportionalities, specifically compounding positive percent changes. If for any group, relative population densities continue

to increase/decrease with increasing granularity, this is an indication that their local population contribution continues to increase/decrease when moving from CONUS average relative densities all the way to particularly impactful tornadoes of a given mode. For these lower relative population density rows in Fig. 6.1, the only positive compounding disproportionality is the increase in local Asian Alone (AS) relative population density from all tornadoes to MCS tornadoes and further to MCS deadly events and total deaths. This indicates that, while the AS demographic is typically lower relative to other groups in the vicinity of tornadoes, it is higher for MCS tornadoes and higher yet for deadly MCS tornadoes. Note that this change is still small enough to disappear when analyzing rounded mean relative population densities.

Moving to Fig. 6.2, absolute changes in percent contributions are shaded, which preferentially highlight changes in demographic groups with high relative population densities. Focusing on rows besides AM to TWO, the Black or African American Alone (BL) demographic has higher relative population densities in the vicinity of tornadoes than their CONUS mean values. Also higher in the vicinity of tornadoes are the Not Hispanic (NHI), Non-Hispanic White (NHW), and Non-Hispanic Black (NHB) populations. Looking at specifically supercellular and MCS tornadoes, the elderly population (Ages 65 to 79 and Age 80 and Older; A6 and A7, respectively) has fairly similar population densities in the vicinity of supercellular and MCS tornadoes compared to all tornadoes, and typically lower density when focus shifts to just tornadoes associated with deaths or injuries. For supercellular tornadoes, White Alone (WH), ORA, Hispanic (HI), and NHW populations have higher densities compared to all tornadoes, and for MCS tornadoes, BL, NHI, and NHB populations have higher densities. Particularly interesting for the WH population are the dark pink gridboxes for supercellular deaths and injuries, representing a 3-7 percentage point increase in relative population density compared to all supercellular tornadoes and indicating deadly supercells occur in areas with disproportionately high WH populations.

Comparing both Figs. 6.1 and 6.2, any gridboxes that are shaded the same color represents a difference that is sufficiently large to be noticed both in a percent and absolute sense. Examples include gridboxes in the A7 and ORA demographic groups, although these are not compounding differences. The demographic groups that stick out the most, both comparing Figs. 6.1 and 6.2 and examining Fig. 6.2 alone are the BL and NHB demographic groups. Both of these groups have positive compounding disproportionalities whether examining demographic density increases by percent or percentage point increase. Specifically, BL and NHB relative densities are higher in the vicinity of tornadoes compared to the CONUS mean; higher in the vicinity of MCS tornadoes compared to all tornadoes; and higher in the vicinity of deadly MCS tornadoes (total deaths for BL; total deaths and by deadly event for NHB) compared to all MCS tornadoes. This means that there is a sizable step-by-step increase from their CONUS mean values to their relative densities in the vicinity of deadly MCS tornadoes, highlighting the disproportionate impact of deadly MCS tornadoes on the BL and NHB communities.

Lastly, SVI is also particularly interesting as data granularity increases. Comparing SVI near all tornadoes to the CONUS mean, they are nearly equal. However, breaking tornadoes down by supercell vs. MCS, the mean supercellular tornado SVI is 0.49 while that for MCS tornadoes is 0.57, a 16% increase. Therefore, MCS tornadoes tend to occur in areas that are more socially vulnerable than supercellular tornadoes. Moving from SVI by mode to SVI by mode and impact, Figs. 6.1 and 6.2 show increases in SVI for most categories compared to all supercellular or all MCS tornadoes. This is perhaps a reflection of the accuracy of the SVI index in identifying areas that are less resilient to natural disasters. An in-depth discussion of the nuances of this work is presented in Chapter 7.

Chapter 7

Discussion

The GR-S database was created to facilitate robust studies of severe weather using radar data from a large sample of storms. Namely, the objective methods used to build the now-public dataset (School of Meteorology, University of Oklahoma 2021) provide an opportunity to easily investigate thousands of severe storms and over 1.3 million total storms with great detail. We believe these data can be used to examine several challenging and important science questions regarding severe weather and we encourage others to use GR-S to explore their own scientific questions as the dataset continues to grow.

Given the strong, positive azimuthal shear collocated with the parent storm updraft at mid- to upper-levels, concerns about contamination of the non-supercellular MCS storm database with spurious (or many) supercell storms are warranted. Magnitudes between azimuthal shear values herein and those in PMM studies of supercells done by Homeyer et al. (2020) are comparable, but that is not particularly surprising; Weisman and Trapp (2003) note that QLCS mesovortices can be "similar in size and strength to mesocyclones associated with supercell storms." They also note that, despite this similarity, supercells are marked by long-lived and persistent midlevel rotation, whereas QLCS mesovortices are not. This is consistent with the supercell classification herein, where any storm that did not have maximum midlevel (4-7 km AMSL) azimuthal shear exceeding $4 * 10^{-3}$ s^{-1} for at least 40 minutes was not classified as a supercell. Murphy et al. (2023) used the GR-S dataset to compare MCS non-supercellular storms to the results of past studies that used manually selected MCS storms. Those comparisons found that the GR-S non-supercellular MCS storms compared closely with the manually identified MCS storms, and GR-S supercellular MCS storms had characteristics more consistent with traditional supercells. For

these reasons, we are confident that the non-supercellular MCS population examined herein is not adversely impacted by contamination from supercellular storms.

Analyses herein show that probability-matched mean radar data, particularly single-polarization data, reveal key low- to mid-level features that can be used to discriminate between tornadic and nontornadic storms. These statistical results may be useful to improving warning decision-making practices for forecasters, since we find unique kinematic signatures that persist at long lead times to tornadogenesis in tornadic storms and differ considerably in comparable nontornadic storms. Given that a vertically-aligned rotating updraft is a major indicator of tornadic potential in storms analyzed herein, three-dimensional mesovortex evaluation in future high-resolution case studies of these events will be important to validate the perceived importance of low- to mid-level updraft alignment in non-mesocyclonic tornadogenesis. This result is particularly notable in that it agrees with comparable prior work focused solely on supercell storms (Homeyer et al. 2020), which suggests that vertical alignment of rotation and the low- to mid-level updraft are common requirements of tornadic storms regardless of storm mode. Relatedly, novel predictive methods such as artificial intelligence may benefit from focusing on the vertical alignment of mesovortices in future studies to improve warning performance for non-supercellular MCS tornadoes (or tornadoes in general), especially their POD.

After determining that single-polarization data are most important in isolating tornadic from nontornadic non-supercellular MCS storms, those single polarization data were used to train, validate, and test a convolutional neural network to ideally identify tornadic storms at skills exceeding what is already common within the NWS. This was not the case; the model only had skill exceeding that of the NWS when the testing set was artificially heavily weighted towards tornadic storms, inflating the perceived performance of the model that, in reality, would have a CSI below 0.01 if tested on the full testing dataset. Potential avenues for future studies to improve model performance may include feeding the model images that cover a larger area of the storm and potentially around it (i.e., the entire MCS);

restricting data to storms within a set distance from the radar; removing storms that have non-negligible missing data at low- and mid-levels; including environmental information (as in Lagerquist et al. 2020); making models specifically trained for different regions of the U.S., times of day, and/or times of the year; or removing weaker storms from models trained and validated on all storms, therefore requiring the model to learn less features that a forecaster already knows and can screen out themselves.

Speaking of forecaster knowledge, the most promising immediate pathway to improving this work may be replacing the null storms with nontornadic storms that were warned but were not matched within the GR-S dataset to a tornado (i.e., a false alarm storm). Regardless of the increase in computational tools to aid in tornado prediction, forecasters will remain an integral part of detecting and warning for tornadoes. Therefore, while a model built purely on objective identification of tornadic storms from all storms of a given subtype is perhaps the most straightforward way to approach this problem, since it pales in comparison to the baseline performance already acheivable by NWS forecasters, it may be advantageous to build a model on the foundation of the knowledge that they already have. By using the library of nontornadic storms that forecasters warned on as the null population, one can compare tornadic storms to those that "appear" tornadic as in Chapters 4-5; however, the appearance of tornadic potential isn't quantitative similarity, but the subjective viewpoint of the people who will actually issue those warnings. This means that some subjectivity, human error, and potential human biases may be a part of the foundation of the model. However, using what forecasters are already doing and building upon that to try and produce a model useful to forecasters may yield more promising results.

Analyses of population densities and SVI scores for communities in the vicinity of tornadoes show an increasing BL and NHB relative population density with data granularity in the vicinity of all tornadoes, MCS tornadoes, and deadly MCS tornadoes. This suggests that MCS tornadoes, including deadly ones, are typically located in areas with higher BL and NHB populations compared to both the CONUS mean and the mean population density

near all tornadoes. Examining WH populations, relative population densities are near the CONUS mean for all, supercellular, and MCS tornadoes, but these relative densities jump in the vicinity of deadly (and injury-causing, although less so) supercellular tornadoes, suggesting deadly supercellular tornadoes preferentially track through areas with higher WH populations. Finally, SVI is near the CONUS mean for all tornadoes but is lower for supercellular tornadoes and higher for MCS tornadoes, highlighting a discrepancy in the underlying vulnerability of communities hit by supercellular vs. MCS tornadoes.

It is reasonable to infer that a community with higher SVI (and therefore one identified as "less likely to respond to, cope with, and recover from a natural disaster"; Cutter and Finch 2008) may suffer higher impacts from severe weather than a community that is quantitatively more resilient. Cutter et al. (2003) outline many of the factors that go into the calculation of a community's SVI in their Table 1, some of which are quite familiar to the meteorological community. For example, they highlight the "Residential Property" as a risk factor, stating that "mobile homes are easily destroyed and less resilient to hazards." Previous meteorological work has shown that mobile homes are the location of the majority of tornado-related deaths in the Southeast and "[t]he likelihood of a tornado fatality in [a mobile home] is 15–20 times greater than in permanent homes" (Strader and Ashley 2018, and references therein). Additionally, Cutter et al. (2003) note that people of color have increased vulnerability, in part due to potential language barriers that may inhibit access to disaster recovery funding. Recent work by Trujillo-Falcón et al. (2021) further discusses the impact of language barriers on severe weather preparedness, focusing on a community's potential inability to access life-saving weather information due to language barriers and/or misunderstandings depending on how a word is translated. Improving MCS tornado warnings (alongside improvements in warning communication and appropriate social response to warnings; see Chapter 1) therefore has the potential to preferentially improve outcomes for the BL and NHB communities and quantitatively more vulnerable communities, highlighting potential improvements in equitable outcomes from severe weather by focusing on

further understanding of MCS tornadogenesis and improvements in MCS tornado warnings.

One limitation of the societal impact portion of the dissertation is the author herself. Switching to a first-person discussion for just a moment, I am fully aware that I am a meteorologist and not a social scientist. The purpose of this section was to discuss the human element of non-supercellular MCS tornadoes–who they disproportionately impact and how their impacts differ from supercellular tornadoes. I see this work as a starting point for a more nuanced study of the human impacts of different types of tornado-producing storms. At the time of writing this document, great care was taken to discuss age, race, and ethnicity in a careful and respectful manner. I recognize that discussing human impacts and comparing totals of deaths and injuries can come across as cold and distant–I emphasize that the goal here was to introduce the relevant data that can be extracted from a combination of the GR-S and SEDAC datasets, given my expertise in data analytics but limited experience in social science, and refrain from diving into further discussions of social sciences outside of my area of expertise. Work is ongoing outside of the scope of this dissertation to continue analyzing GR-S and SEDAC data in tandem with a team of researchers with diverse skill sets such that a final product can represent the full nuances of this type of discussion and not be limited to just number crunching. Please contact me with any questions, comments, or concerns about this work.

Chapter 8

Conclusions

In this study, the GR-S dataset was introduced, a dataset centered on CONUS-wide radar data for ∼100 of the most severe days per year from 2010-2019, inclusive. After determining which days to include in the database, spatiotemporal domains for radar data are selected objectively, and all storms within the domain are tracked throughout their lifetimes. Storms are matched with severe reports and both storm mode classification (single cell, multicell, or MCS) and supercell classification (mesocyclonic or non-mesocyclonic, for all three storm modes) are performed. Based on the analysis presented, the following conclusions can be drawn:

3.1) The GR-S dataset captures a majority of SED reports from 2010-2019 inclusive, and captures ∼90% of reports that exist within the spatiotemporal bounds of GR-S (Table 3.1). The reports captured are analogous in spatial and temporal distribution to the SED reports, and the relative contribution of tornado, hail, and wind reports to all reports per month also mirror that of the SED database quite well (Figs. 3.1–3.3). Therefore, the storm-matched reports within the GR-S database are a representative sample of the complete SED dataset.

3.2) Pronounced annual and diurnal variability was evident for tornado, hail, and wind reports: (i) MCS-classified cells produced the most tornadoes throughout the entire year when compared to other storm modes, and mesocyclonic storms were found to be the primary tornado contributors during peak tornado frequency in spring and summer (Fig. 3.4a). During the overnight and early morning when total number of reports is low, MCS contribution is maximized, and during the time of peak reports, mesocyclonic storms produce the majority of tornado reports (Fig. 3.5a). (ii) Single cell storms account for the largest

fraction of hail reports by storm mode for spring through fall, and mesocyclonic storms account for a majority of all hail reports both year-round (Fig. 3.4b) and throughout the day (Fig. 3.5b). Hail reports have the most pronounced diurnal cycle of the three report types, meaning that reports are highly concentrated around their time of peak occurrence ($\sim$4-5 hours after local solar noon). Unlike tornadoes and wind reports, there is no one storm mode that stands out as the primary producer of severe hail reports over the whole year or whole day. (iii) MCS-classified cells produce the most severe wind reports throughout the entire year (Fig. 3.4c) and day (Fig. 3.5c) when compared to other storm modes. There is a peak in non-mesocyclonic single cell and multicell contribution to wind reports in the late summer and early afternoon, likely due to decay of severe weakly forced thunderstorms.

3.3) For both tornado and hail reports, as EF-rating and hail size increase, so does the relative contribution of mesocyclonic storms to total reports (Fig. 3.6). Wind reports show a similar trend, but the contribution of mesocyclonic storms does not monotonically increase with increasing wind speed. However, due to small sample sizes at the highest intensities, such breakdowns should be interpreted carefully.

3.4) GR-S was found to broadly reproduce the findings of Trapp et al. (2005) (Fig. 3.7) and Ashley et al. (2019) (Table 3.3), lending credence to the usefulness of the GR-S dataset and quality of storm mode classification applied herein.

Additionally, this study examined the radar appearance of tornadic and nontornadic non-supercell MCS storm cells, focusing on nontornadic storms that otherwise look favorable for tornadogenesis (i.e., the null population). The major conclusions of this work are the following:

4.1) Comparisons of tornadic and nontornadic kinematic characteristics show that low-level, mid-level, and column-maximum azimuthal shear, as well as low-level divergence, have the greatest quantitative difference (as measured by K-S tests) between the tornadic and nontornadic storm populations. Focusing on low-level azimuthal shear values, null storms were randomly selected such that the distribution of null low-level azimuthal shear

values matches the distribution of tornadic values. Sensitivity tests showed no meaningful differences in the distribution of tornadic and null values of other kinematic quantities based on the random sample. This resulted in 1627 tornadic and 1625 nontornadic (null) storms retained for analysis.

4.2) Analyzing PMM data at constant altitudes above ground level, Z_{H}, Z_{DR}, and K_{DP} data show a more concentrated lower-left to upper-right (relative to storm motion) axis for tornadic storms compared to null storms, potentially indicating a preferred storm motion-relative orientation for MCS line structures containing tornadic cells.

4.3) Comparing the kinematic characteristics of tornadic and null storms, much of the differences arise in the vertical alignment of low-level features conducive to stretching of vertical vorticity by the parent storm's mid- to upper-level updraft. In tornadic storms, maxima of azimuthal shear (rotation) and convergence are collocated, such that low-level rotation can be advected/stretched into the vertical by the low-level updraft. Low-level signatures of vertical motion are collocated with the mid- to upper-level updraft, allowing for further intensification of vorticity stretching and creating a conducive environment for tornadogenesis. Null storms differ in that they are characterised by a more diffuse field of low-level positive azimuthal shear horizontally displaced from a similarly diffuse field of low-level convergence. These signatures are also displaced from the mid- to upper-level updraft, so even weak vertical tilting of vorticity at low levels is likely decoupled from the parent storm updraft, limiting tornadogenesis potential via stretching. Therefore, enhanced low-level convergence and azimuthal shear coincident with one another and with the parent storm updraft are likely strong single-polarization indicators of tornadic potential for MCS non-supercellular storms. Low-level azimuthal shear appearing as a discriminating characteristic between tornadic and null storms–despite using the distribution of low-level azimuthal shear values in tornadic storms to select the null population–emphasizes that not only the magnitude, but also the location of low-level azimuthal shear is important for identifying potentially tornadic storms.

4.4) In contrast to single-polarization data, dual-polarization fields show little discriminating abilities between tornadic and null MCS storms. The main differences between fields are with the storm's orientation relative to storm motion, which is also observable in Z_{H} data. For this reason, dual-polarization data offer little information beyond that available in single-polarization data to aid in identifying tornadic storms.

Once relevant radar fields for MCS non-supercellular tornado detection were identified, these fields were fed into a machine learning model–specifically, a CNN–to measure the ability of a ML model to predict whether or not a storm will produce a tornado in the next 20 minutes. Hyperparameters were chosen through an extensive test to identify those that maximize model performance. This resulted in the following key conclusions:

5.1) Models that were tested on datasets that included all tornadic and nontornadic non-supercellular MCS storms had performance metrics (i.e., POD, FAR, and CSI) that paled in comparison to the NWS baseline, regardless of whether the model's training and validation datasets included all nontorandic storms or just strong nontornadic storms. These metrics looked far better when testing on a dataset with exclusively tornadic and strong nontornadic storms; however, the skewed tornadic:nontornadic ratio appeared to be the main driver of this artificially inflated performance, not the preferentially selected nontornadic storms. Therefore, the models as presented herein have little to no standalone skill at differentiating tornadic and nontornadic non-supercellular MCS storms, and any skill they do have is far lower than the skill already demonstrated by NWS forecasters.

Finally, understanding that warnings and warning performance do not operate in a vacuum and are a combination of meteorological and social factors, the demographics underlying the locations of severe reports were examined. Using census data, both the local densities of certain demographic groups and the local social vulnerability were examined in the vicinity of both MCS non-supercellular tornadoes and discrete supercellular tornadoes

that were linked with injuries and/or deaths. In this way, the typical demographic characteristics of locations hit by particularly impactful tornadoes can be examined. Conclusions from these analyses are as follows:

6.1) Examining the underlying demographics in areas collocated with tornado reports, the BL and NHB populations are disproportionately likely to have higher densities in the vicinity of all tornadoes compared to their CONUS mean population densities. Furthermore, this density increases when isolating non-supercellular MCS tornadoes, and remains constant or increases when isolating deadly non-supercellular MCS storms. For WH populations, population density in the vicinity of tornadoes is fairly near to the CONUS mean regardless of which type(s) of tornadoes are examined, but density jumps when considering deadly discrete supercellular tornadoes. This points to deadly and injurous supercells preferentially occuring in areas with disproportionately high WH populations, which makes anecdotal sense given the preference for supercells to occur in the Great Plains and the relatively higher WH population in that region (e.g., Center For International Earth Science Information Network-CIESIN-Columbia University 2017b).

6.2) In addition to higher BL and NHB populations, the data show that, while SVI is comparable between CONUS mean values and values in the vicinity of all tornadoes, SVI increases sharply for locations of non-supercellular MCS tornadoes. This is in contrast to SVI near supercellular tornadoes, which decreases slightly from the CONUS mean and from SVI values near all tornadoes. SVI for all deadly and injurous tornadoes is equal to or elevated above SVI for all tornadoes of that particular mode and supercell classification, likely highlighting the ability of SVI to identify populations that are particularly vulnerable to natural hazards.

Bibliography

Agee, E. and E. Jones, 2009: Proposed Conceptual Taxonomy for Proper Identification and Classification of Tornado Events. *Weather and Forecasting*, **24**, 609 – 617, doi: 10.1175/2008WAF2222163.1.

Agee, E. and L. Taylor, 2019: Historical Analysis of U.S. Tornado Fatalities (1808–2017): Population, Science, and Technology. *Weather, Climate, and Society*, **11**, 355 – 368.

Ahlborn, L., J. M. Franc, and D. S. Med, 2012: Tornado Hazard Communication Disparities among Spanish-Speaking Individuals in an English-Speaking Community. *Prehospital and Disaster Medicine*, **27 (1)**, 98–102, doi:10.1017/S1049023X12000015.

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

Amburn, S. A. and P. L. Wolf, 1997: VIL Density as a Hail Indicator. *Weather and Forecasting*, **12**, 473 – 478.

Anderson-Frey, A. K. and H. Brooks, 2019: Tornado Fatalities: An Environmental Perspective. *Weather and Forecasting*, **34**, 1999 – 2015.

Anderson-Frey, A. K. and H. Brooks, 2021: Compared to What? Establishing Environmental Baselines for Tornado Warning Skill. *Bulletin of the American Meteorological Society*, **102**, E738 – E747.

Anderson-Frey, A. K., Y. P. Richardson, A. R. Dean, R. L. Thompson, and B. T. Smith, 2016: Investigation of Near-Storm Environments for Tornado Events and Warnings. *Weather and Forecasting*, **31**, 1771 – 1790.

Ashley, W. S., 2007: Spatial and Temporal Analysis of Tornado Fatalities in the United States: 1880–2005. *Weather and Forecasting*, **22**, 1214 – 1228.

Ashley, W. S., A. M. Haberlie, and V. A. Gensini, 2023: The Future of Supercells in the United States. *Bulletin of the American Meteorological Society*, **104**.

Ashley, W. S., A. M. Haberlie, and J. Strohm, 2019: A Climatology of Quasi-Linear Convective Systems and Their Hazards in the United States. *Weather and Forecasting*, **34**, 1605 – 1631.

Ashley, W. S., A. J. Krmenec, and R. Schwantes, 2008: Vulnerability due to Nocturnal Tornadoes. *Weather and Forecasting*, **23**, 795 – 807.

Aydin, K., T. A. Seliga, and V. Balaji, 1986: Remote Sensing of Hail with a Dual Linear Polarization Radar. *Journal of Applied Meteorology and Climatology*, **25**, 1475 – 1484.

Bentley, M. L. and T. L. Mote, 1998: A Climatology of Derecho-Producing Mesoscale Convective Systems in the Central and Eastern United States, 1986–95. Part I: Temporal and Spatial Distribution. *Bulletin of the American Meteorological Society*, **79**, 2527 – 2540.

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. *Electronic J. Severe Storms Meteor.*, **6 (7)**, 1–30.

Bluestein, H. B. and M. H. Jain, 1985: Formation of Mesoscale Lines of Pirecipitation: Severe Squall Lines in Oklahoma during the Spring. *Journal of Atmospheric Sciences*, **42**, 1711 – 1732.

Bluestein, H. B. and S. S. Parker, 1993: Modes of Isolated, Severe Convective Storm Formation along the Dryline. *Monthly Weather Review*, **121**, 1354 – 1372.

Boukabara, S.-A., et al., 2021: Outlook for exploiting artificial intelligence in the earth and environmental sciences. *Bulletin of the American Meteorological Society*, **102**, E1016 – E1032, doi:10.1175/BAMS-D-20-0031.1.

Brooks, H. and C. A. Doswell, 2001: Some aspects of the international climatology of tornadoes by damage classification. *Atmospheric Research*, **56 (1)**, 191–201.

Brooks, H. E. and J. Correia, 2018: Long-Term Performance Metrics for National Weather Service Tornado Warnings. *Weather and Forecasting*, **33**, 1501 – 1511.

Brooks, H. E., C. A. Doswell, and M. P. Kay, 2003: Climatological Estimates of Local Daily Tornado Probability for the United States. *Weather and Forecasting*, **18**, 626 – 640.

Brotzge, J. A., S. E. Nelson, R. L. Thompson, and B. T. Smith, 2013: Tornado Probability of Detection and Lead Time as a Function of Convective Mode and Environmental Parameters. *Weather and Forecasting*, **28**, 1261 – 1276.

Brown, R. A., L. R. Lemon, and D. W. Burgess, 1978: Tornado Detection by Pulsed Doppler Radar. *Monthly Weather Review*, **106**, 29 – 38.

Browning, K. A., 1964: Airflow and Precipitation Trajectories Within Severe Local Storms Which Travel to the Right of the Winds. *Journal of Atmospheric Sciences*, **21**, 634 – 639.

Bunkers, M. J., M. R. Hjelmfelt, and P. L. Smith, 2006: An Observational Examination of Long-Lived Supercells. Part I: Characteristics, Evolution, and Demise. *Weather and Forecasting*, **21**, 673 – 688.

Burke, S., J. W. Bethel, and A. F. Britt, 2012: Assessing Disaster Preparedness among Latino Migrant and Seasonal Farmworkers in Eastern North Carolina. *International Journal of Environmental Research and Public Health*, **9 (9)**, 3115–3133, doi:10.3390/ijerph9093115.

Byers, H. R. and R. R. Braham, 1949: The Thunderstorm: Report of the Thunderstorm Project. U.S. Government Printing Office.

Center For International Earth Science Information Network-CIESIN-Columbia University, 2017a: Documentation for the U.S. Census Grids (Summary File 1), 2010. doi: 10.7927/H4MP517W.

Center For International Earth Science Information Network-CIESIN-Columbia University, 2017b: U.S. Census Grids (Summary File 1), 2010. doi:10.7927/H40Z716C.

Center For International Earth Science Information Network-CIESIN-Columbia University, 2021a: Documentation for the U.S. Social Vulnerability Index Grids. doi:10.7927/FJR9-A973.

Center For International Earth Science Information Network-CIESIN-Columbia University, 2021b: U.S. Social Vulnerability Index Grids. doi:10.7927/6S2A-9R49.

Chaney, P. L. and G. S. Weaver, 2010: The Vulnerability of Mobile Home Residents in Tornado Disasters: The 2008 Super Tuesday Tornado in Macon County, Tennessee. *Weather, Climate, and Society*, **2**, 190 – 199, doi:https://doi.org/10.1175/2010WCAS1042.1.

Chaney, P. L., G. S. Weaver, S. A. Youngblood, and K. Pitts, 2013: Household Preparedness for Tornado Hazards: The 2011 Disaster in DeKalb County, Alabama. *Weather, Climate, and Society*, **5**, 345 – 358, doi:https://doi.org/10.1175/WCAS-D-12-00046.1.

Chase, R. J., D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, 2022: A machine learning tutorial for operational meteorology. part i: Traditional machine learning. *Weather and Forecasting*, **37**, 1509 – 1529, doi:10.1175/WAF-D-22-0070.1.

Chase, R. J., D. R. Harrison, G. M. Lackmann, and A. McGovern, 2023: A machine learning tutorial for operational meteorology, part ii: Neural networks and deep learning. *Weather and Forecasting*, doi:10.1175/WAF-D-22-0187.1.

Chiu, C. H., A. H. Schnall, C. E. Mertzlufft, R. S. Noe, A. F. Wolkin, J. Spears, M. Casey-Lockyer, and S. J. Vagi, 2013: Mortality From a Tornado Outbreak, Alabama, April 27, 2011. *American Journal of Public Health*, **103**, e52–e58, doi:10.2105/AJPH.2013.301291.

Cintineo, J. L., M. J. Pavolonis, and J. M. Sieglaff, 2022: Probsevere lightningcast: A deep-learning model for satellite-based lightning nowcasting. *Weather and Forecasting*, **37**, 1239 – 1257, doi:https://doi.org/10.1175/WAF-D-22-0019.1.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, L. Cronce, and J. Brunner, 2020: NOAA ProbSevere v2.0—ProbHail, ProbWind, and ProbTor. *Weather and Forecasting*, **35**, 1523 – 1543, doi:10.1175/WAF-D-19-0242.1.

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. *Weather and Forecasting*, **27**, 1235 – 1248.

Coffer, B. E., M. D. Parker, R. L. Thompson, B. T. Smith, and R. E. Jewell, 2019: Using Near-Ground Storm Relative Helicity in Supercell Tornado Forecasting. *Weather and Forecasting*, **34**, 1417 – 1435.

Coleman, T. A. and P. G. Dixon, 2014: An Objective Analysis of Tornado Risk in the United States. *Weather and Forecasting*, **29**, 366 – 376.

Coniglio, M. C., S. F. Corfidi, and J. S. Kain, 2011: Environment and early evolution of the 8 may 2009 derecho-producing convective system. *Monthly Weather Review*, **139**, 1083 – 1102, doi:10.1175/2010MWR3413.1, URL `https://journals.ametsoc.org/view/journals/mwre/139/4/2010mwr3413.1.xml`.

Coniglio, M. C. and R. E. Jewell, 2022: SPC Mesoscale Analysis Compared to Field-Project Soundings: Implications for Supercell Environment Studies. *Monthly Weather Review*, **150**, 567 – 588.

Coniglio, M. C. and D. J. Stensrud, 2004: Interpreting the Climatology of Derechos. *Weather and Forecasting*, **19**, 595 – 605.

Corfidi, S. F., M. C. Coniglio, A. E. Cohen, and C. M. Mead, 2016: A Proposed Revision to the Definition of "Derecho". *Bulletin of the American Meteorological Society*, **97**, 935 – 949.

Crum, T. D. and R. L. Alberty, 1993: The WSR-88D and the WSR-88D Operational Support Facility. *Bulletin of the American Meteorological Society*, **74**, 1669 – 1688.

Cuomo, J. and V. Chandrasekar, 2021: Use of deep learning for weather radar nowcasting. *Journal of Atmospheric and Oceanic Technology*, **38**, 1641 – 1656, doi:10.1175/JTECH-D-21-0012.1.

Cutter, S. L., B. J. Boruff, and W. L. Shirley, 2003: Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, **84 (2)**, 242–261, URL `http://www.jstor.org/stable/42955868`.

Cutter, S. L. and C. Finch, 2008: Temporal and spatial changes in social vulnerability to natural hazards. *Proceedings of the National Academy of Sciences*, **105 (7)**, 2301–2306, doi:10.1073/pnas.0710375105.

DeWald, V. L. and T. W. Funk, 2002: WSR-88D reflectivity and velocity trends of a damaging squall line event on 20 April 1996 over south-central Indiana and central Kentucky. *20th Conf. on Severe Local Storms*, American Meteorological Society, 177–180, 6.3.

Doswell, C. A., H. E. Brooks, and N. Dotzek, 2009: On the implementation of the enhanced Fujita scale in the USA. *Atmospheric Research*, **93 (1)**, 554–563.

Doviak, R. J. and D. S. Zrnić, 1993: *Doppler Radar and Weather Observations*. 2d ed., Academic Press.

Ebert, E. E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Monthly Weather Review*, **129**, 2461 – 2480.

Edwards, R., J. G. LaDue, J. T. Ferree, K. Scharfenberg, C. Maier, and W. L. Coulbourne, 2013: Tornado Intensity Estimation: Past, Present, and Future. *Bulletin of the American Meteorological Society*, **94**, 641 – 653.

Ernst, S., J. Ripberger, M. J. Krocak, H. Jenkins-Smith, and C. Silva, 2021: Colorful Language: Investigating Public Interpretation of the Storm Prediction Center Convective Outlook. *Weather and Forecasting*, **36**, 1785 – 1797, doi:10.1175/WAF-D-21-0001.1.

Feng, Z., R. A. Houze, L. R. Leung, F. Song, J. C. Hardin, J. Wang, W. I. Gustafson, and C. R. Homeyer, 2019: Spatiotemporal characteristics and large-scale environments of mesoscale convective systems east of the rocky mountains. *Journal of Climate*, **32**, 7303 – 7328, doi:10.1175/JCLI-D-19-0137.1, URL `https://journals.ametsoc.org/view/journals/clim/32/21/jcli-d-19-0137.1.xml`.

Feng, Z., L. R. Leung, R. A. Houze Jr., S. Hagos, J. Hardin, Q. Yang, B. Han, and J. Fan, 2018: Structure and Evolution of Mesoscale Convective Systems: Sensitivity to Cloud Microphysics in Convection-Permitting Simulations Over the United States. *Journal of Advances in Modeling Earth Systems*, **10 (7)**, 1470–1494, doi:10.1029/2018MS001305, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001305`, `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018MS001305`.

Flora, M. L., C. K. Potvin, P. S. Skinner, S. Handler, and A. McGovern, 2021: Using Machine Learning to Generate Storm-Scale Probabilistic Guidance of Severe Weather Hazards in the Warn-on-Forecast System. *Monthly Weather Review*, **149**, 1535 – 1557, doi:10.1175/MWR-D-20-0194.1.

Flournoy, M. D. and M. C. Coniglio, 2019: Origins of Vorticity in a Simulated Tornadic Mesovortex Observed during PECAN on 6 July 2015. *Monthly Weather Review*, **147**, 107 – 134, doi:10.1175/MWR-D-18-0221.1.

Fowle, M. A. and P. J. Roebber, 2003: Short-Range (0–48 h) Numerical Prediction of Convective Occurrence, Mode, and Location. *Weather and Forecasting*, **18**, 782 – 794.

Fujita, T. T. and H. R. Byers, 1977: Spearhead Echo and Downburst in the Crash of an Airliner. *Monthly Weather Review*, **105**, 129 – 146.

Fujita, T. T. and R. M. Wakimoto, 1982: Effects of Miso- and Mesoscale Obstructions on PAM Winds Obtained during Project NIMROD. *Journal of Applied Meteorology and Climatology*, **21**, 840 – 858.

Gallus, W. A., N. A. Snook, and E. V. Johnson, 2008: Spring and Summer Severe Weather Reports over the Midwest as a Function of Convective Mode: A Preliminary Study. *Weather and Forecasting*, **23**, 101 – 113.

Geerts, B., 1998: Mesoscale convective systems in the southeast united states during 1994–95: A survey. *Weather and Forecasting*, **13**, 860 – 869, doi:10.1175/1520-0434(1998)013⟨0860:MCSITS⟩2.0.CO;2.

Gensini, V. A. and H. E. Brooks, 2018: Spatial trends in United States tornado frequency. *npj Climate and Atmospheric Science*, **1 (38)**, 1 – 5.

Gensini, V. A., C. Converse, W. S. Ashley, and M. Taszarek, 2021: Machine Learning Classification of Significant Tornadoes and Hail in the United States Using ERA5 Proximity Soundings. *Weather and Forecasting*, **36**, 2143 – 2160.

Géron, A., 2019: *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Incorporated, URL `https://books.google.com/books?id=OCS1twEACAAJ`.

Gibbs, J. G., 2021: Evaluating precursor signals for QLCS tornado and higher impact straight-line wind events. *J. Operational Meteor.*, **9 (5)**, 62–75, doi:10.15191/nwajom.2021.0905.

Grassotti, C., R. N. Hoffman, E. R. Vivoni, and D. Entekhabi, 2003: Multiple-Timescale Intercomparison of Two Radar Products and Rain Gauge Observations over the Arkansas–Red River Basin. *Weather and Forecasting*, **18**, 1207 – 1229.

Greene, D. R. and R. A. Clark, 1972: Vertically Integrated Liquid Water—A New Analysis Tool. *Monthly Weather Review*, **100**, 548 – 552.

Gutierrez, R. E. and M. R. Kumjian, 2021: Environmental and Radar Characteristics of Gargantuan Hail–Producing Storms. *Monthly Weather Review*, **149**, 2523 – 2538, doi:10.1175/MWR-D-20-0298.1.

Hales Jr., J. E., 1988: Improving the watch/warning program through use of significant event data. Preprints, 15th Conf. on Severe Local Storms, Baltimore, MD, Amer. Meteor. Soc., 165–168.

Heinselman, P. L., et al., 2023: Warn-on-Forecast System: From Vision to Reality. *Weather and Forecasting*, doi:https://doi.org/10.1175/WAF-D-23-0147.1.

Hoekstra, S., K. Klockow, R. Riley, J. Brotzge, H. Brooks, and S. Erickson, 2011: A Preliminary Look at the Social Perspective of Warn-on-Forecast: Preferred Tornado Warning Lead Time and the General Public's Perceptions of Weather Risks. *Weather, Climate, and Society*, **3**, 128 – 140, doi:10.1175/2011WCAS1076.1.

Homeyer, C. R. and K. P. Bowman, 2022: Algorithm Description Document for Version 4.2 of the Three-Dimensional Gridded NEXRAD WSR-88D Radar (GridRad) Dataset.

Homeyer, C. R., J. D. McAuliffe, and K. M. Bedka, 2017: On the Development of Above-Anvil Cirrus Plumes in Extratropical Convection. *Journal of the Atmospheric Sciences*, **74**, 1617 – 1633.

Homeyer, C. R., E. M. Murillo, and M. R. Kumjian, 2023: Relationships Between 10 Years of Radar-Observed Supercell Characteristics and Hail Potential. *Monthly Weather Review*, in press, doi:10.1175/MWR-D-23-0019.1.

Homeyer, C. R., T. N. Sandmæl, C. K. Potvin, and A. M. Murphy, 2020: Distinguishing Characteristics of Tornadic and Nontornadic Supercell Storms from Composite Mean Analyses of Radar Observations. *Monthly Weather Review*, **148**, 5015 – 5040.

Houze Jr., R. A., 2004: Mesoscale convective systems. *Reviews of Geophysics*, **42 (4)**.

Hurlbut, M. M. and A. E. Cohen, 2014: Environments of Northeast U.S. Severe Thunderstorm Events from 1999 to 2009. *Weather and Forecasting*, **29**, 3 – 22.

Jauernic, S. T. and M. S. V. D. Broeke, 2017: Tornado Warning Response and Perceptions among Undergraduates in Nebraska. *Weather, Climate, and Society*, **9**, 125 – 139, doi: https://doi.org/10.1175/WCAS-D-16-0031.1.

Jergensen, G. E., A. McGovern, R. Lagerquist, and T. Smith, 2020: Classifying Convective Storms Using Machine Learning. *Weather and Forecasting*, **35**, 537 – 559.

Johns, R. H. and W. D. Hirt, 1987: Derechos: Widespread Convectively Induced Windstorms. *Weather and Forecasting*, **2**, 32 – 49.

Klimowski, B. A., M. J. Bunkers, M. R. Hjelmfelt, and J. N. Covert, 2003: Severe Convective Windstorms over the Northern High Plains of the United States. *Weather and Forecasting*, **18**, 502 – 519.

Kousky, C., 2013: Facts about FEMA Household Disaster Aid: Examining the 2008 Floods and Tornadoes in Missouri. *Weather, Climate, and Society*, **5**, 332 – 344, doi:https://doi.org/10.1175/WCAS-D-12-00059.1.

Krocak, M. J., J. N. Allan, J. T. Ripberger, C. L. Silva, and H. C. Jenkins-Smith, 2021: An Analysis of Tornado Warning Reception and Response across Time: Leveraging Respondents' Confidence and a Nocturnal Tornado Climatology. *Weather and Forecasting*, **36**, 1649 – 1660.

Krocak, M. J. and H. E. Brooks, 2018: Climatological Estimates of Hourly Tornado Probability for the United States. *Weather and Forecasting*, **33**, 59 – 69.

Kumjian, M. R., 2013: Principles and Applications of Dual-Polarization Weather Radar. Part II: Warm- and Cold-Season Applications. *Journal of Operational Meteorology*, **1 (20)**, 243 – 264.

Kumjian, M. R. and A. V. Ryzhkov, 2008: Polarimetric Signatures in Supercell Thunderstorms. *Journal of Applied Meteorology and Climatology*, **47**, 1940 – 1961.

Kurdzo, J. M., et al., 2017: Observations of Severe Local Storms and Tornadoes with the Atmospheric Imaging Radar. *Bulletin of the American Meteorological Society*, **98**, 915 – 935.

Lagerquist, R., A. McGovern, C. R. Homeyer, D. J. G. II, and T. Smith, 2020: Deep Learning on Three-Dimensional Multiscale Data for Next-Hour Tornado Prediction. *Monthly Weather Review*, **148**, 2837 – 2861.

Lemon, L. R. and C. A. Doswell, 1979: Severe Thunderstorm Evolution and Mesocyclone Structure as Related to Tornadogenesis. *Monthly Weather Review*, **107**, 1184 – 1197.

Loeffler, S. D. and M. R. Kumjian, 2018: Quantifying the Separation of Enhanced ZDR and KDP Regions in Nonsupercell Tornadic Storms. *Weather and Forecasting*, **33**, 1143 – 1157, doi:10.1175/WAF-D-18-0011.1.

Loeffler, S. D., M. R. Kumjian, M. Jurewicz, and M. M. French, 2020: Differentiating Between Tornadic and Nontornadic Supercells Using Polarimetric Radar Signatures of Hydrometeor Size Sorting. *Geophysical Research Letters*, **47 (12)**, e2020GL088 242.

Lyza, A. W., M. D. Flournoy, and E. N. Rasmussen, 2022: Observed Characteristics of the Tornadic Supercells of 27–28 April 2011 in the Southeast United States. *Monthly Weather Review*, **150**, 2883 – 2910, doi:10.1175/MWR-D-21-0274.1.

Markowski, P. M. and Y. P. Richardson, 2009: Tornadogenesis: Our current understanding, forecasting considerations, and questions to guide future research. *Atmospheric Research*, **93 (1)**, 3–10, doi:10.1016/j.atmosres.2008.09.015, 4th European Conference on Severe Storms.

Mason, L. R., K. N. Ellis, B. Winchester, and S. Schexnayder, 2018: Tornado Warnings at Night: Who Gets the Message? *Weather, Climate, and Society*, **10**, 561 – 568, doi:10.1175/WCAS-D-17-0114.1.

McCarthy, J., J. W. Wilson, and T. T. Fujita, 1982: The Joint Airport Weather Studies Project. *Bulletin of the American Meteorological Society*, **63**, 15 – 22.

McGovern, A. and A. J. Broccoli, 2022: Editorial. *Artificial Intelligence for the Earth Systems*, **1**, e220 014, doi:10.1175/AIES-D-22-0014.1.

McGovern, A., R. J. Chase, M. Flora, D. J. Gagne, R. Lagerquist, C. K. Potvin, N. Snook, and E. Loken, 2023: A Review of Machine Learning for Convective Weather. *Artificial Intelligence for the Earth Systems*, **2**, e220 077, doi:10.1175/AIES-D-22-0077.1.

McGovern, A., R. Lagerquist, D. J. Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, **100**, 2175 – 2199, doi:10.1175/BAMS-D-18-0195.1.

McGovern, A., et al., 2022: Nsf ai institute for research on trustworthy ai in weather, climate, and coastal oceanography (ai2es). *Bulletin of the American Meteorological Society*, **103**, E1658 – E1668, doi:10.1175/BAMS-D-21-0020.1.

Mecikalski, J. R., T. N. Sandmæl, E. M. Murillo, C. R. Homeyer, K. M. Bedka, J. M. Apke, and C. P. Jewett, 2021: A Random-Forest Model to Assess Predictor Importance and Nowcast Severe Storms Using High-Resolution Radar–GOES Satellite–Lightning Observations. *Monthly Weather Review*, **149**, 1725 – 1746, doi:10.1175/MWR-D-19-0274.1.

Miller, P. W. and T. L. Mote, 2017: Standardizing the Definition of a "Pulse" Thunderstorm. *Bulletin of the American Meteorological Society*, **98**, 905 – 913.

Mitchell, E. D. W., S. V. Vasiloff, G. J. Stumpf, A. Witt, M. D. Eilts, J. T. Johnson, and K. W. Thomas, 1998: The National Severe Storms Laboratory Tornado Detection Algorithm. *Weather and Forecasting*, **13**, 352 – 366, doi:10.1175/1520-0434(1998)013⟨0352:TNSSLT⟩2.0.CO;2.

Murillo, E. M. and C. R. Homeyer, 2019: Severe Hail Fall and Hailstorm Detection Using Remote Sensing Observations. *Journal of Applied Meteorology and Climatology*, **58**, 947 – 970.

Murillo, E. M., C. R. Homeyer, and J. T. Allen, 2021: A 23-Year Severe Hail Climatology Using GridRad MESH Observations. *Monthly Weather Review*, **149**, 945 – 958.

Murphy, A. M. and C. R. Homeyer, 2023: Comparison of Radar-Observed Tornadic and Nontornadic MCS Cells using Probability Matched Means. *Journal of Applied Meteorology and Climatology*, in press, doi:https://doi.org/10.1175/JAMC-D-23-0070.1.

Murphy, A. M., C. R. Homeyer, and K. Q. Allen, 2023: Development and Investigation of GridRad-Severe, a Multi-Year Severe Event Radar Dataset. *Monthly Weather Review*, conditionally accepted.

NCEI, 2023: U.S. Billion-Dollar Weather and Climate Disasters (2023). URL `https://www.ncdc.noaa.gov/billions/`, URL `https://www.ncdc.noaa.gov/billions/`.

NCEI/NOAA, 2022: NOAA's Storm Events Database. National Centers for Environmental Information, accessed December 2019 to January 2022.

NOAA/NWS/ROC, 1991: NOAA Next Generation Radar (NEXRAD) Level II Base Data. NOAA National Centers for Environmental Information, accessed December 2019 to January 2022.

Parker, M. D. and R. H. Johnson, 2000: Organizational Modes of Midlatitude Mesoscale Convective Systems. *Monthly Weather Review*, **128**, 3413 – 3436.

Paul, B. K. and M. Stimers, 2014: Spatial Analyses of the 2011 Joplin Tornado Mortality: Deaths by Interpolated Damage Zones and Location of Victims. *Weather, Climate, and Society*, **6**, 161 – 174, doi:https://doi.org/10.1175/WCAS-D-13-00022.1.

Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the high-resolution rapid refresh model's ability to predict mesoscale convective systems using object-based evaluation. *Weather and Forecasting*, **30**, 892 – 913, doi:10.1175/WAF-D-14-00118.1, URL https://journals.ametsoc.org/view/journals/wefo/30/4/waf-d-14-00118_1.xml.

Ravuri, S., et al., 2021: Skilful precipitation nowcasting using deep generative models of radar. *Nature*, **597 (7878)**, 672–677, doi:10.1038/s41586-021-03854-z.

Ripberger, J. T., M. J. Krocak, W. W. Wehde, J. N. Allan, C. Silva, and H. Jenkins-Smith, 2019: Measuring Tornado Warning Reception, Comprehension, and Response in the United States. *Weather, Climate, and Society*, **11**, 863 – 880, doi:10.1175/WCAS-D-19-0015.1.

Rockwood, A. A. and R. A. Maddox, 1988: Mesoscale and Synoptic Scale Interactions Leading to Intense Convection: The Case of 7 June 1982. *Weather and Forecasting*, **3**, 51 – 68, doi:10.1175/1520-0434(1988)003⟨0051:MASSIL⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/3/1/1520-0434_1988_003_0051_massil_2_0_co_2.xml.

Roebber, P. J., 2009: Visualizing Multiple Measures of Forecast Quality. *Weather and Forecasting*, **24**, 601 – 608, doi:https://doi.org/10.1175/2008WAF2222159.1.

Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A theory for strong, long-lived squall lines. *Journal of Atmospheric Sciences*, **45**, 463 – 485, doi:10.1175/1520-0469(1988)045⟨0463:ATFSLL⟩2.0.CO;2.

Ryzhkov, A., D. Burgess, D. Zrnic, T. Smith, and S. Giangrande, 2002: Polarimetric Analyses of a 3 May 1999 Tornado. *21st Conf. on Severe Local Storms*, American Meteorological Society, 14.2.

Ryzhkov, A. V., T. J. Schuur, D. W. Burgess, and D. S. Zrnic, 2005: Polarimetric Tornado Detection. *Journal of Applied Meteorology*, **44**, 557 – 570.

Sandmæl, T. N., 2017: An Evaluation of Radar- and Satellite-Data Based Products to Discriminate Between Tornadic and Non-Tornadic Storms. M.S. thesis, University of Oklahoma.

Sandmæl, T. N., et al., 2023: The Tornado Probability Algorithm: A Probabilistic Machine Learning Tornadic Circulation Detection Algorithm. *Weather and Forecasting*, **38**, 445 – 466, doi:10.1175/WAF-D-22-0123.1.

Schaefer, J. T., 1990: The Critical Success Index as an Indicator of Warning Skill. *Weather and Forecasting*, **5**, 570 – 575.

Schiesser, H. H., R. A. Houze, and H. Huntrieser, 1995: The Mesoscale Structure of Severe Precipitation Systems in Switzerland. *Monthly Weather Review*, **123**, 2070 – 2097, doi:10.1175/1520-0493(1995)123⟨2070:TMSOSP⟩2.0.CO;2.

School of Meteorology, University of Oklahoma, 2021: GridRad-Severe - Three-Dimensional Gridded NEXRAD WSR-88D Radar Data for Severe Events.

Schumacher, R. S. and K. L. Rasmussen, 2020: The formation, character and changing nature of mesoscale convective systems. *Nat. Rev. Earth Environ.*, **1**, 300 – 314.

Sherman-Morris, K., C. Vaughn, J. C. Senkbeil, and S. Wooten, 2022: The Influence of Demographic and Place Variables on Personalized Tornado Risk Area. *Weather, Climate, and Society*, **14**, 1261 – 1272, doi:https://doi.org/10.1175/WCAS-D-22-0073.1.

Silver, A. and J. Andrey, 2014: The Influence of Previous Disaster Experience and Sociodemographics on Protective Behaviors during Two Successive Tornado Events. *Weather, Climate, and Society*, **6**, 91 – 103, doi:https://doi.org/10.1175/WCAS-D-13-00026.1.

Simmons, K. M. and D. Sutter, 2008: Tornado Warnings, Lead Times, and Tornado Casualties: An Empirical Investigation. *Weather and Forecasting*, **23**, 246 – 258.

Skinner, P. S., et al., 2018: Object-Based Verification of a Prototype Warn-on-Forecast System. *Weather and Forecasting*, **33**, 1225 – 1250.

Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective Modes for Significant Severe Thunderstorms in the Contiguous United States. Part I: Storm Classification and Climatology. *Weather and Forecasting*, **27**, 1114 – 1135.

Snively, D. V. and W. A. Gallus, 2014: Prediction of Convective Morphology in Near-Cloud-Permitting WRF Model Simulations. *Weather and Forecasting*, **29**, 130 – 149.

Starzec, M., C. R. Homeyer, and G. L. Mullendore, 2017: Storm Labeling in Three Dimensions (SL3D): A Volumetric Radar Echo and Dual-Polarization Updraft Classification Algorithm. *Monthly Weather Review*, **145**, 1127 – 1145.

Steinkruger, D., P. Markowski, and G. Young, 2020: An Artificially Intelligent System for the Automated Issuance of Tornado Warnings in Simulated Convective Storms. *Weather and Forecasting*, **35**, 1939 – 1965, doi:10.1175/WAF-D-19-0249.1.

Stensrud, D. J., et al., 2009: Convective-Scale Warn-on-Forecast System: A Vision for 2020. *Bulletin of the American Meteorological Society*, **90**, 1487 – 1500.

Stensrud, D. J., et al., 2013: Progress and challenges with Warn-on-Forecast. *Atmospheric Research*, **123**, 2–16, doi:10.1016/j.atmosres.2012.04.004, 6th European Conference on Severe Storms 2011. Palma de Mallorca, Spain.

Strader, S. M. and W. S. Ashley, 2018: Finescale assessment of mobile home tornado vulnerability in the central and southeast united states. *Weather, Climate, and Society*, **10**, 797 – 812, doi:10.1175/WCAS-D-18-0060.1.

Taszarek, M., N. Pilguj, J. T. Allen, V. Gensini, H. E. Brooks, and P. Szuster, 2021: Comparison of Convective Parameters Derived from ERA5 and MERRA-2 with Rawinsonde Data over Europe and North America. *Journal of Climate*, **34 (8)**, 3211 – 3237, doi: 10.1175/JCLI-D-20-0484.1.

Theodore Fujita, T., 1990: Downbursts: meteorological features and wind field characteristics. *Journal of Wind Engineering and Industrial Aerodynamics*, **36**, 75–86.

Thielen, J. E. and W. A. Gallus, 2019: Influences of Horizontal Grid Spacing and Microphysics on WRF Forecasts of Convective Morphology Evolution for Nocturnal MCSs in Weakly Forced Environments. *Weather and Forecasting*, **34**, 1495 – 1517.

Thompson, R. L., B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective Modes for Significant Severe Thunderstorms in the Contiguous United States. Part II: Supercell and QLCS Tornado Environments. *Weather and Forecasting*, **27**, 1136 – 1154, doi:10.1175/WAF-D-11-00116.1.

Torres, S. M. and C. D. Curtis, 2007: Initial Implementation of Super-Resolution Data on the NEXRAD Network. 23rd Conf. on International Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, San Antonio, TX, American Meteorological Society, 5B.10.

Trapp, R. J., S. A. Tessendorf, E. S. Godfrey, and H. E. Brooks, 2005: Tornadoes from Squall Lines and Bow Echoes. Part I: Climatological Distribution. *Weather and Forecasting*, **20**, 23 – 34.

Trapp, R. J. and M. L. Weisman, 2003: Low-Level Mesovortices within Squall Lines and Bow Echoes. Part II: Their Genesis and Implications. *Monthly Weather Review*, **131**, 2804 – 2823, doi:10.1175/1520-0493(2003)131⟨2804:LMWSLA⟩2.0.CO;2.

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. *Weather and Forecasting*, **21**, 408 – 415, doi:10.1175/WAF925.1.

Trujillo-Falcón, J. E., O. Bermúdez, K. Negrón-Hernández, J. Lipski, E. Leitman, and K. Berry, 2021: Hazardous Weather Communication En Español: Challenges, Current Resources, and Future Practices. *Bulletin of the American Meteorological Society*, **102**, E765 – E773, doi:10.1175/BAMS-D-20-0249.1.

Uccellini, L. W. and J. E. T. Hoeve, 2019: Evolving the National Weather Service to Build a Weather-Ready Nation: Connecting Observations, Forecasts, and Warnings to Decision-Makers through Impact-Based Decision Support Services. *Bulletin of the American Meteorological Society*, **100**, 1923 – 1942, doi:10.1175/BAMS-D-18-0159.1.

Van Den Broeke, M. S., 2020: A Preliminary Polarimetric Radar Comparison of Pretornadic and Nontornadic Supercell Storms. *Mon. Wea. Rev.*, **148**, 1567–1584, doi: 10.1175/MWR-D-19-0296.1.

Vaughan, M. T., B. H. Tang, and L. F. Bosart, 2017: Climatology and Analysis of High-Impact, Low Predictive Skill Severe Weather Events in the Northeast United States. *Weather and Forecasting*, **32**, 1903 – 1919.

Wakimoto, R. M., 2001: *Severe Convective Storms*, chap. Convectively Driven High Wind Events, 255 – 298. Meteorological Monographs, American Meteorological Society, Boston, MA.

Weisman, M. L. and C. A. Davis, 1998: Mechanisms for the Generation of Mesoscale Vortices within Quasi-Linear Convective Systems. *Journal of the Atmospheric Sciences*, **55**, 2603 – 2622, doi:10.1175/1520-0469(1998)055⟨2603:MFTGOM⟩2.0.CO;2.

Weisman, M. L. and R. J. Trapp, 2003: Low-level mesovortices within squall lines and bow echoes. part i: Overview and dependence on environmental shear. *Monthly Weather Review*, **131**, 2779 – 2803, doi:10.1175/1520-0493(2003)131⟨2779:LMWSLA⟩2.0.CO;2.

Wendt, N. A. and I. L. Jirak, 2021: An Hourly Climatology of Operational MRMS MESH-Diagnosed Severe and Significant Hail with Comparisons to Storm Data Hail Reports. *Weather and Forecasting*, **36**, 645 – 659.

Wilson, J. W., J. A. Moore, G. B. Foote, B. Martner, A. R. Rodi, T. Uttal, and J. M. Wilczak, 1988: Convection Initiation and Downburst Experiment (CINDE). *Bulletin of the American Meteorological Society*, **69**, 1328 – 1347.

Wilson, J. W. and R. M. Wakimoto, 2001: The Discovery of the Downburst: T. T. Fujita's Contribution. *Bulletin of the American Meteorological Society*, **82**, 49 – 62.

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. *Weather and Forecasting*, **13**, 286 – 303.

Wurman, J., J. M. Straka, and E. N. Rasmussen, 1996: Fine-Scale Doppler Radar Observations of Tornadoes. *Science*, **272 (5269)**, 1774–1777.