2015

Thunderstorm Observation by Radar (ThOR): An Algorithm to Develop a Climatology of Thunderstorms

Adam L. Houston
University of Nebraska—Lincoln, ahouston2@unl.edu

Noah A. Lock
Weather Decision Technologies

Jamie Lahowetz
University of Nebraska—Lincoln, jlahowetz2@unl.edu

Brian L. Barjenbruch
National Weather Service

George L. Limpert
University of Nebraska-Lincoln, george.limpert@unl.edu

See next page for additional authors

Authors
Adam L. Houston, Noah A. Lock, Jamie Lahowetz, Brian L. Barjenbruch, George L. Limpert, and Cody Oppermann

Thunderstorm Observation by Radar (ThOR): An Algorithm to Develop a Climatology of Thunderstorms

ADAM L. HOUSTON,* NOAH A. LOCK,+ JAMIE LAHOWETZ,# BRIAN L. BARJENBRUCH,@ GEORGE LIMPERT,* AND CODY OPPERMANN&

* *Department of Earth and Atmospheric Sciences, University of Nebraska–Lincoln, Lincoln, Nebraska*
+ *Weather Decision Technologies, Norman, Oklahoma*
# *High Plains Regional Climate Center, Lincoln, Nebraska*
@ *National Weather Service, Topeka, Kansas*
& *Utah Department of Transportation, Salt Lake City, Utah*

(Manuscript received 29 May 2014, in final form 24 November 2014)

ABSTRACT

The Thunderstorm Observation by Radar (ThOR) algorithm is an objective and tunable Lagrangian approach to cataloging thunderstorms. ThOR uses observations from multiple sensors (principally multisite surveillance radar data and cloud-to-ground lightning) along with established techniques for fusing multisite radar data and identifying spatially coherent regions of radar reflectivity (clusters) that are subsequently tracked using a new tracking scheme. The main innovation of the tracking algorithm is that, by operating offline, the full data record is available, not just previous cluster positions, so all possible combinations of object sequences can be developed using all observed object positions. In contrast to Eulerian methods reliant on thunder reports, ThOR is capable of cataloging nearly every thunderstorm that occurs over regional-scale and continental United States (CONUS)-scale domains, thereby enabling analysis of internal properties and trends of thunderstorms.

ThOR is verified against 166 manually analyzed cluster tracks and is also verified using descriptive statistics applied to a large (~35 000 tracks) sample. Verification also relied on a benchmark tracking algorithm that provides context for the verification statistics. ThOR tracks are shown to match the manual tracks slightly better than the benchmark tracks. Moreover, the descriptive statistics of the ThOR tracks are nearly identical to those of the manual tracks, suggesting good agreement. When the descriptive statistics were applied to the ~35 000-track dataset, ThOR tracking produces longer (statistically significant), straighter, and more coherent tracks than those of the benchmark algorithm. Qualitative assessment of ThOR performance is enabled through application to a multiday thunderstorm event and comparison to the behavior of the Storm Cell Identification and Tracking (SCIT) algorithm.

1. Introduction

The thunderstorm is a phenomenon in the physical climate system that can produce a severe impact on humans. Between 1995 and 2013, the hazards produced by thunderstorms (tornado, lightning, hail, flash flooding, and thunderstorm winds) were responsible for 4037 deaths, 40 371 injuries, and $99.9 billion (2013 dollars) in damage (NOAA 2013). It is therefore essential to understand the spatiotemporal variability of thunderstorms and expose spatiotemporal patterns that are a consequence of climate variability and climate change. In turn, trends in such patterns could serve as barometers for climate change. A robust thunderstorm climatology requires a method for identifying and tracking thunderstorms that uses data with a fidelity that is capable of resolving thunderstorm structure and life cycle. The complexity of identifying and tracking thunderstorms and the size of the dataset required to undertake this task have meant that a true thunderstorm climatology for the United States has not yet been produced.

To properly frame the approach required to fill this data void, it is necessary to establish a rigorous definition for a thunderstorm. The *Glossary of Meteorology* (American Meteorological Society 2013) defines a thunderstorm as a mesometeorological "disturbed state

Corresponding author address: Dr. Adam L. Houston, Department of Earth and Atmospheric Sciences, University of Nebraska–Lincoln, 214 Bessey Hall, Lincoln, NE 68588.
E-mail: ahouston2@unl.edu

DOI: 10.1175/JTECH-D-14-00118.1

of the atmosphere'' produced by a cumulonimbus cloud and is always accompanied by lightning and thunder. This definition departs somewhat from the definitions of Houze (1993), Cotton and Anthes (1989), and others, who state that a thunderstorm is not *produced by* a cumulonimbus cloud but *is* the cumulonimbus cloud. We assert here that the most robust definition of a thunderstorm is deep moist convection-producing thunder. In this definition, the thunderstorm is treated as a *process* (deep moist convection) and the cumulonimbus cloud is treated as the most direct manifestation of this process. Unfortunately, it is nearly impossible to extend this definition to the development of a thunderstorm climatology, since conventional observing systems are only capable of resolving the manifestations of thunderstorms, not the process that defines them. Therefore, in this work, the definition of a thunderstorm as a thundering cumulonimbus cloud will be adopted. According to this definition, a robust climatology of thunderstorms cannot be based solely on Eulerian observations of thunderstorm products (e.g., lightning, thunderstorm, precipitation) but requires a Lagrangian approach that embraces the principle that a thunderstorm is an entity with a spatial scale, motion, and life cycle.

Previous ''thunderstorm'' climatologies have been predominantly developed using Eulerian approaches. These studies were directed toward establishing whether a given location in the study region experienced a thunderstorm, termed a thunderstorm event. Such climatologies (e.g., Alexander 1915, 1935; Kuo and Orville 1973; Wallace 1975; Reap and Foster 1979; Court and Griffiths 1981; Falconer 1984; Changnon 1985; Easterling and Robinson 1985; Michaels et al. 1987; Changnon 1988a,b) have served as the benchmarks for the spatiotemporal distributions of thunderstorms. However, to develop a robust thunderstorm climatology, it is necessary to adopt a Lagrangian-based methodology that is capable of cataloging every thunderstorm that occurs over the spatial domain and period of record. With these data it will be possible to catalog internal properties of individual thunderstorms and to examine thunderstorm characteristics and trends (e.g., how has the per-thunderstorm lightning productivity changed in the United States over the last decade?).

Input data for a Lagrangian approach to developing a thunderstorm climatology must satisfy the following requirements: the spatial granularity must be capable of resolving thunderstorm structure, the temporal granularity must be capable of resolving thunderstorm evolution, and lightning and/or thunder must be identified. Additionally, to answer some of the most pressing questions regarding thunderstorm distribution and behavior, the geographic domain must be large enough to minimize the influence of local or regional effects and the period of record must be long enough to resolve diurnal and seasonal cycles.

In general, radar data are ideally suited for the development of thunderstorm climatologies because 1) they are characterized by spatiotemporal granularity capable of resolving thunderstorm structure and evolution, 2) even a single radar can collect data over a sizeable geographic footprint [$O(10^5\,\mathrm{km}^2)$], and 3) the retrieved volumetric dataset offers a more complete perspective of the thunderstorm structure than the 2D perspective offered by satellite data.

To minimize attenuation, surveillance radars operate at frequencies that produce negligible scattering from cloud particles. As such, surveillance radars do not detect the cumulonimbus cloud that defines the thunderstorm but instead detect precipitation-sized hydrometeors produced *by* the cloud. Cumulonimbus clouds that fail to produce precipitation somewhere within the cloud column are assumed too rare to challenge the assertion that radar data can be used to accurately identify thunderstorms.

Since radar data cannot detect lightning, approaches that rely solely on radar data cannot reveal whether a cumulonimbus cloud is actually producing thunder. Reap and Foster (1979) demonstrate that high-reflectivity radar returns are usually associated with thunder and that low-reflectivity returns usually are not. However, the most robust method for identifying thundering cumulonimbus clouds is to synthesize radar-derived cumulonimbus cloud identifications with observations of lightning or thunder.

Like any observational network, there are limitations to using radar data; principally, radar horizon, beam spreading, attenuation (particularly for higher-frequency radars), and anomalous returns (ground clutter, multitrip echoes, etc.). Moreover, the amount of data that compose the volumetric datasets places a significant burden on computational resources tasked with data processing and storage.

Previous radar-based approaches to developing thunderstorm climatologies (e.g., Wiggert et al. 1981; López et al. 1984; MacKeen et al. 1999; Potts et al. 2000; May and Ballinger 2007; Clements and Orville 2008; Mohee and Miller 2010; Mosier et al. 2011; Davini et al. 2012; Seroka et al. 2012; Goudenhoofdt and Delobbe 2013) capitalize on the many benefits of using radar data. Moreover, each of these studies relies on an objective algorithm for thunderstorm identification that, in contrast to manual identification strategies, is repeatable and tunable. However, none of these methods operates on multiple radars covering a large region *and* includes

lightning. Mohee and Miller (2010) and Davini et al. (2012) utilize multiple radars over a regional footprint (North Dakota and northwestern Italy, respectively) but neither integrates lightning into their thunderstorm identification logic. Clements and Orville (2008), Mosier et al. (2011), and Seroka et al. (2012) utilize lightning for thunderstorm identification but use only a single radar for their analysis.

The Thunderstorm Observation by Radar (ThOR) algorithm presented herein similarly relies on radar data in its objective and tunable Lagrangian approach to cataloging thunderstorms. ThOR uses observations from multiple sensors (principally multisite surveillance radar data and cloud-to-ground lightning) and established techniques for fusing multisite radar data and identifying spatially coherent regions of radar reflectivity (clusters) that are subsequently tracked using a new and innovative tracking scheme. The tracking scheme developed for ThOR is designed to work on archived data and is therefore not constrained by the limitation of *real-time* tracking that the future state of the thunderstorm is unknown. The design of the tracking component has been guided by best practices of object tracking and has undergone rigorous quantitative verification.

The algorithm is described in section 2 and is organized as follows: overview (section 2a); event selection, data preprocessing, and quality control (section 2b); radar data merging and composite reflectivity (section 2c); stratiform precipitation attenuation (section 2d); reflectivity cluster identification (section 2e); cluster tracking (section 2f); and lightning association (section 2g). Section 3 presents results from the application of ThOR to a multiday event in April 2007. The summary is in section 4.

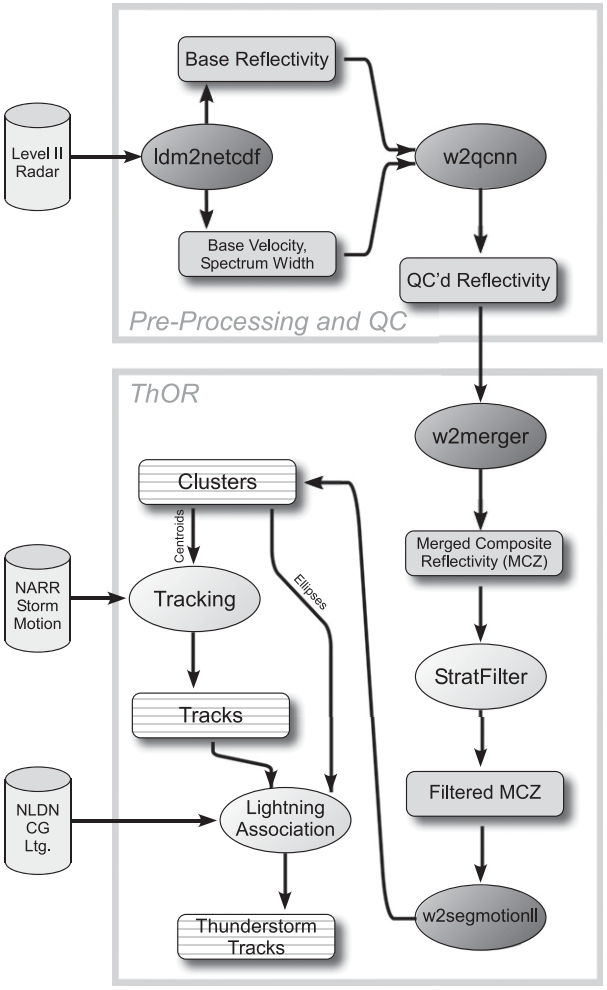

FIG. 1. Data and process flowchart for ThOR. Raw data sources appear as cylinders to the left, processed data are indicated with rounded rectangles, modules within the WDSS-II suite of tools are indicated with dark gray ellipses, and modules developed by the authors specifically for ThOR are indicated with light gray ellipses.

2. Algorithm description

a. Overview

ThOR involves five principal steps (Fig. 1): 1) merge radar data from multiple radars to a single volume on a common Cartesian grid, 2) attenuate stratiform precipitation to improve thunderstorm detection, 3) identify spatially coherent regions in radar reflectivity (i.e., clusters), 4) track clusters to develop candidate thunderstorms, and 5) associate cloud-to-ground (CG) lightning to candidate thunderstorms in order to classify tracked clusters as thunderstorms. Input datasets are level II radar volumes for individual WSR-88D, CG flashes from Vaisala's National Lightning Detection Network (NLDN), and the North American Regional Reanalysis (NARR, Mesinger et al. 2006). The retrieval and processing of radar data presents the largest challenge to the efficient execution of ThOR. Thus, to minimize the amount of radar data, a procedure for identifying "events" likely to contain thunderstorms has been implemented outside of ThOR but is discussed in the following subsection in the interest of completeness. Moreover, since input data for ThOR are expected to be quality controlled and formatted appropriately, the preprocessing of data is also discussed.

The omission of in-cloud (IC) lightning from the lightning data used in ThOR will invariably result in an underestimation of the number of thunderstorms identified. The results of MacGorman et al. (2011) offer guidance for estimating the potential impact of this omission. Using total flashes detected by regional VHF

lightning mapping systems along with CG flashes detected by the NLDN, MacGorman et al. found that 88%–93% of the Oklahoma thunderstorms they studied produced at least one CG flash within one hour of the first IC flash. They found a similar ratio for thunderstorms analyzed in North Texas and a slightly lower ratio (80%) for thunderstorms analyzed over the high plains of Colorado, Kansas, and Nebraska. These ratios probably indicate the upper bound of the fraction of thunderstorms that can be accurately detected when IC flashes are not used in thunderstorm identification. MacGorman et al. also found that the elapsed time to the first CG for 75% of the thunderstorms analyzed in Oklahoma, North Texas, and the high plains was 12, 23, and 44 min, respectively. Based on preliminary results, it is likely that the median thunderstorm duration will be approximately 40 min; thus, it is expected that the lower bound for the detection rate of thunderstorms without including IC lightning is probably not far from 75%, though this is likely to be sensitive to region as well as time of year.

The CG flash detection efficiency of the NLDN could also impact thunderstorm detection rates. CG flash detection by the NLDN exceeds 90% (Biagi et al. 2007) and, thus, missed CG detections are unlikely to result in significant numbers of missed thunderstorms.

The regional operations range of networks like the arrays used by MacGorman et al. (2011) are not suitable for the development of a national or even large regional catalog of thunderstorms. Total lighting detection systems covering the United States do exist [the reader is referred to the work of Rakov (2013) for a review of modern 3D lightning mapping networks], but short periods of record limit their utility for building a robust thunderstorm climatology. Future versions of ThOR could incorporate in-cloud lightning from local/regional LMAs and/or national total lightning detection systems in order to estimate the impact of using only CG lightning for thunderstorm identification.

b. *Event selection, data preprocessing, and quality control*

To minimize the amount of radar data processed by ThOR, radar data are acquired if and only if a CG lightning strike occurs within 230 km of a WSR-88D radar site. A 4-h buffer, centered on the time of the CG lightning, is imposed to account for thunderstorm tracks that may initiate or terminate within 2 h of the occurrence of CG lightning. Temporally contiguous blocks of data from all radars in range of CG lightning are aggregated into events. Events are likely to contain many hours of data from multiple radars but all data are in spatiotemporal proximity to CG lightning. The event designation simplifies the processing, since the development of a given thunderstorm track relies on data occupying a small spatiotemporal window; access to data covering the entire domain and period of record is not necessary.

The quality control of the radar data that are inputs to ThOR can be accomplished through any method that mitigates nonmeteorological echoes. For the results presented in section 3, the *w2qcnn* algorithm, packaged within the Warning Decision Support System–Integrated Information (WDSS-II; Lakshmanan et al. 2007) suite of radar processing/analysis algorithms, is used. This algorithm relies on a trained neural network based on 28 inputs from the reflectivity, velocity, and spectrum width fields, mainly from the lowest scans, to remove nonmeteorological errors. Heuristic modules are used to identify and remove echoes from insect "bloom," volumes collected in clear-air mode, and hardware problems. The algorithm provides the option to retain or discard pixels on either a pixel-by-pixel basis or on the basis of the majority classification of contiguous regions of pixels. In ThOR, the former method is used, as the latter sometimes erroneously removes areas of precipitation far from the radar. The algorithm is effective in removing most nonprecipitating echo from ground clutter and biological scatterers while retaining almost all precipitation echoes.

c. *Radar data merging and composite reflectivity*

Radar reflectivity fields from the multiple radars that constitute a given event must be merged into a single volume on a common Cartesian grid. This is the first step in ThOR. Merging in ThOR relies on the *w2merger* algorithm, which follows the approach described by Lakshmanan et al. (2006) and is included in WDSS-II. The w2merger algorithm accounts for varying radar beam geometry with range, vertical gaps between radar scans, and asynchronism between radars (Lakshmanan et al. 2006). It takes the most recent full volume from each radar in the event, using the time of the last elevation scan as the end of the full volume, and merges them together into a volume of reflectivity mosaics on a latitude–longitude–height grid. The blending of data from multiple radars relies on a distance-dependent weighting, $\exp(-r^2/\sigma)$, where r is the distance and σ is set to 50 km. Data are mapped to a 0.0137°, 0.011°, 1 km (latitude, longitude, height, respectively) grid. The algorithm also calculates the column-maximum reflectivity (composite reflectivity) at each grid point, resulting in a merged composite reflectivity (MCZ; Fig. 2a).

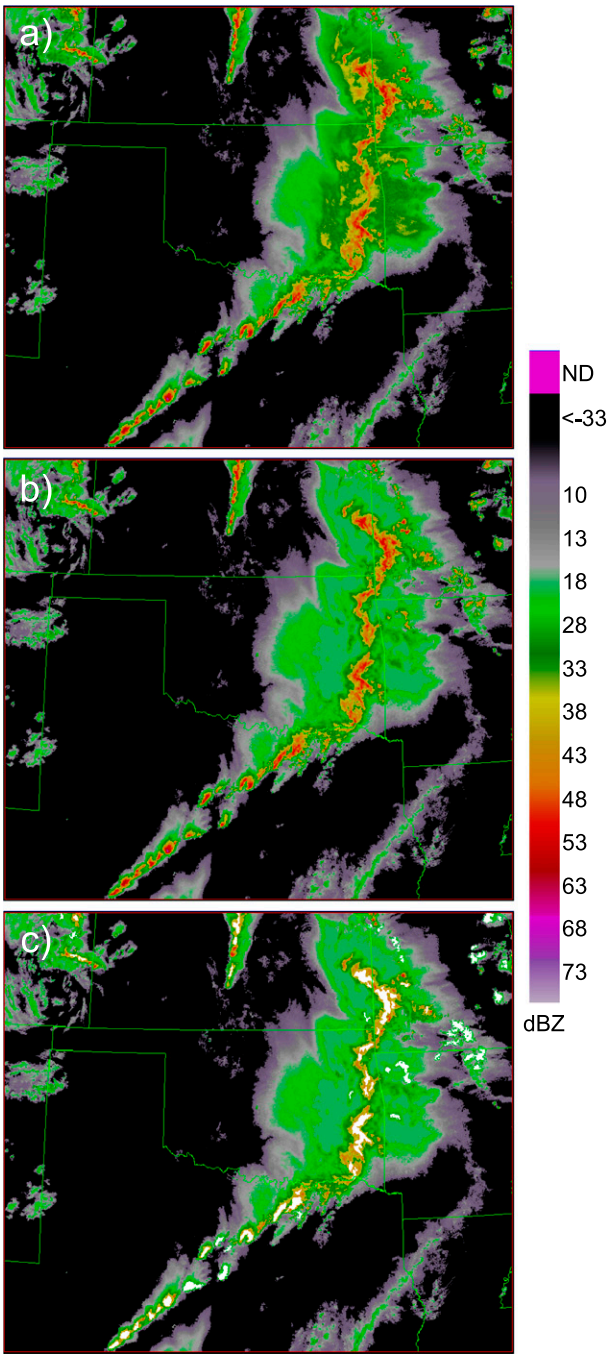

FIG. 2. (a) Merged composite reflectivity (MCZ) from 24 Apr 2007, (b) MCZ following stratiform filtering, and (c) filtered MCZ with clusters in white.

d. Stratiform precipitation attenuation

Preliminary testing of the reflectivity cluster identification algorithm used in ThOR, and described in detail in the following subsection, revealed that large stratiform regions are often identified as single clusters that the algorithm occasionally combines with smaller-scale convective regions. Since smaller-scale convective regions are taken to be the more direct manifestation of thunderstorms than the larger-scale stratiform regions (even though some stratiform precipitation may be a direct consequence of nearby deep convection), the artificial inflation of cluster size through consolidation needs to be prevented. By reducing the reflectivity values in stratiform regions below the threshold used to identify reflectivity clusters (30 dBZ for ThOR), these areas should no longer be detected. This attenuation of reflectivity values within stratiform regions requires first distinguishing stratiform precipitation from convective precipitation.

The stratiform–convective discrimination adopted in ThOR is similar to the approach of Biggerstaff and Listemaa (2000). The algorithm results in a stratiform–convective score for each column on the common Cartesian grid that is used to adjust the column's MCZ value using a fuzzy logic approach. The quantities incorporated into the algorithm are 1) the value of MCZ; 2) the magnitude of the horizontal gradient of MCZ, calculated using a sixth-order centered difference; 3) the vertical gradient of reflectivity calculated as the difference between the maximum reflectivity in the column and the reflectivity 3 km above the level of the maximum, divided by the distance between them;[1] and 4) the horizontal reflectivity gradient at a height 3 km above the level of maximum reflectivity.

For all columns in which the MCZ exceeds 25 dBZ, the stratiform–convective algorithm assigns a score (S) between zero and one, where zero indicates definitely stratiform and one indicates definitely convective. Columns with an MCZ less than 25 dBZ are not scored. Refer to the appendix for a description of the logic used to score columns. The values of MCZ are adjusted based on the associated stratiform–convective score using a fuzzy logic approach. A column with $S < 0.25$ is classified as ''definitely stratiform'' and the MCZ is set to the ''stratiform value,'' which is either 20 dBZ or the value of reflectivity 2 km above the level of maximum reflectivity, whichever is less. A column with $S > 0.55$ is classified as ''definitely convective'' and retains its original MCZ value, defined as the ''convective value.'' The MCZ value for a column with an intermediate value of S is assigned a reflectivity that is a weighted average of its convective value (original value) and its

[1] Similar to the approach used by Zipser and Lutz (1994), in order to account for storm tilt, the reflectivity value at 3 km above the level of maximum reflectivity is the maximum reflectivity within a 9 km × 9 km lateral box centered on the grid point.

stratiform value (defined above). The resultant filtered MCZ field (Fig. 2b) is then used to identify reflectivity clusters.

e. *Reflectivity cluster identification*

A reflectivity cluster is defined for this work as a spatially coherent region of reflectivity in the MCZ field. ThOR relies on the *w2segmotionll* algorithm of WDSS-II, which utilizes *k*-means clustering and a watershed transform that collectively enable cluster identification that is more adaptive than traditional methods of object identification that rely exclusively on global and arbitrarily defined field thresholds (Kolodziej Hobson et al. 2012). Specifically, cluster identification in w2segmotionll involves two principal steps (Lakshmanan et al. 2002, 2009; Kolodziej Hobson et al. 2012): 1) quantization of the reflectivity field using *k*-means clustering and 2) segmentation of the quantized field using a watershed transform [both *k*-means clustering and the watershed transform are image segmentation techniques that can be used for cluster identification; however, *k*-means clustering is used in w2segmotionll solely for field quantization (Kolodziej Hobson et al. 2012)].

The *k*-means clustering is an iterative procedure that aggregates gridded data into clusters with a unique, quantized *k* value. All grid points assume the *k* value of the parent cluster. Gridpoint assignment to a candidate cluster is achieved through minimization of a cost function defined according to 1) the difference between the grid point's texture vector and the mean texture of the candidate cluster (self-similarity; Lakshmanan and Smith 2009) and 2) the number of gridpoint neighbors that have a *k* value that is different from the *k* value of the candidate cluster (spatial coherence; Lakshmanan et al. 2002; Lakshmanan and Smith 2009). In w2segmotionll the texture vector is defined using local neighborhood statistics including mean, variance, and coefficient of variance (Lakshmanan et al. 2002). Once all grid points are assigned in a given iteration, the mean texture value of each cluster is recalculated using the new aggregate of grid points and the process of gridpoint assignment is repeated. This approach results in a quantized field that, unlike a simple rounding approach, considers the spatial arrangement of the data (Lakshmanan et al. 2002).

Following *k*-means clustering, the extended watershed transform of Lakshmanan et al. (2009) is used to segment the quantized field into clusters. The w2segmotionll watershed transform is a maxima-finding technique that operates by identifying regions that encompass local maxima and satisfy a saliency criterion defined by a preset minimum region size. Segmented regions that fall below the saliency criterion are combined with neighboring regions provided their separation is within a preset maximum distance and the regions are connected by pixels exceeding a minimum reflectivity value. In the ThOR implementation, the default saliency criterion is 50 km^2, the default maximum distance between "adjacent" clusters that can be combined is ~1.76 km, and the default minimum reflectivity value is 30 dB*Z*. All of these values can be adjusted by the user. An example of clusters identified from the filtered MCZ is illustrated in Fig. 2c.

The saliency criterion used for ThOR was determined through extensive, albeit largely qualitative, evaluation of cluster identification performance using a range of saliency criteria and a collection of manually analyzed cases with a variety of convective organizations. The 50-km^2 criterion was found to be the optimal compromise between smaller thresholds that allowed for too many small clusters that degraded track accuracy and increased the algorithm's computational burden, and larger thresholds that filtered out "reasonably" small-scale storms. Since the cluster tracking step occurs prior to lightning attribution (Fig. 1) small short-lived deep convection not associated with lighting must still be included in the cluster tracking step. Based on empirical evidence collected as part of this work, even with a carefully crafted tracking logic, described below, these small-scale clusters prove to degrade tracking accuracy. Moreover, as a consequence of ThOR's approach to tracking, increased cluster numbers have the potential to exponentially increase the computation time.

The default saliency criterion used in ThOR should render satisfactory results over the United States. While the regional dependence of the saliency criterion is beyond the scope of this work, the user should be aware that additional tuning of this threshold may be necessary if ThOR is applied to regions with climates that tend to support small (typically air mass) thunderstorms (e.g., Florida).

The 30-dB*Z* minimum reflectivity value used for ThOR has precedent in the work of Potts et al. (2000), Gallus et al. (2008), and Duda and Gallus (2010), but it is lower than the thresholds used by Hocker and Basara (2008) and Goudenhoofdt and Delobbe (2013), who used a threshold of 40 dB*Z*, and May and Ballinger (2007), Davini et al. (2012), and Smith et al. (2012), who used a threshold of 35 dB*Z*. Convective precipitation is generally assumed to have reflectivity values exceeding 40 dB*Z* (Steiner et al. 1995; Geerts 1998; Parker and Johnson 2000); however, the lower threshold used here is justified because stratiform precipitation is attenuated and lightning is included.

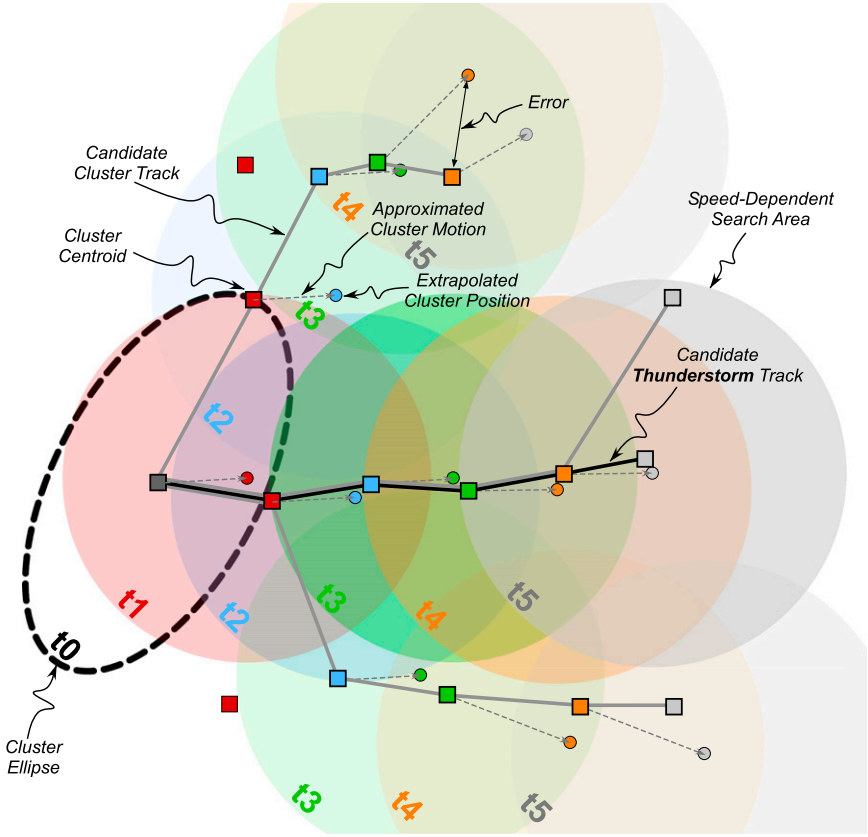

FIG. 3. Illustration of the tracking procedure used in ThOR. Cluster centroids at the initial time (t0) and five subsequent times (t1–t5) are illustrated with dark gray, red, blue, green, orange and light gray squares. The initial cluster ellipse is illustrated in black. Extrapolated cluster positions from t0–t5 cluster centroids appear as colored circles. Search areas around extrapolated cluster positions appear as semitransparent circles colored according to the time that they are valid. Candidate cluster tracks are illustrated with gray lines, and the candidate thunderstorm track is illustrated with black lines.

f. Cluster tracking

Like the Storm Cell Identification and Tracking (SCIT; Johnson et al. 1998) algorithm, the Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN; Dixon and Wiener 1993) algorithm, and other object tracking approaches (e.g., Stumpf et al. 1998; Root et al. 2011), ThOR tracks object centroids. Centroids in ThOR are defined as the center of an ellipse fit to the shape of the cluster. Ellipses are used instead of the irregularly shaped clusters because they facilitate a superior interpolation of cluster position and size (used for lightning attribution, described in section 2g).

Unlike SCIT, TITAN, and the mesocyclone detection algorithm (MDA; Stumpf et al. 1998), ThOR is designed for *offline* processing of archived data, not for *online* (real time) nowcasting. For online algorithms, the track of an object must be developed using only the previous positions of the object; future positions are unknown. However, offline algorithms have access to the full data record, so that all possible combinations of object sequences can be developed using all possible observed object positions. The result of offline tracking algorithms like ThOR is a collection of candidate tracks that branch away from each object being tracked (Fig. 3).

The candidate track, within the set of possible tracks, that most likely corresponds to the correct track is generally taken to be the track that minimizes a cost function defined according to intratrack variability in object characteristics. In ThOR, focus is placed on minimizing intratrack variability in object velocity by selecting as the best track the *candidate cluster track* with the lowest track-averaged error (Fig. 3). Track-averaged error is defined as the geodesic difference between the position of an observed cluster centroid along the track and the extrapolated position from the previous cluster

position, averaged over the duration of the candidate cluster track.

Prioritizing the minimization of intratrack variability in object velocity requires an accurate first guess for cluster motion. In ThOR, the 0–6-km mean wind from NARR (Mesinger et al. 2006) is used as the initial motion estimate. NARR data are distributed on a Lambert conic conformal projection grid with 32-km gridpoint spacing. Data are available at a 3-h temporal granularity. NARR data are used instead of Rapid Update Cycle (RUC) analyses because NARR data cover the entire NEXRAD period of record. The mean wind is temporally interpolated to the cluster observation time, and the value from the NARR grid point nearest to the observed cluster centroid is used for the first guess of cluster motion. NARR winds are used for an increment of time that is set using a user-tunable *narrbound* variable. Based on track parameter training, described in section 2.f.1), narrbound is set to a default of 10 min. As the track gets longer, a reliable motion estimate can be obtained from the position history of the cluster. Therefore, between narrbound and 30 min, the motion estimate is the weighted average of the NARR mean wind and the previous cluster motion along the candidate cluster track. After 30 min, the motion estimate is entirely based on the average motion calculated from the position history over a time interval equivalent to the variable *motiontime*, set to a default of 45 min [section 2f(1)].

To bound the exponential growth in the number of candidate cluster tracks (and attendant processing time) in offline algorithms, a gating function is often implemented to limit the number of objects that can be used to develop candidate cluster tracks (Stumpf et al. 1998; Scharenbroich et al. 2010). A gating function constrains the search area for objects to associate to tracks. Its use is not only justified as a way to bound processing time but is also justified based on uncertainties in object motion estimates; high-confidence estimates of object motion necessitate a smaller search area. For ThOR, the two most likely sources of thunderstorm motion estimate uncertainty are as follows: 1) inconsistencies between the NARR mean wind and the actual cluster motion resulting from (i) natural variability in storm motion owing to the inherent unsteadiness of thunderstorm propagation and/or (ii) inconsistencies between the NARR mean wind and the actual steering current; and 2) random variations in the location of the object centroid stemming from changes in the size and/or shape of clusters (the NARR mean wind may exactly match the actual thunderstorm motion but centroid wobbles caused by morphological changes in the cluster cause the detected centroid position to depart from the extrapolated position). A robust gating function must be structured to allow for these motion estimate uncertainties while being restrictive enough to limit the number of candidate thunderstorm tracks (processing time) and to minimize the chance that a cluster is assimilated into the wrong track.

The gating function used in ThOR is defined according to a dynamic search radius (R). Based on the results from algorithm training discussed below, R depends on the storm speed and the length of time over which the cluster position extrapolation is calculated.

An important attribute of the ThOR tracking is the allowance made for missed cluster detections. All object identification algorithms are prone to miss objects and the w2segmotionll algorithm is no exception. To prevent these missed identifications from prematurely terminating tracks and thereby inflating thunderstorm numbers and deflating mean thunderstorm duration, time skipping is incorporated into the algorithm: ThOR will search both one and two time steps ahead for clusters to associate to a track.

A summary of the tracking procedure is illustrated in Fig. 4. Time skipping is reflected in the steps involving the calculation of extrapolated cluster positions and search radii for which time intervals (Δt) of both one and two time steps (δt) are used. The result of the tracking procedure is a collection of candidate thunderstorm tracks.

The following two subsections describe the training [section 2f(1)] and verification [section 2f(2)] of ThOR. Algorithm training was conducted to tune both the gating function and the parameter values for integrating the NARR winds into the motion estimates (narrbound and motiontime), while algorithm verification was designed to assess the overall accuracy of ThOR tracking. Both the training and verification relied on manual analysis of tracks developed collaboratively by a committee of three meteorologists. These manual tracks were designed to represent the ''best practices'' in tracking and were used as the reference for ''correct'' tracks. To ensure that the manual tracks were superior to algorithmically derived tracks, a committee of three meteorologists was afforded access to data besides just cluster centroids—specifically, cluster boundaries and the MCZ field from which the clusters were derived. Since the focus of the training was on the tracking only, not the cluster identification component, the committee was not responsible for identifying clusters; the algorithms and the manual tracks were based on the same sets of clusters.

1) TRACK PARAMETER TRAINING

A novel approach to training was adopted for this work. For each value of narrbound and motiontime

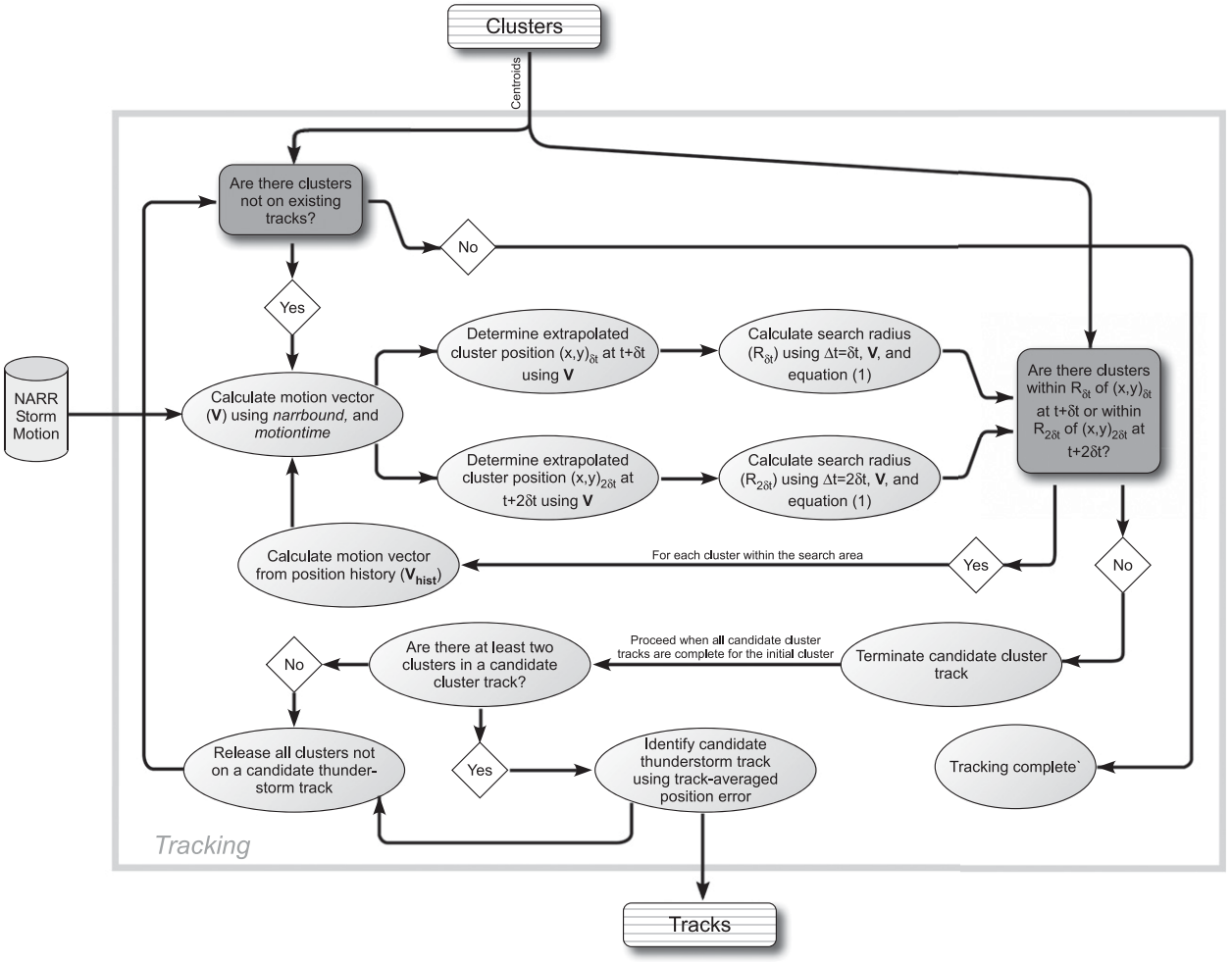

FIG. 4. Data and process flowchart for the tracking component of ThOR. The time interval between merged reflectivity volumes is denoted δt.

tested [narrbound = (0, 10, 20 min), motiontime = (30, 45, 60 min)], an estimated track was created for each segment of each manually analyzed cluster track. For clusters *within* a manual track and not at the track terminus, the minimum search radius (r_{min}) that would have been required to create each segment of the track (i.e., the value required to "catch" the next cluster in the manual track) was calculated. For clusters at the terminus of a manual track, r_{min} was the minimum value that would incorrectly extend a track (i.e., incorrectly associate a cluster to the track). Properly tuned values of narrbound and motiontime should yield cluster motions with a maximum separation between the small r_{min} values that are required to correctly continue a track and the large r_{min} values that would result in the incorrect continuation of a track beyond its terminus. A properly tuned gating function should have an R that is larger than the r_{min} values that are required to correctly continue a track and smaller than the r_{min} values that incorrectly continue a track beyond its terminus.[2] To represent a variety of storm modes and storm speeds in the training set, a collection of 62 tracks from 16 events was considered (Table 1). Differences in the separation between r_{min} for correct and incorrect track continuations across the combinations of narrbound and motiontime were small (not shown). A slightly better discrimination was found for narrbound = 10 min and motiontime = 45 min; thus, these were chosen as the default values.

[2] Incorrect continuations do not include the erroneous association of clusters on nearby tracks because the logic of the algorithm will likely filter out erroneous tracks developed through incorporation of nearby tracks through the minimization of the track-averaged error. In contrast, the track-averaged error will not filter out incorrect continuations beyond the actual track terminus.

TABLE 1. List of events used for ThOR training.

Start	End	Region	Type of storm
2026 UTC 9 Jul 2005	0140 UTC 10 Jul 2005	West TX	Multicell
1800 UTC 13 Jul 2005	1943 UTC 13 Jul 2005	East OK, West AR	Ordinary cells
1709 UTC 14 Jul 2005	2015 UTC 14 Jul 2005	East OK, West AR	Ordinary cells
1906 UTC 15 Jul 2005	2106 UTC 15 Jul 2005	East OK, West AR	Ordinary cells
0100 UTC 13 Jan 2006	0800 UTC 13 Jan 2006	AR	Squall line
0055 UTC 9 Mar 2006	1005 UTC 9 Mar 2006	TX, OK, AR	Supercell plus squall line
1740 UTC 11 Mar 2006	0000 UTC 12 Mar 2006	OK, MO	Supercells
0010 UTC 13 Mar 2006	0710 UTC 13 Mar 2006	OK, KS, MO	Supercells
1935 UTC 2 Apr 2006	0140 UTC 3 Apr 2006	North AR, MO, IL	Supercell plus squall line
2120 UTC 12 Apr 2006	0020 UTC 13 Apr 2006	AR	Backbuilding multicell
1840 UTC 16 May 2006	2105 UTC 16 May 2006	IA, MO	Ordinary cells, weak multicell
2305 UTC 28 Feb 2007	0515 UTC 1 Mar 2007	KS, MO	Supercell
1905 UTC 17 Apr 2007	2235 UTC 17 Apr 2007	TX	Squall line
0505 UTC 24 Apr 2007	0820 UTC 24 Apr 2007	KS, NE, IA	Multicell
1355 UTC 24 Apr 2007	1535 UTC 24 Apr 2007	CO	Elevated cells near upper low
2105 UTC 24 Apr 2007	0320 UTC 25 Apr 2007	South TX	Supercell

Using the default values of narrbound and motion-time, the gating function was tuned by first considering the distribution of $r_{\min}$ to develop a threshold value/formula for R. Since missed clusters along a track are allowed, the gating function must account for both a standard time interval between clusters and a larger time interval over which missed clusters may exist. As such, track projections from the clusters along the manually analyzed tracks were separately analyzed for a single time step ($\Delta t = 5$ min) and two time steps ($\Delta t = 10$ min).

Preliminary analysis of the search radii across the set of tracks revealed that correct continuations required larger search radii as storm speed increased (Fig. 5a, "plus" symbols denote correct continuations). Search radii for incorrect continuations were largely independent of storm speed (Fig. 5a). Using Fisher's linear discriminant (FLD; Wilks 2011)[3] applied to the distribution of correct continuations and incorrect continuations as a function of storm speed (V), a formula was derived for R (km) as a function of V (m s^{-1}) and Δt (min):

$$R = 9.04 + a(\Delta t - 5)/5 + 0.1427V, \qquad (1)$$

where a is set to 3 km by default. The term $a(\Delta t - 5)/5$ is included as a succinct way of imposing adaptability in R to a variable time interval and emerged from FLD analysis of the distribution of search radii relative to V for $\Delta t = 10$ min.

Based on the recommendation of Johnson et al. (1998), the impact of a dynamic search cone width (referred to as directional thresholding) within the gating function was examined. Directional thresholding was proposed by Johnson et al. as a way of limiting incorrect track continuations in the SCIT algorithm. The search cone formulation examined was based on R and for the distance a cluster was projected to travel since its last known location (D): $\Delta\theta = \tan^{-1}(R/D)$. Noting that, according to (1), $R = 10.47$ km for $V = 10$ m s^{-1} ($D = 3$ km for $\Delta t = 5$ min); without directional thresholding there is sizable region, both ahead of and behind the centroid, that could contain clusters that, by virtue of the large directional change that would be required for the cluster to be included as a continuation of the track, are likely incorrect continuations. Thus, ostensibly, some measure of directional thresholding in the gating function would seem to be justified. However, the training revealed very few incorrect continuations within the search radius and outside the search cone (i.e., incorrect continuations that required directional thresholding to be filtered out), but numerous correct continuations that were within the search radius but outside the search cone (i.e., correct continuations that would have been filtered out with the inclusion of directional thresholding). As a result, directional thresholding was not included in the final formulation of the gating function.

It is important to acknowledge that the results of the training depend strongly on the approach to cluster identification adopted for this work. For example, analysis of SCIT clusters (section 3) revealed a

[3] The incorrect continuations associated with search radii greater than 30 km were removed prior to calculating the FLD in order to satisfy the assumption of similar covariance matrices between the correct continuation and the incorrect continuation sets. This is justified, since there were no correct continuations at such a large range.

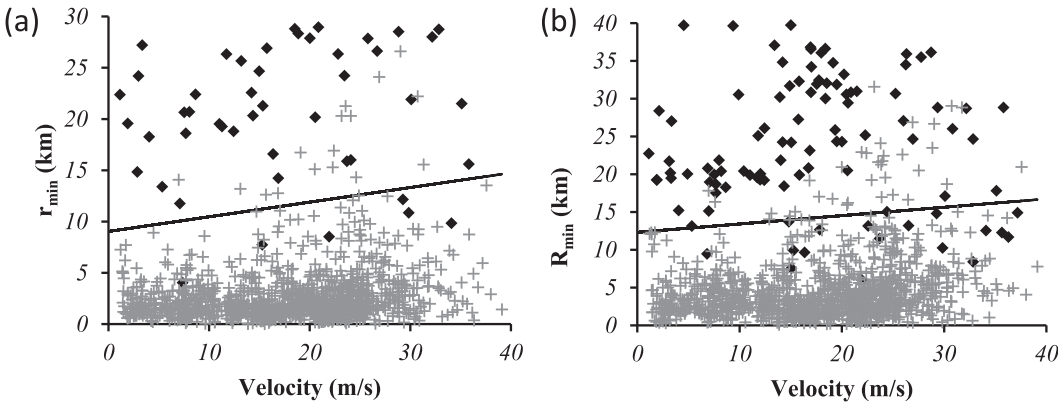

FIG. 5. Search radii (r_{min}) as a function of storm speed for correct continuations (gray "plus" symbols) and incorrect continuations (black diamonds) for a time step of (a) 5 and (b) 10 min. The black lines represent the function expressed in (1) for the dynamic search radius (R) derived from Fisher's line discriminant.

significantly higher spatial density compared to clusters identified using the approach described in section 2e. As such, while directional thresholding proved to be unnecessary for ThOR, it might improve the tracking using SCIT centroids (Stumpf et al. 1998). Moreover, the values of narrbound and motiontime, and the formulation of the gating function will likely change if the procedure for cluster identification is altered.

2) TRACKING VERIFICATION

A two-part approach to verifying the ThOR tracking algorithm is adopted. The first component is loosely modeled off the work of Johnson et al. (1998). Johnson et al. verified the SCIT algorithm by comparing SCIT tracks to manually analyzed tracks, assessing how many times SCIT made the correct association from one time to the next. While a comparison to a correct track is an attractive approach, the Johnson et al. method is limited by the lack of specificity, overestimation of skill, and the labor-intensive nature of the manual tracking (Lakshmanan and Smith 2010). For ThOR tracking verification, ThOR tracks are compared to manually analyzed tracks from a small set of events; however, additional specificity and more accurate skill assessment are enabled through a descriptive contingency table (Table 2).

The second component of tracking verification is modeled off of the work of Lakshmanan and Smith (2010), who propose using a set of descriptive statistics that describe overall track behavior and can be used to compare different tracking algorithms. Specifically, they propose using 1) median track duration, 2) the RMSE of the track compared to a linear fit (curvature), and 3) the standard deviation of vertically integrated liquid (VIL) along the track (incoherence). The philosophy of this strategy is that a better algorithm should produce longer, straighter, and more coherent

TABLE 2. Contingency table for paired tracks.

		Cluster on ThOR track at a given time	
Cluster on manually analyzed track at a given time		Yes	No
	Yes	Hit (h)	Miss (m) Type 1 ($m1$): ThOR track starts too late Type 2 ($m2$): ThOR track ends too early Type 3 ($m3$): Manual and ThOR tracks exist and have different clusters Type 4 ($m4$): Manual track includes a cluster that ThOR track skips
	No	False alarm (fa) Type 1 (fa1): ThOR track starts too early Type 2 (fa2): ThOR track keeps going Type 3 (fa3): ThOR track includes a cluster that manual track skips*	Correct negative (cn)

* As with ThOR tracking, manual tracking could include a "skip": a time step along a track does not have an associated cluster.

TABLE 3. List of events used for the manual tracking component of ThOR verification.

Start	End	Radar	Storm mode	Speed
0000 UTC 23 Jun 2003	0200 UTC 23 Jun 2003	KUEX	Supercells	Very slow
2130 UTC 7 Jul 2005	2330 UTC 7 Jul 2005	KUDX	Multicell	Slow
2000 UTC 15 Nov 2005	2200 UTC 15 Nov 2005	KNQA	Supercells and line segments	Fast
0200 UTC 06 Jun 2008	0400 UTC 06 Jun 2008	KTLX	Squall line	Moderate

(less incoherent) tracks. Since this approach does not require comparison to manually analyzed correct tracks, it can be applied to a much larger sample of tracks.

A potential disadvantage of the descriptive approach of Lakshmanan and Smith is that both curvature and incoherence may be correlated with the track duration (Reed and Trostel 2012). To correct for this dependence while still embracing the philosophy that a superior tracking algorithm produces straighter, more coherent tracks, the curvature is calculated as a local curvature, referred to as *jitter*, defined as $\{[1/(N-2)]\sum_{t=2}^{N-1} d_t^2\}^{1/2}$, where N is the total number of clusters along a track and d_t is the distance between a cluster at time t and the linearly interpolated position at time t using centroid positions at $t-1$ and $t+1$. *A straighter track would exhibit smaller jitter*. For the sample of ThOR-derived candidate thunderstorm tracks (tracks prior to lightning attribution) used for verification based on descriptive statistics (~35 000 tracks from 2005; more information on this sample is provided below), the sample Pearson correlation coefficient between jitter and duration was found to be −0.06.

Incoherence was assessed for this work using the standard deviation of cluster-maximum radar reflectivity factor ($\sigma_{Z_{max}}$). In contrast to the approach of Lakshmanan and Smith, maximum radar reflectivity factor was used instead of VIL in the calculation of incoherence. The sample Pearson correlation coefficient between incoherence and duration for the ThOR-derived candidate thunderstorm tracks was found to be 0.38. However, by normalizing incoherence by the square root of track duration, the correlation decreases to −0.15. Thus, the descriptive statistics used for verification are track duration, jitter, and normalized incoherence.

For both components of the verification, the ThOR tracks were compared with a benchmark tracking algorithm to provide context for the verification statistics. This benchmark algorithm is essentially a "poor man's" tracking algorithm:

- Tracks are developed by choosing the cluster nearest to, and no farther than 12 km from, the projected location at each time step.
- NARR mean wind is used as the motion estimate at all points along the track.
- Cluster skipping is not allowed.

For all events used for verification, both ThOR and the benchmark algorithm were run on the same set of clusters.

A benchmark tracking algorithm could have been chosen from existing, well-established "storm" identification and tracking algorithms such as SCIT or TITAN. However, an alternative approach was adopted for the following reasons. First, unlike ThOR, both SCIT and TITAN are online tracking algorithms; a "fairer" benchmark should be an offline algorithm. Second, both SCIT and TITAN involve tight integration of cluster identification and tracking. As such, teasing out differences solely attributable to the tracking logic (the focus of the verification) would be difficult. ThOR tracking could be applied to the clusters identified through another algorithm, but this would necessitate retraining ThOR, since cluster characteristics such as density and volume-to-volume behavior are likely to be very different from the characteristics of clusters identified via w2segmotionll. Finally, it is our assertion that neither SCIT nor TITAN nor ThOR should serve as a benchmark to assess tracking algorithm performance. Instead, the ideal benchmark should adopt straightforward logic and be easy to implement. The poor man's tracking algorithm used here satisfies these requirements.

The manually analyzed tracks used for verification come from four events not included in the events used for training. Each event consists of a 2-h window of

TABLE 4. Hits, misses, false alarms (following the contingency table described in Table 2) for the four verification events for which tracks were manually analyzed.

	ThOR	Benchmark
h	1016	975
$m1$	60	27
$m2$	53	63
$m3$	6	75
$m4$	8	9
fa1	39	98
fa2	55	2
fa3	29	16

clusters based on data from a single radar (Table 3). The events are intended to represent different speeds and storm modes but constitute a necessarily small sample, since manually tracking is labor intensive and time consuming. The first step in comparing ThOR tracks to manually analyzed tracks is to pair a manual track to the ThOR track that shares the most clusters. Each point along each track in a pair is cataloged as either a hit, a miss, or a false alarm following the contingency table described in Table 2. "Correct negative" points could be treated as hits but are instead ignored so as not to inflate skill. All points along a ThOR (manually analyzed) track that are not paired with a manually analyzed (ThOR) track are counted as false alarms (misses).

TABLE 5. Skill scores for ThOR and benchmark tracks using manually analyzed tracks as verification.

Method	POD	FAR	CSI
ThOR	0.889	0.108	0.803
Benchmark	0.849	0.106	0.771

Results for ThOR tracks and benchmark tracks for the four events considered are summarized in Table 4.

From the total hits, misses, and false alarms for all events, the probability of detection (POD), the false alarm rate (FAR), and the critical success index (CSI) were computed using the following expressions (refer to Table 2 for nomenclature):

$$\mathrm{POD} = \frac{\sum h}{\sum h + \sum (m1 + m2 + m3 + m4)}$$

$$\mathrm{FAR} = \frac{\sum (fa1 + fa2 + fa3)}{\sum h + \sum (fa1 + fa2 + fa3)}$$

$$\mathrm{CSI} = \frac{\sum h}{\sum h + \sum (fa1 + fa2 + fa3) + \sum (m1 + m2 + m3 + m4)}.$$

Results (Table 5) indicate that ThOR matches the manual tracks slightly better than the benchmark tracks.

As noted above, one of the chief benefits of using descriptive statistics is that they do not rely upon manually analyzed correct tracks and, consequently, can be applied to a much larger sample. However, to determine the reference values of these statistics, they are first applied to the small set of events used to determine manual tracks (Table 6). [As adopted by Lakshmanan and Smith (2010), the mean jitter and mean normalized coherence are calculated only for tracks at or above the median duration.] These results indicate that while the differences in jitter and incoherence between the three tracking methods are not significant (p values from a Student's t test are around 0.3), the median duration of the benchmark tracks is considerably lower than that of the other two. This difference is likely a consequence of the exclusion of cluster skipping and the use of a less sophisticated calculation of cluster motion compared to ThOR tracking. Overall, the descriptive statistics of the ThOR tracks are nearly identical to those of the manual tracks, suggesting good agreement.

When the descriptive statistics are evaluated for the much larger set of 35 370 tracks derived from a range of multiradar events in 2005 (Table 7), the ThOR tracking once again produces statistically significantly longer tracks than the benchmark tracking (Table 8). Both mean jitter and mean normalized incoherence are smaller for the ThOR algorithm, but these differences are not statistically significant based on a Student's t test.

TABLE 6. Descriptive statistics applied to the four events used for manual tracking (Table 3).

Method	Median duration (s)	Mean jitter (km)	Mean normalized incoherence (dBZ s^{-1})	No. of tracks (total, at or above median duration)
Manual	1238	2.62	0.053	167, 89
ThOR	1241	2.64	0.051	166, 89
Benchmark	942	2.41	0.051	184, 95

g. *Lightning association*

Candidate thunderstorms, whose tracks are defined through the tracking algorithm described above, are

TABLE 7. List of events in the large sample of candidate thunderstorm tracks that are used for track verification based on descriptive statistics. These are multiradar events occurring in 2005.

Start	End	Candidate thunderstorm tracks
1253 UTC 1 Jan	0303 UTC 6 Jan	6737
1721 UTC 22 Feb	1304 UTC 25 Feb	2564
2200 UTC 12 Jul	2325 UTC 21 Jul	16 583
1847 UTC 18 Oct	2108 UTC 21 Oct	2618
0405 UTC 29 Oct	0919 UTC 1 Nov	2686
2354 UTC 25 Nov	2117 UTC 28 Nov	4182

TABLE 8. Descriptive statistics applied to the large dataset of tracks identified from events in 2005 (Table 7).

Method	Median duration (s)	Mean jitter (km)	Mean normalized incoherence ($dBZ\,s^{-1}$)	No. of tracks (total; at or above median duration)
ThOR	1191	1.34	0.0407	35 370; 17 118
Benchmark	894	1.35	0.0421	39 392; 19 768

designated as thunderstorms if at least one cloud-to-ground lightning strike is observed ''within'' the bounds of the candidate thunderstorm at any point in its lifetime. Candidate thunderstorm bounds at a particular time are defined using an ellipse whose major axis, minor axis, and orientation are set by the w2segmotionll algorithm to best represent the size and shape of the associated cluster. Because the cluster positions making up the candidate thunderstorm are recorded at radar observation times only (generally every 5 min), while the cloud-to-ground lightning observations in the dataset used for this work have a time interval of 1 min, and because ellipse boundaries will evolve between the MCZ data times, candidate thunderstorm bounds are temporally interpolated to the lightning observation times. This is accomplished through linear interpolation of the four primary attributes of the representative ellipse (centroid position, major axis length, minor axis length, and orientation). Any candidate thunderstorm with at least one cloud-to-ground lightning strike within any cluster ellipse along its track will be designated a thunderstorm.

3. 23–28 April 2007 case

A qualitative evaluation of ThOR is presented in this section through analysis of thunderstorms identified during a multiday event in April 2007. This event (1525 UTC 23 April to 0936 UTC 28 April)[4] spanned most of the U.S. central and southern plains (Fig. 6), and was characterized by a variety of convective organization types including supercells, disorganized multicells, and quasi-linear systems. The merged composite reflectivity was derived on a Cartesian grid with a grid spacing of 0.014° latitude × 0.011° longitude (approximately 1 km × 1 km).

A total of 3151 thunderstorms, consisting of 27 394 reflectivity clusters, were identified. The event was composed of 8314 candidate thunderstorms; thus, only 38% of candidate thunderstorms were associated with CG flashes. On average, 52.9 CG flashes were associated with a given thunderstorm during this event, equating to 1.16 flashes per minute. Mean storm duration was 41.5 min, and the mean storm size was 339 km^2 (which corresponds to a diameter of 20.8 km for a circular thunderstorm).

As expected, nonthunderstorm tracks (illustrated in blue in Fig. 7) are generally relegated to northern latitudes and the ''back side'' of the synoptic-scale system, where convective available potential energy was ostensibly insufficient to support CG lightning. The absence of IC lightning data may have also led to some missed identifications in this region. Clusters of observed thunderstorms tend to exhibit similar motion vectors, which should also be expected. There are few if any clusters that, by virtue of their high radar reflectivity, should have been associated with a thunderstorm track but are omitted by ThOR. There are several examples of quasi-linear segments that have been consolidated into a single thunderstorm (e.g., located in eastern Oklahoma, western Arkansas, and southwestern Missouri in Fig. 7). This result is consistent with the definition of a thunderstorm that underpins the logic of the ThOR algorithm but is also a consequence of any centroid-based method of thunderstorm identification that requires that even spatially expansive clusters must be represented as a single point for the purposes of cataloging and tracking.

Closer examination of observed CG lightning and thunderstorm tracks (Fig. 6) reveals instances of CG lightning that, because of its spatial separation from thunderstorm tracks, has not been associated with thunderstorms. Lightning not associated with a ThOR-identified thunderstorm is either a consequence of the algorithm's failure to catch small and/or low-reflectivity clusters that are producing CG lightning, lightning producing stratiform regions, or spurious lightning observations. Three representative examples of the first two causes are illustrated in Fig. 8.

In Fig. 8a, CG lightning associated with the stratiform precipitation region of a mesoscale convective system is allowed to occur without thunderstorm identification. Consistent with the definition of a thunderstorm adopted here (a thundering cumulonimbus cloud), it is assumed that CG-producing stratiform precipitation is a part of the same thundering cumulonimbus cloud whose most direct manifestation is the (smaller scale)

[4] Level II WSR-88D data and NARR data were downloaded from the National Climatic Data Center (NCDC), while the CG flashes from the NLDN were downloaded from Iowa State's Internet Data Distribution archive.

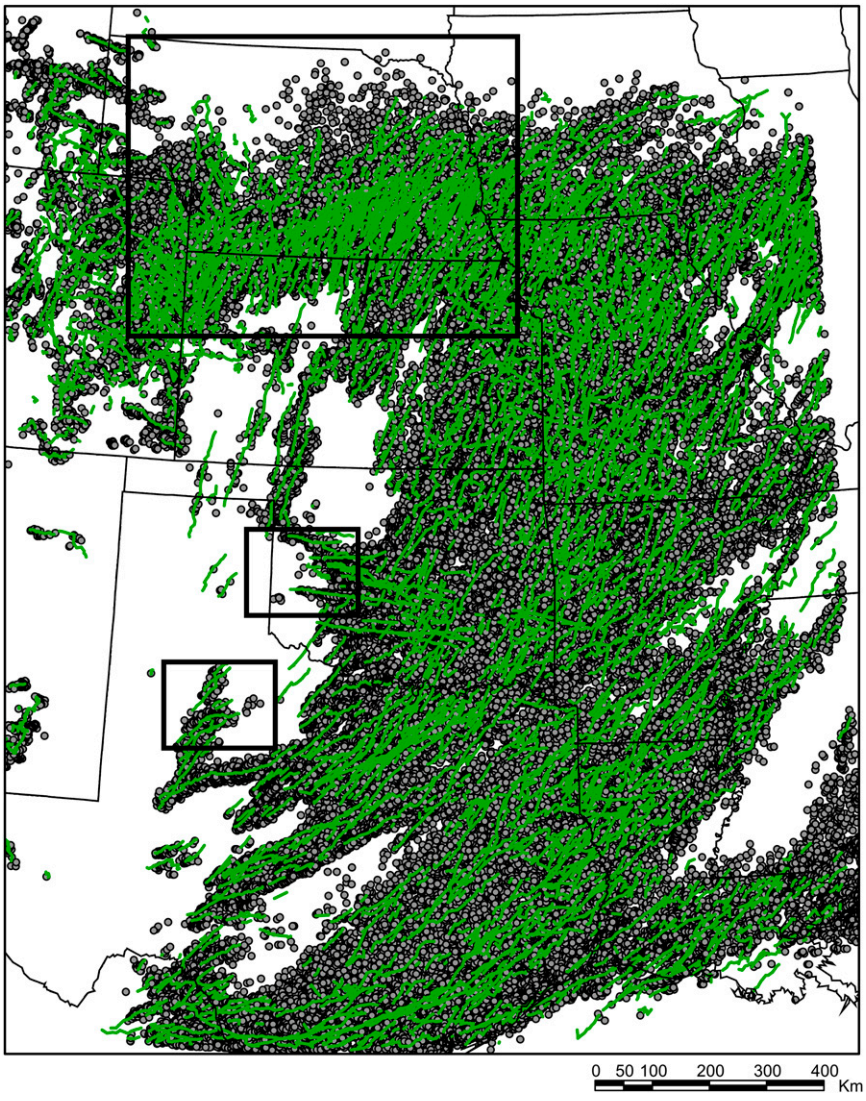

FIG. 6. CG lightning flashes (gray circles) and thunderstorm tracks for the 23–28 Apr 2007 case. The three inset boxes are the regions illustrated in Fig. 8.

convective regions. As discussed in section 2d, this filtering is enabled in ThOR through attenuation of stratiform precipitation. Another illustration in which a cluster of CG flashes within stratiform precipitation is not associated with a distinct thunderstorm appears in Fig. 8b. The spatial separation between the convective and stratiform regions is much smaller (~50 km) in this example than in the prior example, but ThOR is still able to correctly recognize that they should not be associated with a distinct thunderstorm.

The final example of CG flashes that have not been associated with thunderstorm tracks reveals a failure of ThOR to identify a small thunderstorm (Fig. 8c). As discussed in section 2e, the saliency criterion utilized in the ThOR implementation of w2segmotionll is set to 50 km^2. The thunderstorm near the southwest corner of the Fig. 8c has a precipitation footprint of ~40–45 km^2 and yet should have been cataloged as a thunderstorm given the occurrence of nearby CG lightning. As addressed in section 2e, the saliency criterion was set to minimize the number of small clusters that degrade track accuracy and increase the algorithm's computational burden. Missed thunderstorms are an inevitable consequence of thresholds used in ThOR and any other automated thunderstorm identification algorithm. Analysis has revealed few obvious examples of thunderstorms missed by ThOR. Those that are missed are deemed to be a justified collateral sacrifice for a consistent/objective approach that can be applied to much larger datasets than possible with manual analysis.

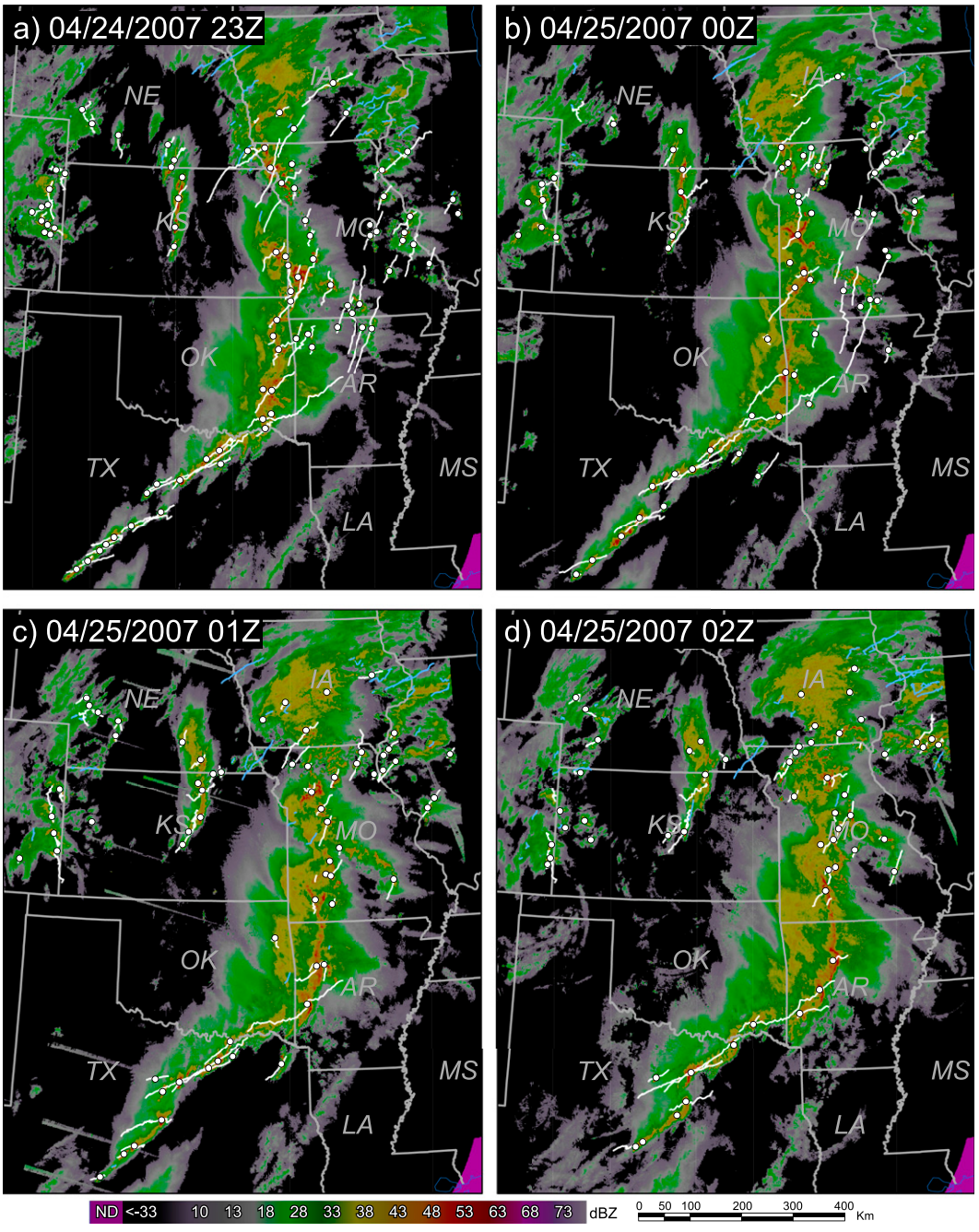

FIG. 7. Example tracks from the 23–28 Apr 2007 case used for qualitative assessment of ThOR. Radar images are the merged composite reflectivity, thunderstorm tracks that exist at the time of the radar image are in white, non-thunderstorm tracks that exist at the time of the radar image are in blue, and thunderstorm clusters positions at the time of the radar image are indicated with white circles. Thunderstorm tracks without a corresponding cluster centroid are a consequence of the allowances ThOR makes for skipped clusters.

The comparison between ThOR and the benchmark tracking algorithm [section 2f(2)] serves as the principal means of verifying the performance of the tracking component in ThOR. To complement this evaluation, the behavior of ThOR is compared to the behavior of SCIT for a subset of the 23–28 April 2007 case. The intent is not to use this comparison as a means of evaluating the accuracy of either scheme. Instead, the comparison is made to offer a mainly qualitative comparison of ThOR to a more familiar identification and tracking algorithm.

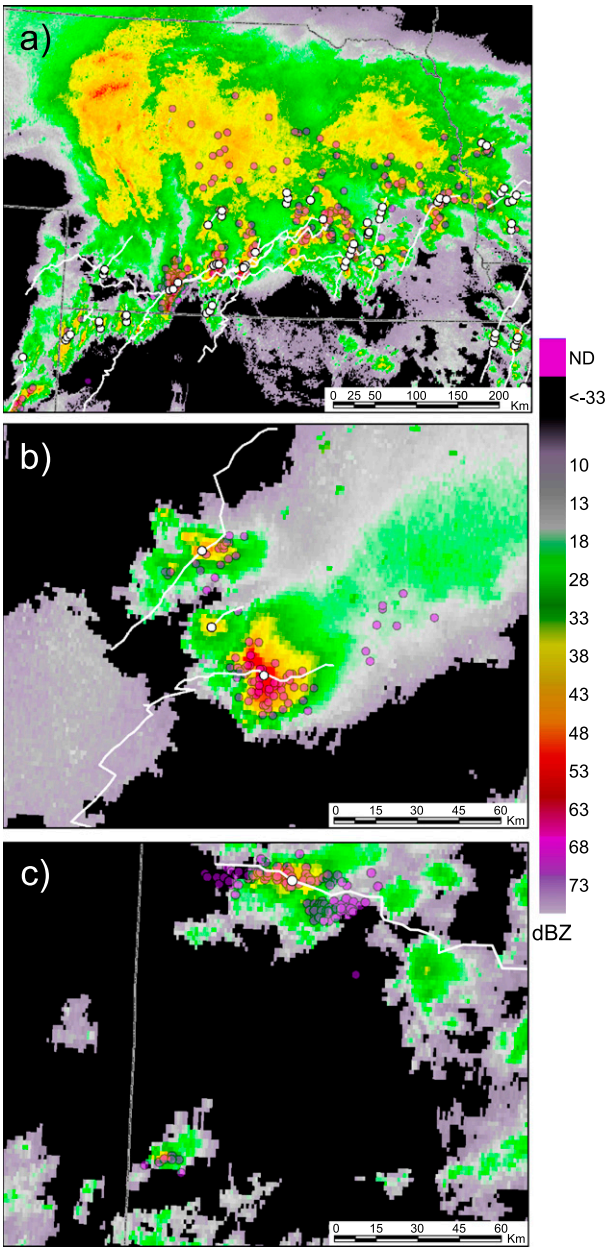

FIG. 8. Examples of CG lightning not associated with ThOR-identified thunderstorms from the 23–28 Apr 2007 case used for qualitative assessment of ThOR. Merged composite reflectivity serves as the radar images, thunderstorm tracks that exist within ±5 min of the time of the radar image are in white, and the positions of CG lightning flashes are indicated with semitransparent purple circles. (a) Region indicated by the northernmost box in Fig. 6: indicated lightning flashes occurred in a 10-min time window centered at 0830 UTC 24 April, and all cluster positions within the time window are indicated with white circles. (b) Region indicated by the southernmost box in Fig. 6: indicated lightning flashes occurred in a 30-min time window centered at 0345 UTC 24 April, and cluster positions at the time of the radar image are indicated with white circles. (c) Region indicated by the middle box in Fig. 6: indicated lightning flashes occurred in a 30-min time window centered at 1335 UTC 27 April, and cluster positions at the time of the radar image are indicated with white circles.

The version of SCIT integrated into the Severe Storms Analysis Package of the WDSS-II suite was run for a single radar (Fort Worth, Texas) for the period 1800 UTC 24 April–00600 UTC 25 April 2007 and compared to the *candidate* thunderstorm tracks identified by ThOR for the same period. As reflected in Fig. 9, SCIT tracks are far more numerous than ThOR tracks: there are 1058 SCIT tracks and only 358 ThOR tracks. The difference appears to depend less on the tracking differences (though, as noted above, this comparison is not capable of determining what role is played by differences in tracking logic) and more on the differences in cluster identification. As illustrated in Fig. 10, there are 3–4 times as many SCIT clusters as ThOR clusters.

4. Summary

The ThOR algorithm is an objective and tunable Lagrangian approach to cataloging thunderstorms that uses surveillance radar data and NARR mean winds to identify and track radar reflectivity clusters that are used to represent candidate thunderstorms. Using observed cloud-to-ground lightning, thunderstorm tracks are identified. Unlike Eulerian methods for developing thunderstorm *event* climatologies, ThOR is capable of cataloging nearly every thunderstorm that occurs over the spatial domain and period of record, thereby enabling analysis of internal properties of thunderstorms.

ThOR involves 1) merging radar data from multiple radars to a single volume on a common Cartesian grid, 2) attenuating stratiform precipitation to improve thunderstorm detection, 3) identifying spatially coherent regions in radar reflectivity (i.e., clusters), 4) tracking clusters to develop candidate thunderstorms, and 5) associating cloud-to-ground (CG) lightning to candidate thunderstorms to classify tracked clusters as thunderstorms.

Training and verification of the cluster tracking component of ThOR were discussed. Both training and verification rely on comparison of ThOR tracks to manually analyzed tracks. Additionally, the verification step involved descriptive statistics applied to a sample of tracks much larger than reasonably possible using manually analyzed tracks. The verification step also relied on a benchmark tracking algorithm that is essentially a ‘‘poor man’s’’ approach to tracking and provides context for the verification statistics.

Tracking verification revealed that ThOR tracks matched the manual tracks slightly better than benchmark tracks. Moreover, the descriptive statistics of the ThOR tracks are nearly identical to those of the manual

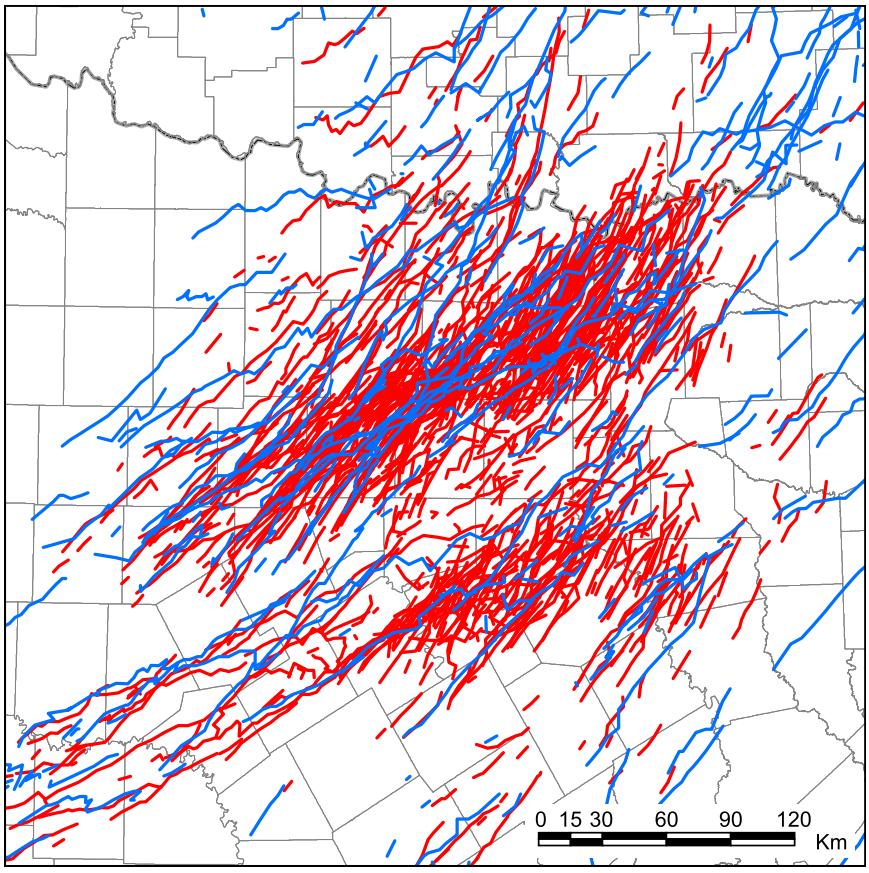

FIG. 9. Comparison of SCIT tracks (red curves) and ThOR tracks (blue curves) for clusters occurring in April 2007 between 1800 UTC 24 Apr and 0600 UTC 25 Apr.

tracks, suggesting good agreement. When the descriptive statistics were applied to a ~35 000-track dataset, ThOR tracking produces statistically significantly longer tracks than the benchmark tracking and smaller mean jitter and mean normalized incoherence, though the latter differences from the benchmark values were not statistically significant.

A largely qualitative examination of ThOR performance was performed using a multiday event in April 2007. ThOR-identified thunderstorms were primarily located in the warm sector, as should be expected, and clusters of thunderstorms tended to exhibit similar motion vectors. Few if any clusters, by virtue of their high radar reflectivity, should have been associated to a thunderstorm track but were omitted by ThOR.

Analysis of the April 2007 event also focused on examples of CG flashes that were not associated with thunderstorm tracks. These unassociated flashes tended to occur in regions of stratiform precipitation that the algorithm assumes to be part of the same thundering cumulonimbus cloud whose most direct manifestation is the (smaller scale) convective regions. An example was also provided of a legitimate thunderstorm that, by virtue of its small scale, failed to be detected by ThOR. Such misses appear to be rare but are an expected trade-off for a consistent/objective approach that can be applied to much larger datasets than possible with manual analysis.

Acknowledgments. This work was funded by NSF Grant AGS-0757189 and Air Force Office of Scientific Research Grant FA9550-12-1-0412, and it utilized computational resources at the Holland Computing Center (HCC) at the University of Nebraska and the Open Science Grid. The authors wish to thank the three anonymous reviewers, whose comments greatly improved this article. The authors are grateful to the HCC staff, particularly Dr. Adam Caprez, for guidance in optimizing ThOR for the HCC systems; and the Open Science Grid. We would also like to thank Alexander Gibbs for his assistance in the development of ThOR and Dr. Valliappa Lakshmanan for assistance with WDSS-II.

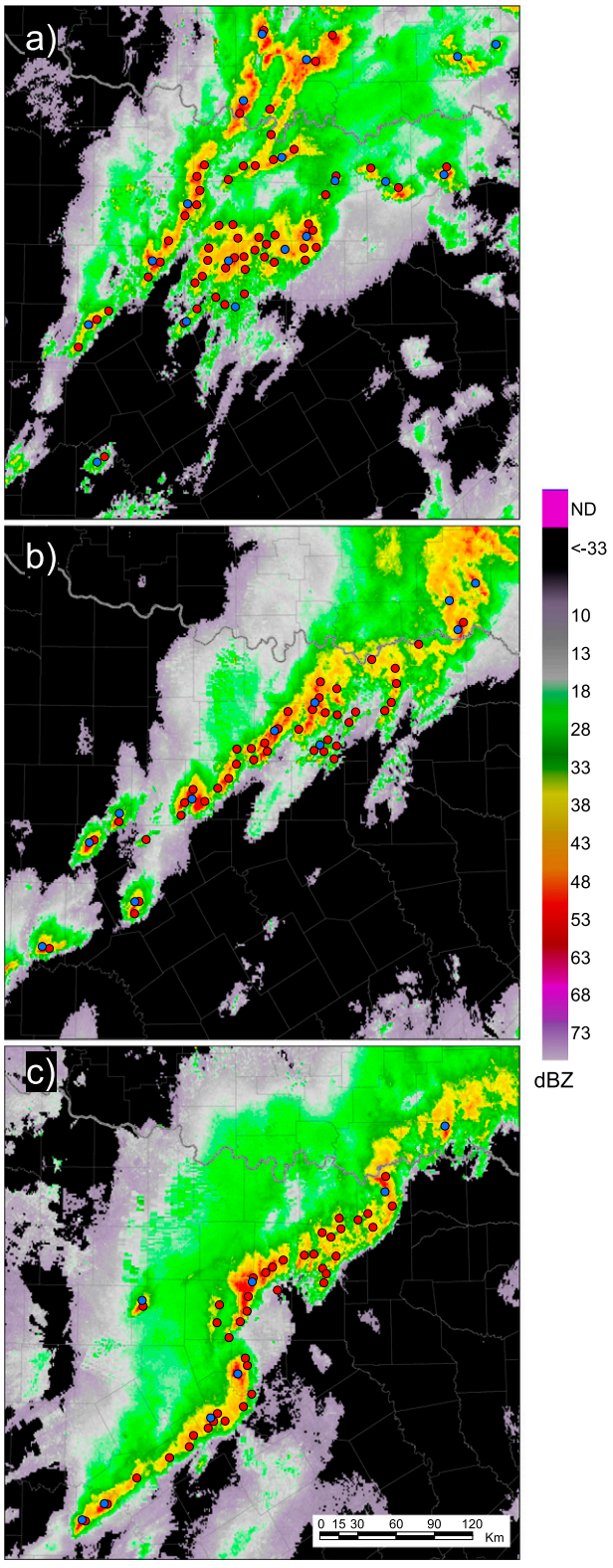

FIG. 10. Comparison of SCIT clusters (red circles) and ThOR clusters (blue circles) for 24–25 Apr 2007 at (a) 2000, (b) 2300, and (c) 0200 UTC. Merged composite reflectivity serves as the radar images.

APPENDIX

Stratiform–Convective Scoring

For each column on the common Cartesian grid, the stratiform–convective score (S), ranging from 0 for definitely stratiform to 1 for definitely convective, is calculated by first setting S to 0 for any column with a vertical reflectivity gradient (calculated between the height of the maximum reflectivity and 3 km above this level) that is $\geq 5\,\mathrm{dBZ\,km^{-1}}$. Any column with a vertical reflectivity gradient $<5\,\mathrm{dBZ\,km^{-1}}$ and either an MCZ $\geq 50\,\mathrm{dBZ}$ or a horizontal gradient of MCZ $\geq 6\,\mathrm{dBZ\,km^{-1}}$ is given a score of 1. All other columns are assigned a score calculated using

$$S = 0.2S_H + 0.4S_V + 0.4S_{H_{\max+3}},$$

where, S_H, S_V, and $S_{H_{\max+3}}$ are the component scores for the magnitude of the horizontal gradient of MCZ, the vertical gradient of reflectivity, and the horizontal reflectivity gradient at a height 3 km above the level of maximum reflectivity, respectively. Term S_H is set to 0 if the horizontal gradient of MCZ is $\leq 1\,\mathrm{dBZ\,km^{-1}}$, 1 if the horizontal gradient is $\geq 3\,\mathrm{dBZ\,km^{-1}}$, and linearly interpolated between 0 and 1 for horizontal gradient values between 1 and $3\,\mathrm{dBZ\,km^{-1}}$. Term S_V is set to 0 if the vertical gradient is $\geq 4\,\mathrm{dBZ\,km^{-1}}$, 1 if the vertical gradient is $\leq 1\,\mathrm{dBZ\,km^{-1}}$, and linearly interpolated between 1 and 0 for vertical gradient values between 1 and $4\,\mathrm{dBZ\,km^{-1}}$. Term $S_{H_{\max+3}}$ is set following the logic used to set S_H. The resultant field of S is then smoothed (to account for issues with the radar's sampling).

REFERENCES

Alexander, W. H., 1915: Distribution of thunderstorms in the United States. *Mon. Wea. Rev.*, **43,** 322–340, doi:10.1175/1520-0493(1915)43<322:DOTITU>2.0.CO;2.

——, 1935: The distribution of thunderstorms in the United States, 1904–33. *Mon. Wea. Rev.*, **63,** 157–158, doi:10.1175/1520-0493(1935)63<157:TDOTIT>2.0.CO;2.

American Meteorological Society, 2013: Thunderstorm. Glossary of Meteorology. [Available online at http://glossary.ametsoc.org/wiki/Thunderstorm.]

Biagi, C. J., K. L. Cummins, K. E. Kehoe, and E. P. Krider, 2007: National Lightning Detection Network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003–2004. *J. Geophys. Res.*, **112,** D05208, doi:10.1029/2006JD007341.

Biggerstaff, M. I., and S. A. Listemaa, 2000: An improved scheme for convective/stratiform echo classification using radar reflectivity. *J. Appl. Meteor.*, **39,** 2129–2150, doi:10.1175/1520-0450(2001)040<2129:AISFCS>2.0.CO;2.

Changnon, S. A., 1985: Secular variations in thunder-day frequencies in the twentieth century. *J. Geophys. Res.*, **90,** 6181–6194, doi:10.1029/JD090iD04p06181.

——, 1988a: Climatography of thunder events in the conterminous United States. Part I: Temporal aspects. *J. Climate,* **1,** 389–398, doi:10.1175/1520-0442(1988)001<0389:COTEIT>2.0.CO;2.

——, 1988b: Climatography of thunder events in the conterminous United States. Part II: Spatial aspects. *J. Climate,* **1,** 399–405, doi:10.1175/1520-0442(1988)001<0399:COTEIT>2.0.CO;2.

Clements, N. C., and R. E. Orville, 2008: The warning time for cloud-to-ground lightning in isolated, ordinary thunderstorms over Houston, Texas. *Third Conf. on Meteorological Applications of Lightning Data,* New Orleans, LA, Amer. Meteor. Soc., 6.4. [Available online at https://ams.confex.com/ams/88Annual/techprogram/paper_132309.htm.]

Cotton, W. R., and R. A. Anthes, 1989: *Storm and Cloud Dynamics.* International Geophysics Series, Vol. 44, Academic Press, 883 pp.

Court, A., and J. F. Griffiths, 1981: Thunderstorm climatology. *Thunderstorm Morphology and Dynamics,* University of Oklahoma Press, 9–39.

Davini, P., R. Bechini, R. Cremonini, and C. Cassardo, 2012: Radar-based analysis of convective storms over northwestern Italy. *Atmosphere,* **3,** 33–58, doi:10.3390/atmos3010033.

Dixon, M., and G. Wiener, 1993: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A radar-based methodology. *J. Atmos. Oceanic Technol.,* **10,** 785–797, doi:10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2.

Duda, J. D., and W. A. Gallus Jr., 2010: Spring and summer midwestern severe weather reports in supercells compared to other morphologies. *Wea. Forecasting,* **25,** 190–206, doi:10.1175/2009WAF2222338.1.

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

Falconer, P. D., 1984: A radar-based climatology of thunderstorm days across New York State. *J. Climate Appl. Meteor.,* **23,** 1115–1120, doi:10.1175/1520-0450(1984)023<1115:ARBCOT>2.0.CO;2.

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, doi:10.1175/2007WAF2006120.1.

Geerts, B., 1998: Mesoscale convective systems in the Southeast United States during 1994–95: A survey. *Wea. Forecasting,* **13,** 860–869, doi:10.1175/1520-0434(1998)013<0860:MCSITS>2.0.CO;2.

Goudenhoofdt, E., and L. Delobbe, 2013: Statistical characteristics of convective storms in Belgium derived from volumetric weather radar observations. *J. Appl. Meteor. Climatol.,* **52,** 918–934, doi:10.1175/JAMC-D-12-079.1.

Hocker, J. E., and J. B. Basara, 2008: A geographic information systems–based analysis of supercells across Oklahoma from 1994 to 2003. *J. Appl. Meteor. Climatol.,* **47,** 1518–1538, doi:10.1175/2007JAMC1673.1.

Houze, R. A., 1993: *Cloud Dynamics.* Academic Press, 573 pp.

Johnson, J. T., P. L. MacKeen, A. Witt, E. D. W. Mitchell, G. J. 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, doi:10.1175/1520-0434(1998)013<0263:TSCIAT>2.0.CO;2.

Kolodziej Hobson, A. G. K., V. Lakshmanan, T. M. Smith, and M. Richman, 2012: An automated technique to categorize storm type from radar and near-storm environment data. *Atmos. Res.,* **111,** 104–113, doi:10.1016/j.atmosres.2012.03.004.

Kuo, J.-T., and H. D. Orville, 1973: A radar climatology of summertime convective clouds in the Black Hills. *J. Appl. Meteor.,* **12,** 359–368, doi:10.1175/1520-0450(1973)012<0359:ARCOSC>2.0.CO;2.

Lakshmanan, V., and T. Smith, 2009: Data mining storm attributes from spatial grids. *J. Atmos. Oceanic Technol.,* **26,** 2353–2365, doi:10.1175/2009JTECHA1257.1.

——, and ——, 2010: An objective method of evaluating and devising storm-tracking algorithms. *Wea. Forecasting,* **25,** 701–709, doi:10.1175/2009WAF2222330.1.

——, V. DeBrunner, and R. Rabin, 2002: Nested partitions using texture segmentation. *SSIAI '02: Proceedings of the Fifth IEEE Southwest Symposium on Image Analysis and Interpretation,* IEEE Computer Society, 153–157.

——, T. Smith, K. Hondl, G. J. Stumpf, and A. Witt, 2006: A real-time, three-dimensional, rapidly updating, heterogeneous radar merger technique for reflectivity, velocity, and derived products. *Wea. Forecasting,* **21,** 802–823, doi:10.1175/WAF942.1.

——, ——, G. Stumpf, and K. Hondl, 2007: The Warning Decision Support System–Integrated Information. *Wea. Forecasting,* **22,** 596–612, doi:10.1175/WAF1009.1.

——, 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, doi:10.1175/2008JTECHA1153.1.

López, R. E., D. O. Blanchard, D. Rosenfeld, W. L. Hiscox, and M. J. Casey, 1984: Population characteristics, development processes and structure of radar echoes in south Florida. *Mon. Wea. Rev.,* **112,** 56–75, doi:10.1175/1520-0493(1984)112<0056:PCDPAS>2.0.CO;2.

MacGorman, D. R., I. R. Apostolakopoulos, N. R. Lund, N. W. S. Demetriades, M. J. Murphy, and P. R. Krehbiel, 2011: The timing of cloud-to-ground lightning relative to total lightning activity. *Mon. Wea. Rev.,* **139,** 3871–3886, doi:10.1175/MWR-D-11-00047.1.

MacKeen, P. L., H. E. Brooks, and K. L. Elmore, 1999: Radar reflectivity–derived thunderstorm parameters applied to storm longevity forecasting. *Wea. Forecasting,* **14,** 289–295, doi:10.1175/1520-0434(1999)014<0289:RRDTPA>2.0.CO;2.

May, P. T., and A. Ballinger, 2007: The statistical characteristics of convective cells in a monsoon regime (Darwin, northern Australia). *Mon. Wea. Rev.,* **135,** 82–92, doi:10.1175/MWR3273.1.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.,* **87,** 343–360, doi:10.1175/BAMS-87-3-343.

Michaels, P. J., R. A. Pielke, J. T. McQueen, and D. E. Sappington, 1987: Composite climatology of Florida summer thunderstorms. *Mon. Wea. Rev.,* **115,** 2781–2791, doi:10.1175/1520-0493(1987)115<2781:CCOFST>2.0.CO;2.

Mohee, F. M., and C. Miller, 2010: Climatology of thunderstorms for North Dakota, 2002–06. *J. Appl. Meteor. Climatol.,* **49,** 1881–1890, doi:10.1175/2010JAMC2400.1.

Mosier, R. M., C. Schumacher, R. E. Orville, and L. D. Carey, 2011: Radar nowcasting of cloud-to-ground lightning over Houston, Texas. *Wea. Forecasting,* **26,** 199–212, doi:10.1175/2010WAF2222431.1.

NOAA, 2013: Natural hazard statistics. Accessed 25 October 2014. [Available online at http://www.nws.noaa.gov/om/hazstats.shtml.]

Parker, M. D., and R. H. Johnson, 2000: Organizational modes of midlatitude mesoscale convective systems. *Mon. Wea. Rev.,* **128,** 3413–3436, doi:10.1175/1520-0493(2001)129<3413:OMOMMC>2.0.CO;2.

Potts, R. J., T. D. Keenan, and P. T. May, 2000: Radar characteristics of storms in the Sydney area. *Mon. Wea. Rev.,* **128,** 3308–3319, doi:10.1175/1520-0493(2000)128<3308:RCOSIT>2.0.CO;2.

Rakov, V. A., 2013: Electromagnetic methods of lightning detection. *Surv. Geophys.,* **34,** 731–753, doi:10.1007/s10712-013-9251-1.

Reap, R. M., and D. S. Foster, 1979: Automated 12–36 hour probability forecasts of thunderstorms and severe local storms. *J. Appl. Meteor.,* **18,** 1304–1315, doi:10.1175/1520-0450(1979)018<1304:AHPFOT>2.0.CO;2.

Reed, J., and J. Trostel, 2012: Evaluation of an improved Storm Cell Identification and Tracking (SCIT) algorithm based on DBSCAN clustering and JPDA tracking methods. *28th Conf. on Interactive Information Processing Systems (IIPS),* New Orleans, LA, Amer. Meteor. Soc., 4B.3. [Available online at https://ams.confex.com/ams/92Annual/webprogram/Paper201783.html.]

Root, B., M. B. Yeary, and T. Y. Yu, 2011: Novel storm cell tracking with multiple hypothesis tracking. *27th Conf. on Interactive Information Processing Systems (IIPS),* Seattle, WA, Amer. Meteor. Soc., 8B.3. [Available online at https://ams.confex.com/ams/91Annual/webprogram/Paper184250.html.]

Scharenbroich, L., G. Magnusdottir, P. Smyth, H. Stern, and C.-c. Wang, 2010: A Bayesian framework for storm tracking using a hidden-state representation. *Mon. Wea. Rev.,* **138,** 2132–2148, doi:10.1175/2009MWR2944.1.

Seroka, G. N., R. E. Orville, and C. Schumacher, 2012: Radar nowcasting of total lightning over the Kennedy Space Center. *Wea. Forecasting,* **27,** 189–204, doi:10.1175/WAF-D-11-00035.1.

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, doi:10.1175/WAF-D-11-00115.1.

Steiner, M., R. A. Houze, 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, doi:10.1175/1520-0450(1995)034<1978:CCOTDS>2.0.CO;2.

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, doi:10.1175/1520-0434(1998)013<0304:TNSSLM>2.0.CO;2.

Wallace, J. M., 1975: Diurnal variations in precipitation and thunderstorm frequency over the conterminous United States. *Mon. Wea. Rev.,* **103,** 406–419, doi:10.1175/1520-0493(1975)103<0406:DVIPAT>2.0.CO;2.

Wiggert, V., G. J. Lockett, and S. S. Ostlund, 1981: Radar rainshower growth histories and variations with wind speed, echo motion, location and merger status. *Mon. Wea. Rev.,* **109,** 1467–1494, doi:10.1175/1520-0493(1981)109<1467:RRGHAV>2.0.CO;2.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* 3rd ed. Elsevier, 676 pp.

Zipser, E. J., and K. R. Lutz, 1994: The vertical profile of radar reflectivity of convective cells: A strong indicator of storm intensity and lightning probability? *Mon. Wea. Rev.,* **122,** 1751–1759, doi:10.1175/1520-0493(1994)122<1751:TVPORR>2.0.CO;2.