

## An Iterative Storm Segmentation and Classification Algorithm for Convection-Allowing Models and Gridded Radar Analyses

COREY K. POTVIN,<sup>a,b</sup> BURKELY T. GALLO,<sup>c,d</sup> ANTHONY E. REINHART,<sup>a</sup> BRETT ROBERTS,<sup>c,a,d</sup>  
 PATRICK S. SKINNER,<sup>c,a,b</sup> RYAN A. SOBASH,<sup>e</sup> KATIE A. WILSON,<sup>c,a</sup> KELSEY C. BRITT,<sup>b,c,a</sup> CHRIS BROYLES,<sup>d</sup>  
 MONTGOMERY L. FLORA,<sup>c,a,b</sup> WILLIAM J. S. MILLER,<sup>c,a</sup> AND CLARICE N. SATRIO<sup>c,a</sup>

<sup>a</sup> NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

<sup>b</sup> School of Meteorology, University of Oklahoma, Norman, Oklahoma

<sup>c</sup> Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma

<sup>d</sup> NOAA/NWS/NCEP/Storm Prediction Center, Norman, Oklahoma

<sup>e</sup> National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 27 September 2021, in final form 3 March 2022)

**ABSTRACT:** Thunderstorm mode strongly impacts the likelihood and predictability of tornadoes and other hazards, and thus is of great interest to severe weather forecasters and researchers. It is often impossible for a forecaster to manually classify all the storms within convection-allowing model (CAM) output during a severe weather outbreak, or for a scientist to manually classify all storms in a large CAM or radar dataset in a timely manner. Automated storm classification techniques facilitate these tasks and provide objective inputs to operational tools, including machine learning models for predicting thunderstorm hazards. Accurate storm classification, however, requires accurate storm segmentation. Many storm segmentation techniques fail to distinguish between clustered storms, thereby missing intense cells, or to identify cells embedded within quasi-linear convective systems that can produce tornadoes and damaging winds. Therefore, we have developed an iterative technique that identifies these constituent storms in addition to traditionally identified storms. Identified storms are classified according to a seven-mode scheme designed for severe weather operations and research. The classification model is a hand-developed decision tree that operates on storm properties computed from composite reflectivity and midlevel rotation fields. These properties include geometrical attributes, whether the storm contains smaller storms or resides within a larger-scale complex, and whether strong rotation exists near the storm centroid. We evaluate the classification algorithm using expert labels of 400 storms simulated by the NSSL Warn-on-Forecast System or analyzed by the NSSL Multi-Radar/Multi-Sensor product suite. The classification algorithm emulates expert opinion reasonably well (e.g., 76% accuracy for supercells), and therefore could facilitate a wide range of operational and research applications.

**SIGNIFICANCE STATEMENT:** We have developed a new technique for automatically identifying intense thunderstorms in model and radar data and classifying storm mode, which informs forecasters about the risks of tornadoes and other high-impact weather. The technique identifies storms that are often missed by other methods, including cells embedded within storm clusters, and successfully classifies important storm modes that are generally not included in other schemes, such as rotating cells embedded within quasi-linear convective systems. We hope the technique will facilitate a variety of forecasting and research efforts.

**KEYWORDS:** Classification; Convective storms; Decision trees

### 1. Introduction

#### *a. Motivation for storm classification algorithms*

Storm mode features prominently in many thunderstorm research and forecast applications. The primary processes driving storm behavior can often be confidently inferred from

---

Flora's current affiliation: Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma.

Miller's current affiliation: Earth Systems Science Interdisciplinary Center, and NOAA/National Environmental Satellite Data and Information Service/Center for Satellite Applications and Research, University of Maryland, College Park, College Park, Maryland.

---

*Corresponding author:* Corey K. Potvin, corey.potvin@noaa.gov

the storm morphology manifested in radar and satellite imagery, or visually. For example, the propagation and strength of a quasi-linear convective system (QLCS) is typically governed largely by a systemwide cold pool (e.g., Weisman and Rotunno 2004), and supercells are largely governed by their mesocyclone (e.g., Lemon and Doswell 1979). These governing processes, in turn, strongly modulate storm predictability, with modes characterized by greater organization (e.g., QLCS, supercells) tending toward higher predictability than disorganized modes (e.g., ordinary cells). Storm mode also correlates with the predictability of particular hazards, including severe hail, severe wind, and tornadoes (e.g., Lakshmanan and Smith 2009; Brotzge et al. 2013), and to the relative likelihood of these hazards (e.g., Andra et al. 2002; Gallus et al. 2008; Duda and Gallus 2010; Smith et al. 2012, 2013). For example, a QLCS is less likely than a supercell to produce a significant [i.e., (E)F2+] tornado (e.g., Trapp et al. 2005), and the timing

DOI: 10.1175/JTECH-D-21-0141.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

and location of tornadogenesis is generally more uncertain in a QLCS than in a supercell. Given the strong relationships between storm mode and essential aspects of storm behavior, mode is one of the primary storm characteristics that forecasters evaluate when predicting storm evolution and severe weather potential, researchers consider when compiling cases for study, and field campaign planners consider when designing storm observing strategies.

Given the relevance of storm mode to so many aspects of convective meteorology, storm classification techniques are a critical tool for both forecast operations and research. Real-time classification of storm mode in convection-allowing model (CAM) output can provide valuable guidance to forecasters, especially when the output is too large to be thoroughly examined in real time. Moreover, storm classifications could be a valuable input to machine learning models that provide forecast guidance on storm hazards, longevity, and other attributes tied to storm mode. When instances of a particular storm mode are sought in large sets of idealized simulations or real-world forecasts, automated methods enable the identification of many more cases than could be identified manually (Lakshmanan and Smith 2009). Therefore, classification algorithms facilitate studies of various important topics, including how storm mode distributions vary with mesoscale environment, how particularly hazardous modes (e.g., supercells) vary with climate, and the accuracy with which CAMs predict different storm modes. Finally, objective techniques are necessary for reproducible research.

#### *b. Storm segmentation and classification algorithms*

Before storms can be classified, they must first be detected and characterized. While it is useful to distinguish between these two tasks—storm segmentation and classification—the two are (ideally) linked. The storm segmentation algorithm should be designed with the storm classification criteria in mind. For example, to classify QLCSs by the nature of their stratiform region (e.g., Parker and Johnson 2000), the storm segmentation algorithm should accurately identify QLCS stratiform regions, and not solely QLCS convective lines. In the technique presented herein, the output of the storm classification algorithm informs the quality control within the storm segmentation algorithm, and so the two tasks are especially coupled.

Designing an objective storm segmentation technique that performs well over a range of convective scenarios is challenging. Two of the greatest obstacles are the tendency for storms to occur in close proximity to one another (e.g., in the case of a recently split supercell), and for convective systems to contain convective cells that the developer wishes to identify. While it is typically easy to visually distinguish between individual storms in such cases, it is difficult to design an objective procedure that reliably does so because there is no single reflectivity (or reflectivity gradient) threshold that demarcates the boundaries of all storms. Using a lower reflectivity threshold allows more storms to be correctly identified, but also increases false identifications and inflates the areas of intense storms. Conversely, using a higher threshold reduces spurious detections and produces smaller storm objects, but causes more storms to be missed.

Several methods have been developed to address this and other limitations of fixed-threshold storm segmentation techniques. The Storm Cell Identification and Tracking algorithm (Johnson et al. 1998) uses multiple reflectivity thresholds to identify more intense cells embedded within convective systems. The enhanced Thunderstorm Identification, Tracking, and Nowcasting algorithm (Dixon and Wiener 1993; Han et al. 2009) uses multiple reflectivity thresholds as well as erosion (an image processing operation) to separate falsely merged objects. Lakshmanan and Smith (2009) introduced a multiscale clustering technique that enables segmentation of storm objects at different scales, including individual storm cells embedded within mesoscale convective systems. Lack et al. (2010) developed a multiscale version of the Procrustes object-oriented verification scheme (Micheas et al. 2007) that achieves the same objective. Lakshmanan et al. (2009) developed the enhanced watershed method for storm segmentation and noted the possibility of implementing a hierarchical version that saves storm objects detected with successive saliency criteria and hysteresis levels. Flora et al. (2021) used a two-pass version of the enhanced watershed method in their creation of ensemble storm track objects.

There is no universal set of storm mode classification criteria. How best to categorize and define storm modes depends upon the application (e.g., distinguishing severe from nonsevere storms; subcategorizing QLCS by stratiform characteristics), input data (e.g., radar or rainfall observations; model output), and whether a subjective or objective approach is used. Thus, a variety of storm classification criteria have been developed (Table 1). Classification criteria can be qualitative (e.g., long line of storms) or quantitative (e.g., line of storms > 100 km long), and often reference properties derived from radar reflectivity data (e.g., contiguous region of composite reflectivity > 40 dBZ exceeds 100 km in length). Process-based criteria are also sometimes used; for example, QLCS classification could require strong evidence of the existence of a systemwide cold pool. While such criteria are useful in conceptual definitions of storm modes or in a modeling framework, some are not generally applicable to real storms due to observing network limitations.

Storm classification criteria, including those adopted herein, have typically been tuned through trial and error. Supervised machine learning approaches automate this process by training the classification model on sets of storms with associated mode labels (e.g., Gagne et al. 2009; Lack and Fox 2012; Jergensen et al. 2020). We briefly discuss the tradeoffs between the manual and machine learning approaches to storm classification model development in section 5.

Storm segmentation algorithms often compute various morphological properties of the detected storms (e.g., aspect ratio, orientation, eccentricity) that can be used by storm classification schemes. We note, however, that storm segmentation has many applications besides storm classification. Initial storm detection algorithms were primarily oriented toward real-time tracking, characterization, and nowcasting of storms in radar data (e.g., Dixon and Wiener 1993; Johnson et al. 1998). A more recent application of storm segmentation is the calculation of storm

TABLE 1. Sampling of previous storm classification schemes.

Study	Application	Data inputs	Primary storm modes
<a href="#">Bluestein and Jain (1985)</a>	Characterization of QLCS modes, environments	WSR-57 reflectivity scans (photographs viewed with time-lapse projector)	Broken line, back building, broken areal, embedded areal
<a href="#">Steiner et al. (1995)</a>	Distinguish convective vs stratiform precipitation	Doppler C-band reflectivity scans	Precipitating convection, precipitating stratiform
<a href="#">Parker and Johnson (2000)</a>	Characterization of QLCS modes, environments	Base radar reflectivity	Trailing stratiform, leading stratiform, parallel stratiform
<a href="#">Baldwin et al. (2005)</a>	Automated classification of rainfall systems	NCEP Stage IV hourly precipitation estimates	Convective linear, convective cellular, stratiform
<a href="#">Trapp et al. (2005)</a>	Tornadic mode climatology	Composite reflectivity (CREF)	Discrete cell, QLCS, other
<a href="#">Gagne et al. (2009)</a>	Machine learning (ML) classification of storm mode	Simulated and observed reflectivity (4 km AGL)	Cells: isolated cells, isolated strong cells, multicells; linear systems: parallel, leading, trailing stratiform
<a href="#">Duda and Gallus (2010)</a>	Severe weather mode climatology	CREF; Mesocyclone Detection Algorithm ( <a href="#">Stumpf et al. 1998</a> )	Isolated cells, cell clusters, broken lines; squall lines with parallel, leading, trailing, or no stratiform region; bow echo; nonlinear convection; supercell
<a href="#">Smith et al. (2012)</a>	Tornadic and significant severe mode climatology	Volumetric reflectivity and radial velocity; mesocyclone nomograms ( <a href="#">Andra 1997</a> )	QLCS: line, bow echo; supercell: discrete, in cluster, in line; disorganized: discrete, in line, cluster; linear hybrid; marginal supercell
<a href="#">Lack and Fox (2012)</a>	ML classification of storm mode	CREF; Rapid Update Cycle (RUC; <a href="#">Benjamin et al. 2002</a> ) environmental fields	Airmass storm, pulse storm (severe potential); small-scale rotating storm—with or without severe potential; linear convective system—with or without rotation; supercell
<a href="#">Starzec et al. (2017)</a>	Improve upon traditional convective–stratiform classification	Volumetric reflectivity and dual-polarimetric variables	Convection, convective updraft, precipitating stratiform, nonprecipitating stratiform, ice-only anvil
<a href="#">Jergensen et al. (2020)</a>	ML classification of storm mode	Multiyear Reanalysis of Remotely Sensed Storms ( <a href="#">Ortega et al. 2012</a> ); RUC proximity soundings	Supercell, QLCS, disorganized

properties for input to machine learned severe weather prediction models (e.g., [Gagne et al. 2017](#); [Lagerquist et al. 2017](#); [Cintineo et al. 2018](#); [Burke et al. 2020](#); [Flora et al. 2021](#)). Storm segmentation methods (in combination with object matching techniques) have also enabled powerful new ways to verify and compare CAM representation and prediction of storms and near-storm

environments (e.g., [Skinner et al. 2018](#); [Potvin et al. 2019, 2020](#); [Johnson et al. 2020](#); [Lawson et al. 2021](#); [Miller et al. 2022](#)).

### c. Overview of this study

Inspired by previous multiscale storm segmentation methods, we have designed an iterative technique to identify storm

cells embedded within larger-scale systems, and to better delineate between clustered storms (section 3). The segmentation technique uses the simpler, fixed-threshold (rather than the enhanced watershed) approach (the motivation for this choice is explained in section 3b), and operates solely on the composite reflectivity (hereafter CREF) field, which is routinely available for both model and observed storms. The storm mode classifier operates on attributes of the reflectivity objects and on cyclonic rotation objects generated from either the 2–5 km AGL updraft helicity (hereafter UH) field for CAM storms, or the 2–5 km AGL azimuthal shear product in the NSSL Multi-Radar/Multi-Sensor (MRMS; Smith et al. 2016) suite for observed storms. The classification model is a hand-developed decision tree. The storm modes in our classification scheme were selected to support a breadth of operational and research activities involving intense convection, but especially efforts within the NOAA Warn-on-Forecast (WoF; Stensrud et al. 2009, 2013) program. We developed and tested our technique using the WoF System (WoFS; Wheatley et al. 2015; Jones et al. 2016; Lawson et al. 2018) ensemble forecasts and MRMS data generated during the 2017–20 NOAA Hazardous Weather Testbed Spring Forecasting Experiments (SFEs; Gallo et al. 2017; Clark et al. 2020). Both the WoFS and MRMS products used for testing have 3-km grid spacing.

To evaluate the classification algorithm, we developed an online survey to collect expert storm mode labels for use as “truth” (section 4). Each survey question presents a storm randomly selected from the WoFS or MRMS output and asks the respondent to select among the seven storm modes in our classification scheme. Preliminary versions of the survey were completed by several of the coauthors, whose responses and subsequent feedback were used to refine the storm mode classes and definitions, the storm segmentation and classification algorithms, and the survey design. The final surveys were completed by 30 experts, the majority of which are NWS forecasters, who collectively labeled 200 WoFS and 200 MRMS storms. Comparing the expert- and algorithm-assigned storm mode labels reveals that the classification algorithm is reasonably accurate, and works well for both MRMS and WoFS storms. Thus, the classification algorithm could be used to facilitate a range of endeavors leveraging CAM and gridded radar datasets.

## 2. WoFS and MRMS datasets

The WoFS is an experimental convection-allowing (3-km grid spacing) ensemble designed to provide thunderstorm hazard guidance at short (e.g., 0–6-h) lead times. The WoFS uses the Advanced Research version of the Weather and Research Forecast Model (WRF-ARW; Skamarock et al. 2008). Radar and satellite data are assimilated every 15 min and conventional observations every hour using an ensemble Kalman filter. The WoFS uses 36 ensemble members for data assimilation, and the first 18 of those members to generate 3- or 6-h lead time forecasts after each data assimilation cycle. During the SFEs, the WoFS domain is daily centered in collaboration with the NOAA Storm Prediction Center on a region forecast to experience severe weather. The size of the WoFS domain was 750 km × 750 km during the 2017 SFE and 900 km × 900 km in

subsequent SFEs. The present study uses 1-h WoFS forecasts initialized hourly from 1900 to 0300 UTC during the 2017–20 SFEs. The CREF field output by each WoFS member is the column maximum of the reflectivity calculated by the NSSL two-moment microphysics scheme (Mansell et al. 2010), which includes both graupel and hail categories. Additional WoFS details are found in Miller et al. (2022) and references therein.

The MRMS system includes severe weather, transportation, and precipitation products that leverage forecast model data and observations from the WSR-88D radar network and satellite, lightning, and conventional platforms. The MRMS CREF product used herein is the column maximum of the exponential inverse-distance-weighted average of the reflectivity from each contributing radar (Smith et al. 2016). The 2–5 km AGL layer maximum cyclonic azimuthal shear (hereafter AzShr) is computed as in Mahalik et al. (2019). Both products were generated at NSSL using the version 12 MRMS codebase. The MRMS CREF and AzShr fields are interpolated from their native 0.5- and 1-km grids, respectively, to the 3-km WoFS grid.

## 3. Storm identification and classification algorithm

### a. Overview of the classification scheme

We designed our set of storm mode classes with a variety of operational and research applications in mind. These applications are primarily oriented toward severe thunderstorm hazards (tornadoes, large hail, and damaging wind), and so we chose to identify and classify convective, and not stratiform, precipitation regions, and to distinguish between organized and disorganized convection (Table 2). We seek to identify QLCS convective lines (QLCS), supercells (SUPERCELL), supercell clusters (SUP\_CLUSTER), and intense but nonmesocyclonic cells (ORDINARY). Since QLCS severe weather potential is often maximized within line-embedded cells, we also identify those cells and classify them as mesocyclonic or nonmesocyclonic (QLCS\_MESO and QLCS\_ORD, respectively). Finally, we define a catch-all class, OTHER, that encompasses all storms that do not exhibit one of the previous six modes.

Our classification criteria involve only radar-derived fields and their model counterparts, namely, CREF and either 2–5 km AGL UH (WoFS) or 2–5 km AGL AzShr (MRMS). We could have included environmental criteria as have some machine learned classification models (e.g., Gagne et al. 2009; Lack and Fox 2012; Jergensen et al. 2020), but chose instead to limit the required inputs for the algorithm to maximize usability. In addition, given the large overlap in environments for different storm modes, we were concerned about unduly biasing the classification algorithm toward the optimal environments for particular modes (especially since one of the intended applications of the algorithm is to analyze relationships between storm mode and environment). As we will show in section 4, the classification algorithm performs quite well with just the CREF and midlevel rotation fields.

TABLE 2. Storm modes in the 7-class scheme.

Class	Description
ORDINARY	Single storm cell lacking sustained intense rotation; not embedded in QLCS convective line (see QLCS_MESO); may be embedded in QLCS stratiform region or be one of a cluster of storms
SUPERCELL	Single storm cell containing sustained intense rotation; not embedded in QLCS convective line (see QLCS_MESO); may be embedded in QLCS stratiform region or be one of a cluster of storms
QLCS	Convective line of a quasi-linear convective system that exhibits systemwide organization
SUP_CLUST	Multiple storms in close proximity, of which <i>two or more</i> are supercells; not a QLCS
QLCS_ORD	QLCS-line-embedded cell lacking sustained intense rotation
QLCS_MESO	QLCS-line-embedded cell containing sustained intense rotation (either a supercell, or a mesovortex-bearing cell)
OTHER	Cluster of ordinary cells (or one ordinary cell and one supercell), <i>or</i> a convective segment lacking systemwide organization, <i>or</i> noncanonical storm mode (e.g., amorphous blob of convection not associated with a single strong updraft)

### b. Motivation for iterative storm segmentation procedure

The simplest way to detect storms in 2D gridded reflectivity data is to identify grid cells exceeding a prescribed intensity threshold (e.g., 40 dBZ), then identify connected regions of those grid cells, then identify those regions exceeding a prescribed area threshold (e.g., 100 km<sup>2</sup>). Prior to imposing the area threshold, objects whose minimum boundary separation falls below a prescribed distance limit (e.g., 10 km) can be merged to ensure identification of convective systems whose reflectivity falls below the intensity threshold between convective cores. Additional criteria can be imposed; for example, a minimum threshold on the maximum reflectivity within the object (e.g., 45 dBZ).

The watershed transform (Roerdink and Meijster 2001) is a popular image processing technique that improves upon the above “fixed-threshold” approach by simultaneously applying a range of intensity thresholds. The traditional watershed transform, however, does not work well for meteorological image segmentation (Lakshmanan et al. 2003). The enhanced watershed transform introduced by Lakshmanan et al. (2009) redefines saliency—the criterion used in watershed methods to retain potential objects, or “basins”—in terms of basin area in order to provide direct control over the minimum size of identified objects. While the fixed-threshold approach could simply be iterated for different intensity thresholds, the enhanced watershed transform is far more efficient for big-data applications.

While the enhanced watershed transform is a particularly sophisticated approach to storm segmentation, the fixed-threshold approach has also proven effective for identifying potentially high-impact discrete cells and convective systems in MRMS and CAM CREF data (Skinner et al. 2018). The success of the simpler approach is due to the fact that a single reflectivity intensity threshold identifies such storms in most meteorological scenarios. Both approaches, however, fail to identify cells embedded within convective systems or that lie in close proximity to one another since each grid cell can be assigned to at most a single object. Addressing this limitation requires a hierarchical approach that allows smaller objects to be identified within larger ones, and that retains both sets of objects.

Both the fixed-threshold or enhanced watershed methods can be made hierarchical in a conceptually straightforward way. In either case, the original procedure must be iterated using different object segmentation thresholds and saving the objects from every iteration. Since multiple objects corresponding to the same storm can be identified in different iterations, criteria are required to identify probable duplicates and select a single object for retention. While the enhanced watershed transform would likely be needed for applications seeking to identify a broad spectrum of convective storms (i.e., including weak storms that are unlikely to imminently produce severe weather), we are exclusively interested in identifying intense storms; therefore, we have developed an iterative version of the simpler, fixed-threshold, approach.

### c. Initial storm objects and rotation objects

The first step of our object segmentation algorithm (Fig. 1) approximately follows the method of Skinner et al. (2018). For both WoFS and MRMS output, we binarize the CREF field on a prescribed threshold: 43 and 39 dBZ, respectively. These thresholds and all the other storm segmentation parameters listed herein were empirically tuned using numerous cases from the 2017–20 SFES. Upon binarizing the CREF field, the Python Scikit-image module (Van der Walt et al. 2014) is used to identify and label storm objects. Then storm objects with area <135 km<sup>2</sup> are discarded. Next, objects whose boundaries lie within 9 km of one another are merged (i.e., assigned the same label). Finally, storm objects whose gridpoint-maximum intensity does not exceed a threshold—45 and 41 dBZ for WoFS and MRMS objects, respectively—are discarded. These thresholds approximate the 99th percentiles computed from the full WoFS and MRMS datasets. Using both a segmentation and maximum-intensity threshold (e.g., 43 and 45 dBZ for WoFS) rather than a single, higher segmentation threshold (e.g., 45 dBZ for WoFS) allows us to discard large storms that nevertheless lack an intense updraft.

After the initial storm object segmentation, the objects are classified using their morphological attributes (Fig. 2a), which are computed by Scikit-image. Most of these classifications are subject to change in the final classification phase (Figs. 2b,c), which occurs after the final iteration of the segmentation–classification

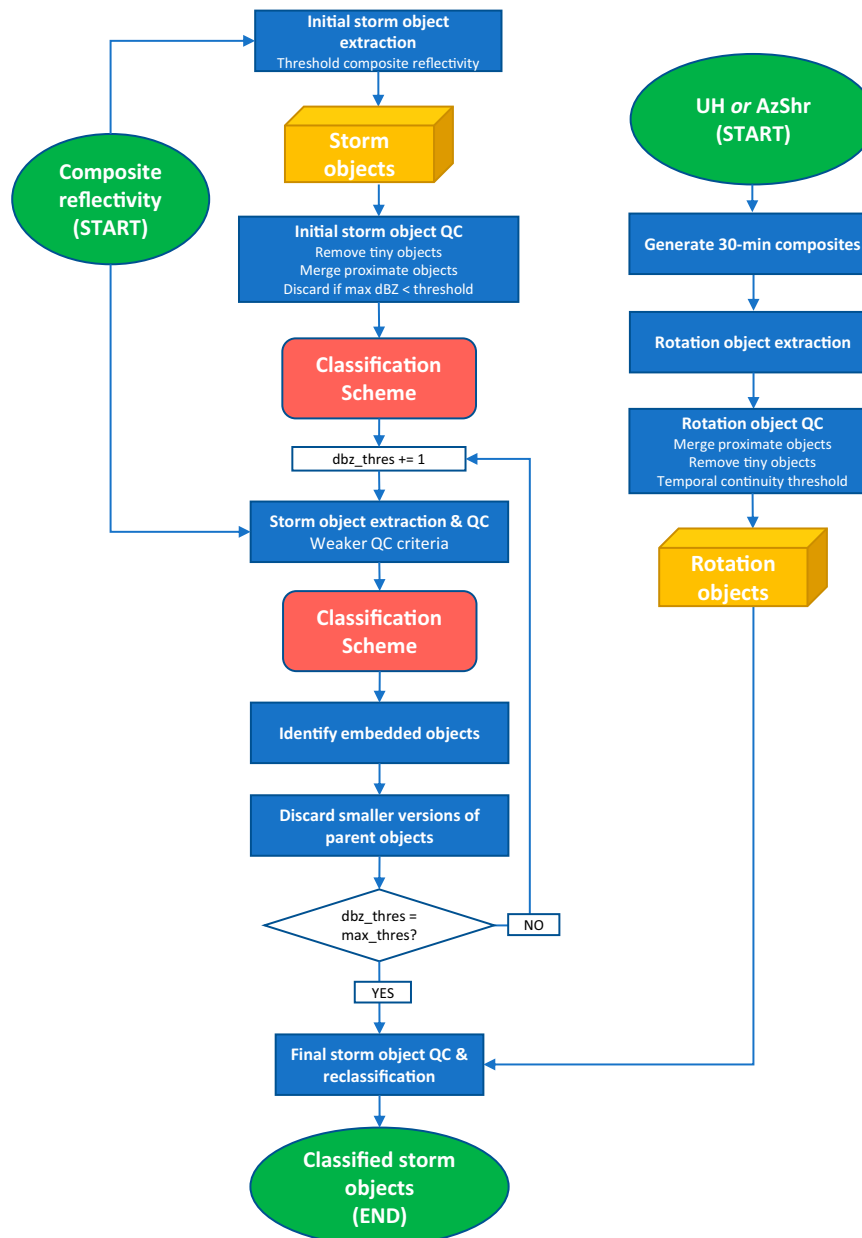


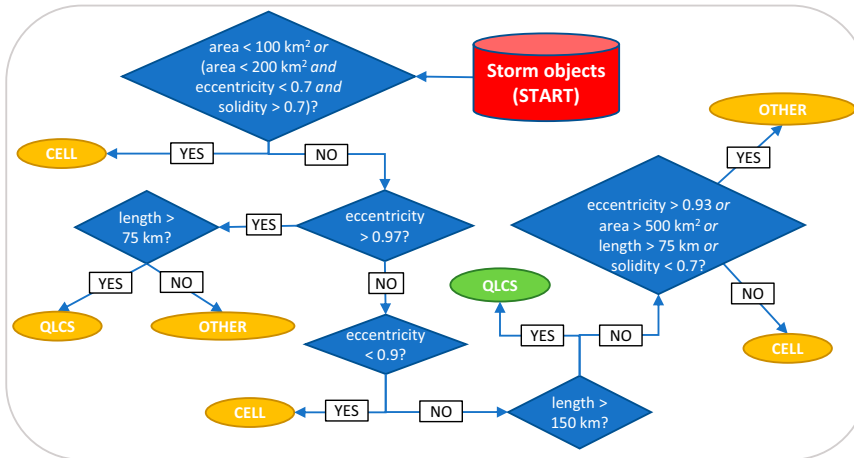
FIG. 1. Flowchart for storm object segmentation algorithm.

procedure (section 3e). Single-cell storm objects identified during the preliminary classification and final multicellular classifications (Figs. 2a,b), hereafter “CELLS,” are ultimately classified as SUPERCELL, ORDINARY, QLCS\_MESO, or QLCS\_ORD (Fig. 2c).

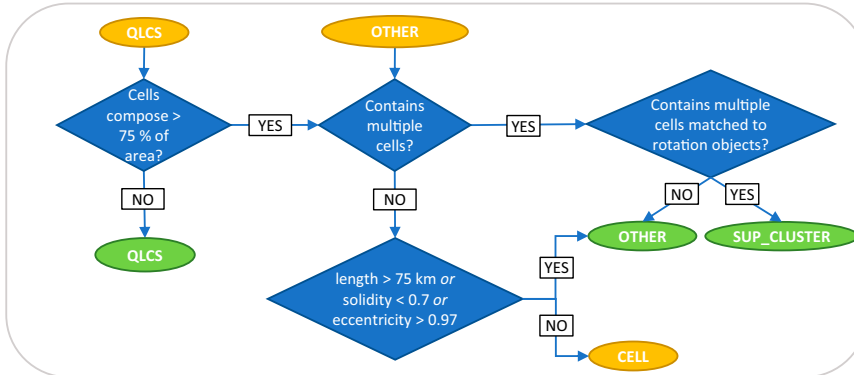
To later distinguish between storms with and without strong sustained rotation (i.e., SUPERCELL versus ORDINARY, and QLCS\_MESO versus QLCS\_ORD), 30-min composite rotation objects are created from the instantaneous UH and AzShr fields valid every 5 min (Fig. 1). Exceedance thresholds of  $50 \text{ m}^2 \text{ s}^{-2}$  and  $0.004 \text{ s}^{-1}$  are used; these approximate the

99.95th percentiles of each respective field. Rotation objects whose boundaries lay within 3 km of each other are merged. Objects are discarded if they are smaller than  $45 \text{ km}^2$  or were generated using rotation from fewer than 3 times (since such objects likely do not represent sustained rotation). To determine whether a reflectivity object contains strong sustained rotation, rotation objects are sought within a square domain centered on the reflectivity object centroid. The diameter of the search domain is set to the greater of 12 km and one-third of the equivalent diameter of the reflectivity object.

**a) Preliminary storm classifications**



**b) Final multicellular classifications**



**c) Final cell classifications**

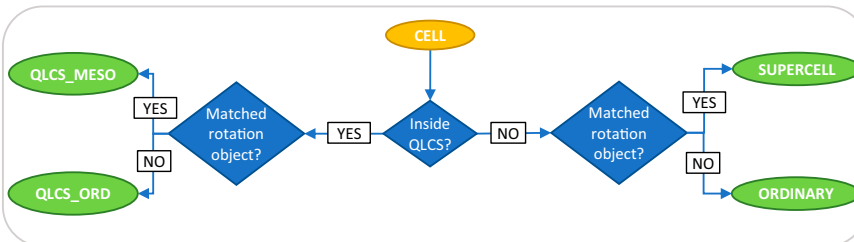


FIG. 2. Flowcharts for storm classification algorithm: (a) Preliminary classification (prior to final iteration), (b) final multicellular classification, and (c) final cellular classification. Green and yellow shading signifies preliminary and final classifications, respectively.

*d. Constituent storm objects*

To identify any CELLS contained within the objects identified in the first step, we iterate the object segmentation and classification, with the CREF threshold (initially 43 and 39 dBZ for WoFS and MRMS objects, respectively) incremented by 1 dBZ for each of 20 iterations (Fig. 1). Candidate new objects are identified at each iteration using a minimum area threshold of 10 grid cells (90 km<sup>2</sup>), with no subsequent object merging. These settings make allowance for the typically small size of, and narrow separation between, CELLS that are clustered together or embedded within QLCSs. New objects are retained if

they exist within a previously identified (parent) object. The new objects are preliminarily classified in the same manner as the initial objects (Fig. 2a). Objects not classified as CELLS are discarded since we aim to identify only the individual cells constituting parent objects (e.g., we do not wish to identify subclusters of cells within larger-scale clusters).

Since the CREF threshold is increased with each iteration, many of the candidate objects will merely be smaller versions of CELL objects identified in previous iterations. We found that in such cases, the centroids of the parent CELL and candidate objects are typically nearly collocated. Thus, at the end of each iteration, we discard objects contained within previously

identified CELLS if the distance between their two centroids is less than 15% of the equivalent diameter of the parent CELL or less than 6 km.

#### *e. Final quality control and classification*

Following the final iteration of the storm segmentation and classification algorithm, we perform a series of quality control (QC) and reclassification steps that account for constituent CELLS and their parent objects (“Final storm object QC & reclassification” in Fig. 1). To facilitate this procedure, each CELL’s parent object (if one exists) is identified, and then the “generation” of each object is determined. Unembedded storm objects are assigned to generation 0, objects within those objects to generation 1, and so forth. In the first step of the final QC and reclassification procedure, CELLS contained within QLCS-embedded CELLS are discarded. Then each QLCS with length  $< 150$  km and whose constituent CELLS compose  $> 75\%$  of the QLCS area is reclassified as a SUP\_CLUSTER if it contains multiple rotating CELLS or OTHER if it does not (Fig. 2b). Next, CELLS embedded within OTHERs are examined, beginning with the “deepest” objects (typically generation 2 or 1) and progressing to unembedded objects (generation 0). If a singleton CELL is embedded within an OTHER or CELL, it is discarded; if the parent object is an OTHER, it is reclassified as a CELL. If multiple objects are embedded within an unembedded CELL, the CELL is reclassified as an OTHER or SUP\_CLUSTER. If multiple objects are embedded within an *embedded* OTHER or CELL, the latter is discarded since we do not wish to classify multicellular objects within larger-scale multicellular objects. Upon treating all instances of OTHERs with embedded CELLS, unembedded CELLS with length  $> 75$  km, solidity<sup>1</sup>  $< 0.7$ , or eccentricity  $> 0.97$  are reclassified as OTHER. Finally, CELLS are assigned their final classifications (ORDINARY, SUPERCELL, QLCS\_ORD, or QLCS\_MESO) based on whether they are embedded within a QLCS and whether they are associated with rotation objects (Fig. 2c). Identifying and classifying all MRMS or WoFS (for a single member) storms within a  $900 \text{ km} \times 900 \text{ km}$  region at a single time average around 1 min on a single Intel Xeon 6140 CPU (2.30-GHz base frequency).

The benefit of the iterative approach is illustrated in Fig. 3. The initial storm segmentation procedure, which is essentially the fixed-threshold method, correctly identifies all of the intense convection in the image, but is incapable of identifying the individual cells constituting the storm complexes (Fig. 3a). Upon iterating through all the reflectivity thresholds and performing the final QC and classification, numerous QLCS-embedded cells have been identified and classified, as have two ORDINARY cells within a cluster near the northeastern corner of the domain (Fig. 3b).

#### **4. Survey-based evaluation of classification algorithm**

The storm classification algorithm was evaluated using expert labels of 400 storms collected by surveys created using the

<sup>1</sup> The ratio of the storm object area to the area of the convex hull (i.e., smallest convex polygon enclosing the storm object).

Qualtrics software. Before describing the details of the final surveys (section 4b) and evaluations (section 4c), we discuss the motivation for key aspects of our survey design.

#### *a. Survey design considerations*

A major limitation of our classification algorithm is that it operates on a single snapshot of the CREF field, and on sparse temporal information about rotation. This limitation increases misclassification of storms that temporarily exhibit features of different modes. Human experts, on the other hand, routinely examine storm evolution while assessing storm mode. Therefore, to obtain expert labels that more fairly evaluate the algorithm, we include animations of CREF and rotation in the surveys.

However, even when provided with detailed information about storm evolution, two experts can reasonably classify the same storm differently. To accommodate this frequent ambiguity in storm mode, we had multiple experts complete each survey, then accounted for the degree of consensus in the expert labels when evaluating the classification algorithm (section 4c). We could alternatively have assigned each storm to a single respondent and asked them to quantify their uncertainty in their classification. The primary advantage of that approach would have been to multiply the number of expert-labeled storms. However, the advantages of our approach include reducing the workload for each respondent, and generating an ensemble of expert labels of each storm, thereby enabling analysis of the degree of subjective agreement on storm mode. Sobash et al. (2021) combined both approaches in their classification algorithm evaluation: multiple experts classified each storm and quantified their uncertainty in their classification.

Several preliminary versions of the classification survey were completed by a group of convective storm experts (all of them coauthors). These expert labels and subsequent discussions were used to iteratively refine the classification scheme, including the choices of storm modes and their definitions, as well as the survey design. Limiting the number of storm mode options and clarifying the definitions of each were critical to achieving expert consensus in a reasonably large proportion of cases. In the preliminary survey, the degree of expert agreement with each other and with the algorithm classifications was found to substantially increase as the respondents labeled more storms, indicating a nontrivial learning curve for survey respondents. To mitigate this ordering effect on experts’ subjective classifications in the final surveys, the question order was randomized for each respondent, thus reducing the odds that a given storm was labeled by all or most of the respondents while they were still in their “warm-up” period.

#### *b. Final survey design*

The storms for the final surveys were automatically selected from the set of 2017 SFE objects, which were minimally used in the algorithm development. Four surveys were created, each with a unique set of 100 storm objects for labeling, for a total of 400 storms. We chose to deploy four surveys with 100 storms rather than one survey with 400 storms to reduce



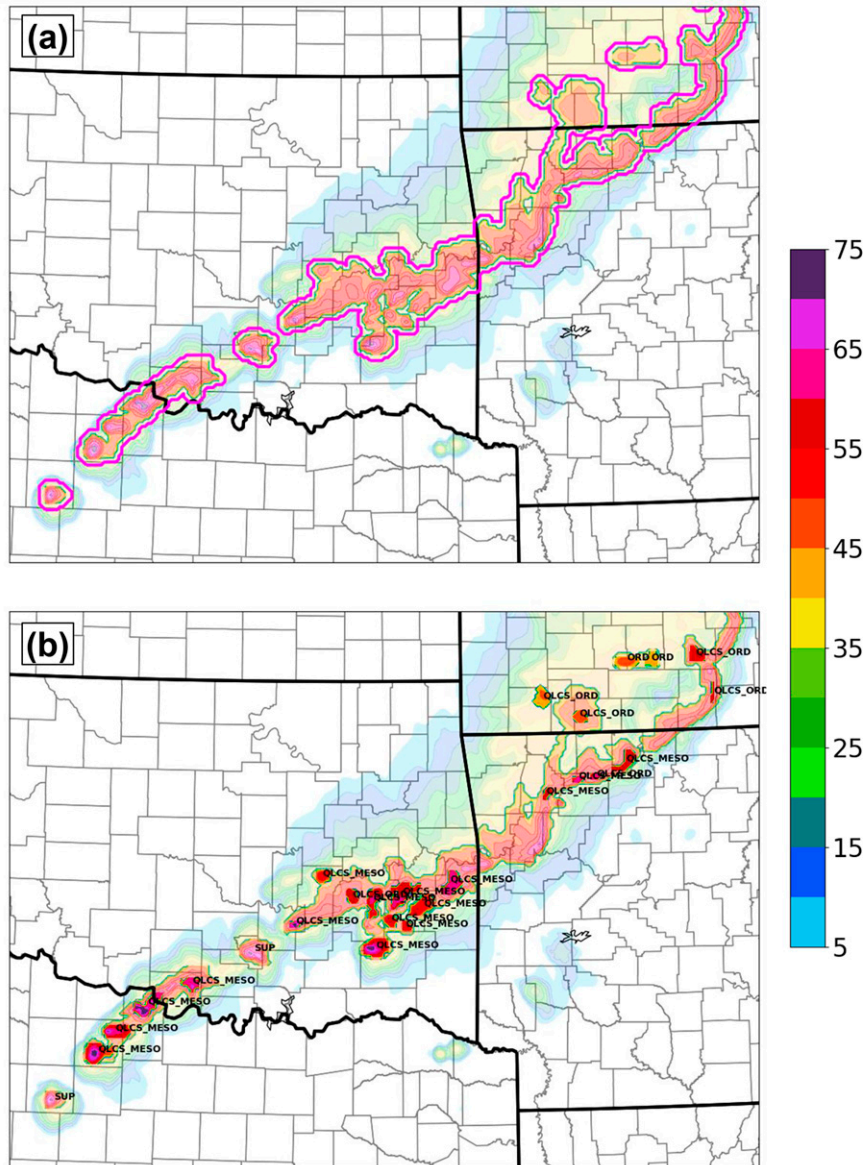


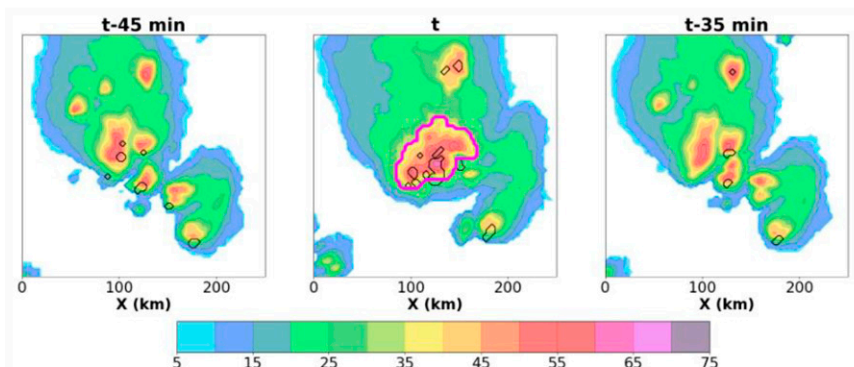
FIG. 3. Demonstration of the storm segmentation and classification algorithm. CREF (dBZ) is shaded in both panels. (a) Storm objects identified in the first step of the algorithm, indicated by darker shading and magenta contours. (b) The final output of the algorithm, with storm objects indicated by darker shading, and embedded objects indicated by the darkest shading and their mode labeled (SUP = SUPERCELL, ORD = ORDINARY).

the workload on each survey participant. To help ensure a sufficiently large evaluation sample for each storm mode, a quasi-arbitrary number of each algorithm-assigned mode was randomly selected from both the WoFS and MRMS datasets for each survey: 9 ORDINARY, 9 SUPERCELL, 7 QLCS, 7 SUP\_CLUST, 5 QLCS\_ORD, 5 QLCS\_MESO, and 8 OTHER. For each storm, respondents were presented with a plot of the CREF field and the 30-min composite rotation objects (section 3c) valid over the 250-km domain centered on the storm of interest, which was prominently indicated (Fig. 4). Respondents were not informed whether

each storm came from the WoFS or the MRMS dataset. To allow the respondents to incorporate information about storm evolution into their label selection, animations of CREF and instantaneous rotation exceeding a threshold (section 3c) were shown on both sides of the static plot, each valid from 60 min before to 30 min after the storm classification time. The left (right) animation showed images valid every 15 (5) min. The respondents were prompted to select the storm mode from a multiple-choice menu containing detailed definitions of each mode. Survey instructions, including examples of each storm mode, were provided both

## STORM MODE CLASSIFICATION

(Storm 2 of 100)



How would you label the indicated storm? [Survey instructions](#)

**SUPERCELL:** single storm cell containing sustained intense rotation; not embedded in QLCS leading line (see QLCS\_MESO); may be embedded in QLCS stratiform region or be one of a cluster of storms.

**ORDINARY:** single storm cell \*not\* containing sustained intense rotation; not embedded in QLCS leading line (see QLCS\_ORD); may be embedded in QLCS stratiform region or be one of a cluster of storms.

**SUP\_CLUSTER:** multiple storms in close proximity, of which two or more are supercells; not a QLCS

**QLCS:** leading convective line of a quasi-linear convective system that exhibits system-wide organization (NOT a segment of a QLCS)

**QLCS\_MESO:** QLCS-line-embedded cell containing sustained intense rotation (either a supercell, or a mesovortex-bearing cell)

**QLCS\_ORD:** QLCS-line-embedded cell \*not\* containing sustained intense rotation

**OTHER:** exhibits a storm mode that is not listed above, such as a cluster of ordinary cells (or one ordinary cell and one supercell), or a convective segment lacking system-wide organization, or no canonical storm mode (e.g., an amorphous blob of convection that does not appear to be associated with a single strong updraft)

FIG. 4. Sample question from the final surveys. (top) CREF (shading) and intense rotation (black outlines), with (center) the storm to be classified outlined in magenta. The top-left and top-right panels are animated in the surveys to show storm evolution shortly before and after classification time.

at the beginning of the survey and as a linked document within each question.

The survey was advertised via email to communities with expertise in convective meteorology. A survey link was sent to

interested individuals who were not involved with the algorithm development, have expertise in convective meteorology, and identified as an NWS forecaster, research scientist, or graduate student. The surveys were anonymous; the only

TABLE 3. Number of participants completing each survey by professional status.

	Forecasters	Researchers	Students	Total
Survey 1	3	2	1	6
Survey 2	5	2	1	8
Survey 3	5	3	0	8
Survey 4	4	3	0	7
Totals	17	10	2	29

participant information collected was professional status (forecaster, scientist, or student). A total of 29 individuals completed a survey: 17 forecasters, 10 researchers, and 2 students (Table 3). The majority of individuals completing each survey were forecasters. All participants signed an electronic informed consent form prior to completing this survey. This study was approved under the University of Oklahoma’s Institutional Review Board (Project 13364).

c. Evaluating the storm classification algorithm

As noted previously, storm mode is often ambiguous, making verification of classification schemes challenging. Our solution was to discard storms for which there was no respondent consensus, initially defined herein as a simple majority (four experts agreeing in surveys 1 and 4, and five experts agreeing in surveys 2 and 3; Table 3). We could have adopted a stronger consensus threshold to improve the accuracy of the storm mode labels used for verification, but this may also have produced an overly optimistic view of the algorithm classification accuracy by biasing the verification toward storms exhibiting more obvious modes. Using the simple-majority threshold, consensus was achieved on 265 of the 400 storms (66%), yielding a robust sample for verification (Table 4). The algorithm classifications matched the consensus for 70% of the verification storms. Averaged across all respondents, individual classifications matched the consensus for 79% of the verification storms. Thus, the algorithm deviated from the expert consensus only moderately more often than did individual experts. Given the relatively small samples of storms of each mode, and the uncertainty in the “true” mode in cases where consensus was barely obtained among the respondents, these verification statistics should not be interpreted too literally. Nevertheless, our results strongly suggest that the classification algorithm is reasonably consistent with expert judgements of storm mode. The rate of respondent consensus and algorithm classification accuracy were similar for WoFS and MRMS storms (72% versus 61%, and 73% versus 66%, respectively). The similar performance of the classification algorithm for WoFS and MRMS storms suggests that it would perform similarly well for

other CAM and gridded radar datasets without excessive tuning of the CREF-based criteria.

Restricting the verification to storms for which a majority of the respondents selected the same mode excludes many storms whose mode is more ambiguous (135 out of 400). Not only does this shrink the verification sample, but excluding these ambiguous storms potentially produces an overly optimistic view of the classification accuracy. To mitigate these potential sampling biases, we expanded the verification to include all storms assigned the same mode by at least three respondents. This “weak consensus” approach yielded 389 storms for evaluation, versus 265 from the original approach. In those cases where one mode received a plurality of respondent classifications, the “true” mode was set to the plurality mode. When two modes were selected by the same number of respondents and one of the modes matched the algorithm classification, the “true” mode was set to the latter; this tiebreaker was invoked for 26 of the 124 new storms (21%). The inclusion of these more ambiguous storms reduced the classification accuracy of the algorithm from 70% to 64%, but the rate of agreement of individual respondents with the consensus decreased by a similar margin: from 79% to 72% (Table 4). With 97% of the surveyed 400 storms included in the expanded evaluation, we conclude that the original evaluation well represents the consistency of the classification algorithm with expert judgements of storm mode.

Since many applications do not require as many storm mode distinctions as our scheme makes, we additionally evaluated the algorithm using two different groupings of our seven storm modes, each composed of four or three classes (Table 5). For each of these schemes, we assigned the respondent and algorithm classifications to their respective classes, then evaluated the new labels as we did the original mode labels. Naturally, the classification accuracy of the algorithm increased as the number of classes decreased (Table 4). The 76% and 88% classification accuracy for the four- and three-category schemes, respectively, strongly recommend the algorithm for high-level storm mode classification. Since the expert consensus for the 3- and 4-class schemes were determined by remapping the original (7-class) expert labels, however, these verification results should be interpreted with caution.

Contingency-table-based verification statistics were computed for each storm mode for each of the classification schemes (Table 6). The three chosen statistics—true skill statistic (TSS; Hanssen and Kuipers 1965), Gilbert skill score (GSS; Gilbert 1884), and Heidke skill score (HSS; Wilks 2011)—are well suited to rare events forecasting since they account for correct forecasts due to random chance. Confusion matrices (Wilks 2011) were also generated for each of the classification schemes (Figs. 5 and 6). Row- and

TABLE 4. Rate of expert consensus, average individual agreement with consensus, and algorithm agreement with consensus.

Classification scheme	Respondent consensus	Respondent agreement with consensus	Algorithm accuracy
7-class	66%	79%	70%
7-class, weak consensus	97%	72%	64%
4-class	—	—	76%
3-class	—	—	88%

TABLE 5. Storm modes in 4-class and 3-class schemes.

4-class scheme	
ORG_MULTICELL	QLCS, SUP_CLUST
DISCRETE	ORDINARY, SUPERCELL
QLCS_EMBED	QLCS_ORD, QLCS_MESO
OTHER	OTHER
3-class scheme	
CELLULAR	ORDINARY, SUPERCELL, QLCS_ORD, QLCS_MESO
ORG_MULTICELL	QLCS, SUP_CLUST
OTHER	OTHER

column-normalized confusion matrices were similar for the original and weak-consensus evaluations of the 7-class scheme (not shown). Examination of the verification statistics and confusion matrices reveals that the algorithm generally better classifies organized than disorganized storm modes. More specifically, in the 7-class scheme, QLCS, SUPERCELL, and SUP\_CLUST are better classified than OTHER and QLCS\_ORD, and in the 3- and 4-class schemes, ORG\_MULTICELL are better classified than OTHER. The confusion matrices are particularly useful in highlighting common pitfalls of the classification algorithm. For example, for the 7-class scheme, ORDINARY storms are often misclassified as QLCS\_ORD or OTHER, OTHER as QLCS, and QLCS\_ORD as ORDINARY or OTHER (Fig. 5). Inspection of individual cases reveals typical reasons for these recurring misclassifications. For example, the frequent misclassification of ORDINARY as QLCS\_ORD (Fig. 5), and of DISCRETE as QLCS\_EMBED (Fig. 6), arises from the frequent proximity of discrete cells to QLCS (not shown). Given the large differences in classification accuracy between modes, the overall accuracy of the classification algorithm will be sensitive to the storm mode distribution of a particular dataset.

## 5. Conclusions and future work

We have presented an iterative storm segmentation and classification algorithm for use with CAM and gridded radar data. The algorithm requires only composite reflectivity and rotation (midlevel updraft helicity or MRMS azimuthal shear) fields as input. The algorithm identifies a wide range of intense storms, most notably, cells embedded within storm clusters or QLCS. Storms are assigned one of seven modes that were selected with severe weather operations and research applications in mind. Based on evaluations with multiple expert labels of 400 storms, the classification algorithm agrees reasonably well with expert judgements of storm mode. Given the small number of inputs required for the technique, and its similarly good performance for both WoFS and MRMS data, we expect it could be valuably applied to other CAM and gridded radar datasets without excessive modifications.

The evaluation of our classification scheme could be extended in at least three ways. First, increasing the number of labeled storms would enable new analyses, including rigorous comparisons of the classification accuracy for WoFS versus MRMS storms. Second, it would be valuable to assess how well the algorithm classifies storms during rapid mode transition

(e.g., upscale growth), during which storms can exhibit traits of both the start and end modes. Third, given the fundamental differences between simulated updraft helicity and radar-derived azimuthal shear, it would be helpful to systematically assess the intercomparability of our WoFS and MRMS rotation objects, perhaps using observing system simulation experiments.

Mode	TSS	GSS	HSS	ACC
7-class scheme				
QLCS	0.77	0.54	0.71	0.82
SUPERCELL	0.72	0.59	0.74	0.76
SUP_CLUST	0.68	0.54	0.70	0.72
QLCS_MESO	0.65	0.52	0.68	0.67
ORDINARY	0.64	0.49	0.66	0.70
OTHER	0.51	0.33	0.49	0.60
QLCS_ORD	0.51	0.30	0.46	0.57
4-class scheme				
ORG_MULTICELL	0.82	0.68	0.81	0.88
DISCRETE	0.68	0.53	0.69	0.78
QLCS_EMBED	0.59	0.40	0.57	0.68
OTHER	0.51	0.32	0.49	0.60
3-class scheme				
CELLULAR	0.89	0.78	0.87	0.93
ORG_MULTICELL	0.83	0.69	0.82	0.88
OTHER	0.52	0.33	0.50	0.60

(e.g., upscale growth), during which storms can exhibit traits of both the start and end modes. Third, given the fundamental differences between simulated updraft helicity and radar-derived azimuthal shear, it would be helpful to systematically assess the intercomparability of our WoFS and MRMS rotation objects, perhaps using observing system simulation experiments.

It might be possible to improve the classification accuracy by incorporating additional fields that can be calculated from both CAM and gridded radar data, such as low-level rotation, reflectivity at various altitudes (e.g., Starzec et al. 2017), and echo-top height. Storm attributes from a series of times could also be used to capture aspects of storm evolution that experts find valuable for classifying mode. A machine learning (ML) approach (e.g., Gagne et al. 2009; Lack and Fox 2012; Jergensen et al. 2020; Sobash et al. 2021) would greatly facilitate any substantial expansion of the feature set, since manual development of complex decision trees is time-consuming. It may not be trivial to develop a ML model that outperforms the present technique, however. While expanding the feature set to include additional types of information should promote higher classification accuracy, the number of expert-labeled storms required for training, hyperparameter tuning, and testing may well exceed the 400 samples we have already collected. This highlights an important trade-off between ML and manual approaches to certain supervised learning problems (i.e., training a machine learning model on labeled datasets): the ML approach can reduce the time spent on model development, but also requires a much larger total human time investment (to gather the requisite labels) to outperform traditional methods. On the other hand, ML models can provide uncertainty estimates for their predictions, which are particularly useful in classification problems like ours where the

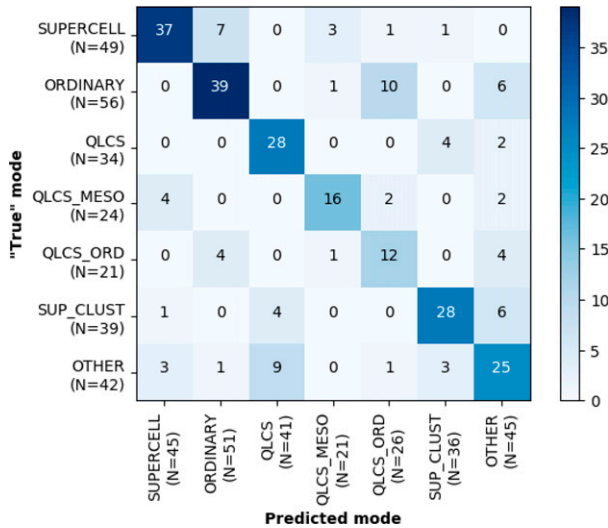


FIG. 5. Confusion matrix for expert (“true”) vs algorithm storm mode classifications. Sample sizes are listed beneath each storm mode label.

targets often do not fit neatly into a single class. Given that many storms do not exhibit a well-defined mode, as evidenced by the lack of expert agreement on many of the storms in the classification surveys, the deterministic nature of our algorithm’s classifications limits its value to certain applications.

We are considering a variety of potential WoFS applications of the new algorithm. One is to provide real-time forecast guidance on the evolution of storms and attendant hazards: for example, the number of ensemble members that are predicting a storm cluster to grow upscale into a QLCS, or for a nonrotating storm to become a supercell. Another potential application is to incorporate storm mode as a new feature in existing WoFS-based ML models for predicting storm hazards (Flora et al. 2021; Chmielewski et al. 2021), and in future ML models (in development) for estimating uncertainty of WoFS forecasts of individual storms. Stratifying the verification of large sets of WoFS forecasts by storm mode would provide valuable information for real-time interpretation, and about the relative practical predictability of different modes. Finally, objective storm classifications would facilitate intermode comparisons of near-storm environments, CAM performance, and the risks of thunderstorm hazards.

*Acknowledgments.* This work was prepared by the authors with funding from the NOAA/National Severe Storms Laboratory (CKP, AER), NOAA/Storm Prediction Center (CB), NOAA HWT Grant NA19OAR4590128 (RAS), and the NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce (PSS, MLF, BTG, BR, KAW, WJSM, KCB, CS). We thank Derek Stratman for informally reviewing a preliminary version of this paper, and three anonymous reviewers who provided valuable suggestions for improvement. Many of the analyses and visualizations were produced using the freely provided Anaconda Python distribution. The

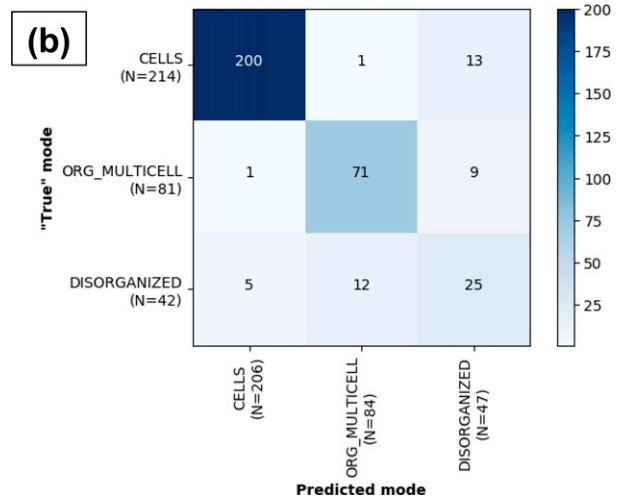
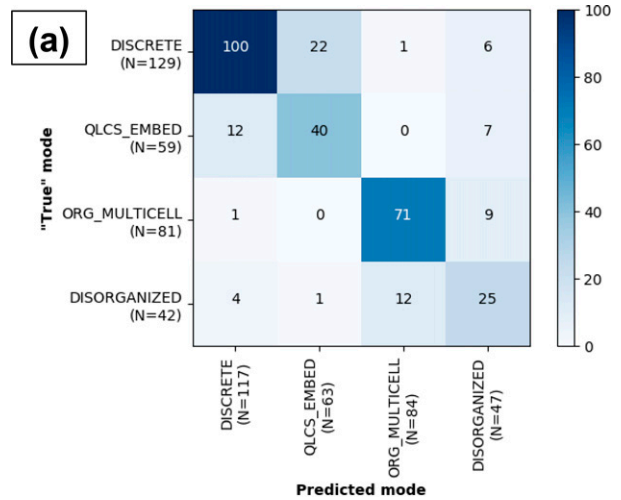


FIG. 6. As in Fig. 5, but for (a) the 4-class scheme and (b) the 3-class scheme.

contents of this paper do not necessarily reflect the views or official position of any organization of the United States.

*Data availability statement.* The experimental WoFS ensemble forecasts and MRMS data used in this study are not currently available in a publicly accessible repository. The data and code used to generate the results herein are available from the authors upon request.

REFERENCES

Andra, D. L., Jr., 1997: The origin and evolution of the WSR-88D mesocyclone recognition nomogram. Preprints, *28th Conf. on Radar Meteorology*, Austin, TX, Amer. Meteor. Soc., 364–365.

—, E. M. Quetone, and W. F. Bunting, 2002: Warning decision making: The relative roles of conceptual models, technology, strategy, and forecaster expertise on 3 May 1999. *Wea.*

- Forecasting*, **17**, 559–566, [https://doi.org/10.1175/1520-0434\(2002\)017<0559:WDMTRR>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0559:WDMTRR>2.0.CO;2).
- Baldwin, M. E., J. S. Kain, and S. Lakshminarayanan, 2005: Development of an automated classification procedure for rainfall systems. *Mon. Wea. Rev.*, **133**, 844–862, <https://doi.org/10.1175/MWR2892.1>.
- Benjamin, S. G., and Coauthors, 2002: RUC20—The 20-km version of the Rapid Update Cycle. NWS Tech. Procedures Bull. 490, 30 pp.
- Bluestein, H. B., and M. H. Jain, 1985: Formation of mesoscale lines of precipitation: Severe squall lines in Oklahoma during the spring. *J. Atmos. Sci.*, **42**, 1711–1732, [https://doi.org/10.1175/1520-0469\(1985\)042<1711:FOMLOP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<1711:FOMLOP>2.0.CO;2).
- Brotzge, J., S. Nelson, R. Thompson, and B. Smith, 2013: Tornado probability of detection and lead time as a function of convective mode and environmental parameters. *Wea. Forecasting*, **28**, 1261–1276, <https://doi.org/10.1175/WAF-D-12-00119.1>.
- Burke, A., N. Snook, D. J. Gagne, S. McCorkle, and A. McGovern, 2020: Calibration of machine learning–based probabilistic hail predictions for operational forecasting. *Wea. Forecasting*, **35**, 149–168, <https://doi.org/10.1175/WAF-D-19-0105.1>.
- Chmielewski, V., C. Potvin, P. S. Skinner, A. E. Reinhart, E. R. Mansell, and K. M. Calhoun, 2021: How well can we forecast cloud-to-ground lightning rates within the NSSL Experimental Warn-on-Forecast system using machine learning? *10th Conf. on the Meteorological Application of Lightning Data*, Online, Amer. Meteor. Soc., 5.4, <https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/380582>.
- Cinteo, J. L., and Coauthors, 2018: The NOAA/CIMSS ProbSevere model: Incorporation of total lightning and validation. *Wea. Forecasting*, **33**, 331–345, <https://doi.org/10.1175/WAF-D-17-0099.1>.
- Clark, A. J., and Coauthors, 2020: A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather. *Bull. Amer. Meteor. Soc.*, **101**, E2022–E2024, <https://doi.org/10.1175/BAMS-D-19-0298.1>.
- Dixon, M., and G. Wiener, 1993: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A radar-based methodology. *J. Atmos. Oceanic Technol.*, **10**, 785–797, [https://doi.org/10.1175/1520-0426\(1993\)010<0785:TTITAA>2.0.CO;2](https://doi.org/10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2).
- Duda, J. D., and W. A. Gallus, 2010: Spring and summer Midwestern severe weather reports in supercells compared to other morphologies. *Wea. Forecasting*, **25**, 190–206, <https://doi.org/10.1175/2009WAF2222338.1>.
- 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. *Mon. Wea. Rev.*, **149**, 1535–1557, <https://doi.org/10.1175/MWR-D-20-0194.1>.
- Gallo, B. T., and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed spring forecasting experiment. *Wea. Forecasting*, **32**, 1541–1568, <https://doi.org/10.1175/WAF-D-16-0178.1>.
- 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. *Wea. Forecasting*, **23**, 101–113, <https://doi.org/10.1175/2007WAF2006120.1>.
- Gagne, D. J., A. McGovern, and J. Brotzge, 2009: Classification of convective areas using decision trees. *J. Atmos. Oceanic Technol.*, **26**, 1341–1353, <https://doi.org/10.1175/2008JTECHA1205.1>.
- , —, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, <https://doi.org/10.1175/WAF-D-17-0010.1>.
- Gilbert, G. F., 1884: Finley's tornado predictions. *Amer. Meteor. J.*, **1**, 166–172.
- Han, L., S. Fu, L. Zhao, Y. Zheng, H. Wang, and Y. Lin, 2009: 3D convective storm identification, tracking, and forecasting—An enhanced TITAN algorithm. *J. Atmos. Oceanic Technol.*, **26**, 719–732, <https://doi.org/10.1175/2008JTECHA1084.1>.
- Hanssen, A. W., and W. J. A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Meded. Verh.*, **81**, 2–15.
- Jergensen, G. E., A. McGovern, R. Lagerquist, and T. Smith, 2020: Classifying convective storms using machine learning. *Wea. Forecasting*, **35**, 537–559, <https://doi.org/10.1175/WAF-D-19-0170.1>.
- Johnson, A., X. Wang, Y. Wang, A. Reinhart, A. J. Clark, and I. L. Jirak, 2020: Neighborhood- and object-based probabilistic verification of the OU MAP ensemble forecasts during 2017 and 2018 Hazardous Weather Testbeds. *Wea. Forecasting*, **35**, 169–191, <https://doi.org/10.1175/WAF-D-19-0060.1>.
- Johnson, J. T., P. L. Mackeen, A. Witt, E. D. Mitchell, G. Stumpf, M. D. Eilts, and K. W. Thomas, 1998: The Storm Cell Identification and Tracking algorithm: An enhanced WSR-88D algorithm. *Wea. Forecasting*, **13**, 263–276, [https://doi.org/10.1175/1520-0434\(1998\)013<0263:TSCIAT>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0263:TSCIAT>2.0.CO;2).
- Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast system. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, <https://doi.org/10.1175/WAF-D-15-0107.1>.
- Lack, S. A., and N. I. Fox, 2012: Development of an automated approach for identifying convective storm type using reflectivity-derived and near-storm environment data. *Atmos. Res.*, **116**, 67–81, <https://doi.org/10.1016/j.atmosres.2012.02.009>.
- , G. L. Limpert, and N. I. Fox, 2010: An object-oriented multiscale verification scheme. *Wea. Forecasting*, **25**, 79–92, <https://doi.org/10.1175/2009WAF2222245.1>.
- Lagerquist, R. A., A. McGovern, and T. M. Smith, 2017: Machine learning for real-time prediction of damaging straight-line convective wind. *Wea. Forecasting*, **32**, 2175–2193, <https://doi.org/10.1175/WAF-D-17-0038.1>.
- Lakshmanan, V., and T. Smith, 2009: Data mining storm attributes from spatial grids. *J. Atmos. Oceanic Technol.*, **26**, 2353–2365, <https://doi.org/10.1175/2009JTECHA1257.1>.
- , R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *Atmos. Res.*, **67–68**, 367–380, [https://doi.org/10.1016/S0169-8095\(03\)00068-1](https://doi.org/10.1016/S0169-8095(03)00068-1).
- , K. Hondl, and R. Rabin, 2009: An efficient, general-purpose technique for identifying storm cells in geospatial images. *J. Atmos. Oceanic Technol.*, **26**, 523–537, <https://doi.org/10.1175/2008JTECHA1153.1>.
- Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, <https://doi.org/10.1175/WAF-D-17-0145.1>.
- , C. K. Potvin, P. S. Skinner, and A. E. Reinhart, 2021: The vice and virtue of increased horizontal resolution in ensemble forecasts of tornadic thunderstorms in low-CAPE, high-shear environments. *Mon. Wea. Rev.*, **149**, 921–944, <https://doi.org/10.1175/MWR-D-20-0281.1>.

- Lemon, L. R., and C. A. Doswell, 1979: Severe thunderstorm evolution and mesocyclone structure as related to tornadogenesis. *Mon. Wea. Rev.*, **107**, 1184–1197, [https://doi.org/10.1175/1520-0493\(1979\)107<1184:STEAMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1979)107<1184:STEAMS>2.0.CO;2).
- Mahalik, M., B. Smith, K. Elmore, D. Kingfield, K. Ortega, and T. Smith, 2019: Estimates of gradients in radar moments using a linear least squares derivative technique. *Wea. Forecasting*, **34**, 415–434, <https://doi.org/10.1175/WAF-D-18-0095.1>.
- Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, <https://doi.org/10.1175/2009JAS2965.1>.
- Micheas, A., N. I. Fox, S. A. Lack, and C. K. Wikle, 2007: Cell identification and verification of QPF ensembles using shape analysis techniques. *J. Hydrol.*, **343**, 105–116, <https://doi.org/10.1016/j.jhydrol.2007.05.036>.
- Miller, W., and Coauthors, 2022: Exploring the usefulness of downscaling free forecasts from the Warn-on-Forecast system. *Wea. Forecasting*, **37**, 181–203, <https://doi.org/10.1175/WAF-D-21-0079.1>.
- Ortega, K., T. Smith, J. Zhang, C. Langston, Y. Qi, S. Stevens, and J. Tate, 2012: The Multi-Year Reanalysis of Remotely Sensed Storms (MYRORSS) project. *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., 74, <https://ams.confex.com/ams/26SLS/webprogram/Paper211413.html>.
- Parker, M. D., and R. H. Johnson, 2000: Organizational modes of midlatitude mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 3413–3436, [https://doi.org/10.1175/1520-0493\(2001\)129<3413:OMOMMC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<3413:OMOMMC>2.0.CO;2).
- Potvin, C. K., and Coauthors, 2019: Systematic comparison of convection-allowing models during the 2017 NOAA HWT Spring Forecasting Experiment. *Wea. Forecasting*, **34**, 1395–1416, <https://doi.org/10.1175/WAF-D-19-0056.1>.
- , and Coauthors, 2020: Assessing systematic impacts of PBL schemes on storm evolution in the NOAA Warn-on-Forecast system. *Mon. Wea. Rev.*, **148**, 2567–2590, <https://doi.org/10.1175/MWR-D-19-0389.1>.
- Roerdink, J., and A. Meijster, 2001: The watershed transform: Definitions, algorithms and parallelization strategies. *Fundam. Inf.*, **41**, 187–228, <https://doi.org/10.3233/FI-2000-411207>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Skinner, P. S., and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast system. *Wea. Forecasting*, **33**, 1225–1250, <https://doi.org/10.1175/WAF-D-18-0020.1>.
- 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. *Wea. Forecasting*, **27**, 1114–1135, <https://doi.org/10.1175/WAF-D-11-00115.1>.
- , T. E. Castellanos, A. C. Winters, C. M. Mead, A. R. Dean, and R. L. Thompson, 2013: Measured severe convective wind climatology and associated convective modes of thunderstorms in the contiguous United States, 2003–09. *Wea. Forecasting*, **28**, 229–236, <https://doi.org/10.1175/WAF-D-12-00096.1>.
- Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, <https://doi.org/10.1175/BAMS-D-14-00173.1>.
- Sobash, R. A., C. S. Schwartz, D. A. Ahijevych, D. J. Gagne II, C. Potvin, M. L. Weisman, and S. J. Weiss, 2021: Building a machine learning-based system to objectively identify the convective mode in convection-allowing model output. *20th Conf. on Artificial Intelligence for Environmental Science*, Online, Amer. Meteor. Soc.
- Starzec, M., C. R. Homeyer, and G. L. Mullendore, 2017: Storm Labeling in Three Dimensions (SLD3): A volumetric radar echo and dual-polarization updraft classification algorithm. *Mon. Wea. Rev.*, **145**, 1127–1145, <https://doi.org/10.1175/MWR-D-16-0089.1>.
- Steiner, M., R. A. Houze Jr., and S. E. Yuter, 1995: Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data. *J. Appl. Meteor.*, **34**, 1978–2007, [https://doi.org/10.1175/1520-0450\(1995\)034<1978:CCOTDS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034<1978:CCOTDS>2.0.CO;2).
- Stensrud, D. J., and Coauthors, 2009: Convective-scale Warn-on-Forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, <https://doi.org/10.1175/2009BAMS2795.1>.
- , and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, <https://doi.org/10.1016/j.atmosres.2012.04.004>.
- Stumpf, G. J., A. Witt, E. D. Mitchell, P. L. Spencer, J. T. Johnson, M. D. Eilts, K. W. Thomas, and D. W. Burgess, 1998: The National Severe Storms Laboratory mesocyclone detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 304–326, [https://doi.org/10.1175/1520-0434\(1998\)013<0304:TNSSLM>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0304:TNSSLM>2.0.CO;2).
- 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. *Wea. Forecasting*, **20**, 23–34, <https://doi.org/10.1175/WAF-835.1>.
- Van der Walt, S., J. L. Schonberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, 2014: Scikit-image: Image processing in Python. *PeerJ*, **2**, e453, <https://doi.org/10.7717/peerj.453>.
- Weisman, M. L., and R. Rotunno, 2004: “A theory for strong long-lived squall lines” revisited. *J. Atmos. Sci.*, **61**, 361–382, [https://doi.org/10.1175/1520-0469\(2004\)061<0361:ATFSLS>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<0361:ATFSLS>2.0.CO;2).
- Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast system. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, <https://doi.org/10.1175/WAF-D-15-0043.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.