A Method for Identifying Midlatitude Mesoscale Convective Systems in Radar Mosaics. Part II: Tracking

ALEX M. HABERLIE[a] AND WALKER S. ASHLEY

Department of Geographic and Atmospheric Sciences, Northern Illinois University, DeKalb, Illinois

(Manuscript received 15 October 2017, in final form 10 April 2018)

ABSTRACT

This research is Part II of a two-part study that evaluates the ability of image-processing and select machine-learning algorithms to detect, classify, and track midlatitude mesoscale convective systems (MCSs) in radar-reflectivity images for the conterminous United States. This paper focuses on the tracking portion of this framework. Tracking is completed through a two-step process using slice (snapshots of instantaneous MCS intensity) data generated in Part I. The first step is to perform spatiotemporal matching, which associates slices through temporally adjacent radar-reflectivity images to generate swaths, or storm tracks. When multiple slices are found to be matches, a difference-minimization procedure is used to associate the most similar slice with the existing swath. Once this step is completed, a second step combines swaths that are spatiotemporally close. Tracking performance is assessed by calculating select metrics for all available swath-building perturbations to determine the optimal approach in tracking. Frequency maps and time series generated from the swaths suggest that the spatiotemporal occurrence of these swaths is reasonable as determined from previous work. Further, these events exhibit a diurnal cycle that is distinct from that of overall convection for the conterminous United States. Last, machine-learning predictions are found to limit areas of high MCS frequency to the central and eastern Great Plains.

1. Introduction

Mesoscale convective systems (MCSs) are thought to produce a significant portion of warm-season precipitation for many regions in the conterminous United States (CONUS) (Zipser 1982; Ashley et al. 2003; Houze 2004). Because of this, MCSs have been, and continue to be, a popular focus for research in the fields of hydrology, climatology, and meteorology (Houze 2004). To assess objectively the spatiotemporal frequency of MCSs and their precipitation, extensive remotely sensed datasets have been analyzed to find events that meet size, intensity, and duration criteria (Parker and Johnson 2000, henceforth PJ00). In specific terms, PJ00 defined MCSs as areas of deep, moist convection (DMC) organized at the mesoscale (e.g., a horizontal extent of at least 100 km) that last at least 3 h. Translating this dynamically based definition of an MCS into an automated detection and tracking process is crucial because of the large size of remotely sensed datasets (Lakshmanan and Smith 2010). The segmentation, classification, and tracking of phenomena driven by DMC remains a challenging problem (Lakshmanan et al. 2009).

The process of objectively characterizing the frequency of MCS events requires the spatiotemporal association (''matching'') of qualifying precipitation clusters (e.g., Haberlie and Ashley 2018, hereinafter Part I) between temporally adjacent radar images (''storm tracking''). This complex, but necessary, step is complicated by the erratic evolution of precipitation clusters (Lakshmanan and Smith 2010): clusters can undergo many unpredictable changes between radar images that can complicate matching decisions; these changes include initiation, splitting, merging, and decay (e.g., Fig. 2 in Vila et al. 2008). This study explores the utility of using machine-learning predictions to reduce the complexity of the matching step by removing cases that meet the PJ00 criteria for size and intensity but are not labeled as MCS by an ensemble of machine-learning algorithms trained and validated using hand-labeled data (Part I). Further, this work examines the impact of segmentation-threshold values on resulting tracks to determine the potential effects on automated MCS ''climatologies.''

---

[a] Current affiliation: Department of Geography and Anthropology, Louisiana State University, Baton Rouge, Louisiana.

Corresponding author: Alex M. Haberlie, ahaberlie1@lsu.edu

DOI: 10.1175/JAMC-D-17-0294.1

The discussion herein will focus on the tracking portion of an MCS segmentation, classification, and tracking framework introduced in Part I, as well as examples of applying the framework. The main contributions of this paper include 1) an objective and subjective assessment of the effect of segmentation parameters and probabilistic classifications on MCS tracking performance, 2) a demonstration of the influence of segmentation parameters and probabilistic classification thresholds on the spatial frequency and statistical attributes of MCS events, and 3) a statistical description of a novel, manually labeled, dataset of radar-derived MCS events from the 2015 and 2016 warm seasons (May–September). An important finding discussed in Part I is that machine learning can be used to generate reliable classification probabilities for detected convective clusters in composite reflectivity images. To be specific, evidence is presented that suggests select machine-learning algorithms can probabilistically distinguish between MCS and non-MCS (i.e., tropical systems, synoptic systems, and unorganized clusters) convective precipitation areas. This work builds on those findings by illustrating and discussing the utility of these predictions by testing various probability thresholds to balance the removal of false-positive clusters with the inclusion of true-positive cases. Although the application of this approach may be limited for general-purpose storm-tracking algorithms, the complexity of identifying and tracking specific meteorological phenomena can be reduced by using this framework to limit the influence of false-positive events on climatologies, case studies, and other products. Case studies for select events, as well as 2015 and 2016 warm-season MCS frequency maps, are used to demonstrate the utility of the framework.

2. Background

Using remotely sensed data, previous research has either implicitly or explicitly tracked MCS occurrence (Fritsch and Forbes 2001; Houze 2004). Implicit approaches use aggregate rainfall products (stage-IV mosaics, Hovmöller diagrams, etc.) to find contiguous ‘‘precipitation objects’’ (e.g., Davis et al. 2006) of sufficient width and duration (Carbone et al. 2002; Hitchens et al. 2012; Pinto et al. 2015). Explicit tracking, on the other hand, extracts contiguous precipitation clusters from each image (e.g., every 15 min), with the added complexity of associating these clusters through time. Although this approach is more computationally expensive, it allows for a more rigorous examination of MCSs at fine temporal scales and is analogous to more-formal definitions of MCSs. Despite the opportunities that automated methods provide, explicit storm-tracking procedures have known issues with identifying (segmentation) and associating precipitation clusters between time steps (tracking), especially when handling splitting and merging events (Lakshmanan and Smith 2010). Workable solutions to these issues exist in the form of tweaking detection and tracking parameters to correct poor tracking behavior on the basis of case studies or summary statistics (Lakshmanan and Smith 2010).

During an MCS-tracking process, slices—instantaneous snapshots of the geographic distribution of contiguous regions of precipitation—and swaths—the progression of slices over time—are generated (Fig. 1). For a slice to be considered for the swath-building process (i.e., a candidate MCS slice), it must contain a region of contiguous or semicontiguous convective precipitation ($\geq 40\,\mathrm{dB}Z$) with a horizontal dimension exceeding 100 km (PJ00; Part I). These regions, defined as MCS cores (e.g., label i in Fig. 1), are generated by spatially aggregating convective cells that contain intense precipitation ($\geq 50\,\mathrm{dB}Z$) that are within a given distance of one another. Nearby areas of precipitation are associated with MCS cores to generate candidate MCS slices (e.g., label ii in Fig. 1). These areas of convection and stratiform precipitation from temporally adjacent radar images are then spatiotemporally associated to generate swaths (label iii in Fig. 1). This study further restricts the MCS-slice detection (and ultimately the swath building) by using probabilistic machine-learning predictions (P_{MCS}; see Part I) to remove certain candidate MCS slices. This is done using an ensemble classifier, containing trained random-forest (Breiman 2001; scikit-learn 0.18 software, Pedregosa et al. 2011), gradient-boosting (scikit-learn 0.18; Pedregosa et al. 2011), and XGBoost (xgboost-python 0.6 software; Chen and Guestrin 2016) classifiers, to predict the likelihood that each detected slice is a candidate MCS slice. For a detailed explanation of how these classifiers were generated and tested, see Part I. Examples of candidate MCS slices that are likely to have a low P_{MCS} are tropical systems, synoptic systems with embedded convection, and unorganized convective clusters (Part I).

There are several ways to track storm cells in a climatological context. Two widely used approaches are centroid matching (Lakshmanan et al. 2015) and spatiotemporal object building [Skok et al. (2009); Method for Object-Based Diagnostic Evaluation–time domain (MODE-TD); Clark et al. (2014)]. Both procedures use the concept of postevent tracking (Lakshmanan et al. 2015), which, in contrast to real-time storm tracking [e.g., Storm Cell Identification and Tracking (SCIT); Johnson et al. 1998], uses full event histories within the climatological record to generate more accurate storm tracks (e.g., Fig. 6 of Lakshmanan et al. 2015). Spatiotemporal object building is generally used for objects at the scale of MCSs (Clark et al. 2014), whereas centroid matching is generally used for tracking objects on the scale of supercells (Gagne et al. 2017). One disadvantage to spatiotemporal object building

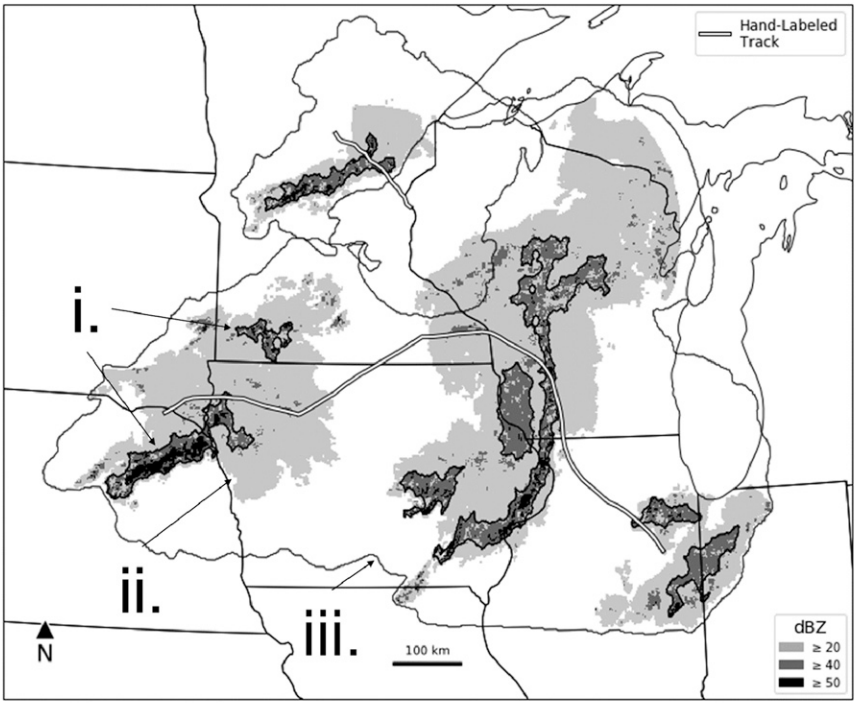

FIG. 1. Manually generated MCS swaths from 0000 to 1700 UTC 7 Jun 2015. The labeled areas include MCS cores (label i; heavy black outlines), MCS slices (label ii; shaded fill), and MCS swaths (label iii; thin gray outlines). Included are centroid tracks (2-h mean position; white line with black outline) for the two MCS swaths and their MCS slices at 0300, 1000, and 1600 UTC. Centroid paths are included only for visualization purposes.

is that the merging (splitting) of multiple, unique, objects will result in a single, overly expansive, storm swath (Skok et al. 2009). For example, Chang et al. (2016) illustrate the "chaining effect" in which small, short-lived cells can incorrectly combine two unique regions of precipitation during the segmentation process. As an alternative, spatiotemporal overlap tracking (Lakshmanan et al. 2009) can be used to apply the spatiotemporal object-building procedure only to storms that spatially overlap between two adjacent radar images. Although tracking using the overlap criteria is thought to be overly conservative in many cases, its performance is similar to, or better than, more complex techniques for objects at the scale of MCSs (Lakshmanan and Smith 2010).

3. Data

The data generated for this study are extracted from the 5-min-temporal-resolution, approximately 2-km-spatial-resolution, National Operational Weather Radar (NOWrad; see Fabry et al. 2017) product, which is a CONUS-wide composite reflectivity mosaic. As in Part I, simplified pixel lengths and areas are defined as 2 km and 4 km^2, respectively. Each pixel value represents the instantaneous precipitation rate for the grid's location, and values are constrained to a range of 4-bit numbers (0–16) representing bins of 5 dBZ from 0 to 80. As in Part I, values representing 20–35 dBZ are labeled as stratiform, those between 40 and 45 dBZ are labeled as convection, and values of 50 dBZ and greater are labeled as intense. These data have been used in several studies that produced and examined climatologies of convection (Fabry et al. 2017). Since the mosaics are generated from NEXRAD reflectivity, the caveats associated with those data are also transferred to the raw data used to generate the product (Smith et al. 1996). Such issues include anomalous propagation, false echoes, attenuation, and other spurious signals relating to the curvature of the Earth and atmospheric conditions. To systematically reduce the occurrence of these problems, the data are initially quality controlled before they are released (Carbone et al. 2002). This work examines the data in 15-min intervals to reduce processing time.

MCS slices generated in Part I are used in this study to build MCS swaths. These data include geographic, intensity, and feature information (see Table 3 in Part I). A total of 48 perturbations were generated to test the sensitivity of the swath-building procedure, including 1) four different search radii for connecting convective cells [convective-region search radius (CRSR)]; 2) three different search radii for connecting stratiform regions to qualifying convective regions [stratiform search radius (SSR)]; and 3) four different P_{MCS} thresholds. In total, 742 242 slices were generated for the period of May–September in 2015 and 773 702 slices were generated for the same months in 2016. Each slice is saved as a lossless, 8-bit, portable network graphics (PNG) image and is indexed within a comma-separated-values (CSV) file, from which attributes such as geographic information, file location, slice features, and P_{MCS} can be queried to generate analyses. These files are available online (https://github.com/ahaberlie/MCS/).

4. Tracking

a. Overview

An important part of the PJ00 definition is that the organization of DMC at a larger scale than an individual updraft must persist long enough for mesoscale circulations to form. Because of this stipulation, studies that have tracked MCSs in remotely sensed data have required 1) that slices that meet size and intensity requirements must be spatiotemporally associated between time steps and 2) that these associations (i.e., swaths) must exist for a minimum amount of time (e.g., Table 1 in Part I). There are several explicit tracking approaches that can be used to generate spatiotemporal associations between MCS slices (e.g., Lakshmanan and Smith 2010, p. 703). The goal of tracking for this study is to spatiotemporally associate candidate MCS slices for the purpose of generating a database of MCS swaths that contain intensity, spatial, and temporal information. These swaths can then be queried, extracted, and analyzed for research applications.

Because there are 48 combinations of CRSR, SSR, and the probability that a slice is a part of an MCS (P_{MCS}), determining the differences (if any) between each perturbation could help to inform an optimal segmentation choice in the context of swath building (see Part I for more information on these values). Lakshmanan and Smith (2010) provides a framework for assessing the performance of storm-tracking algorithms by using summary statistics from all available swaths. They suggest that, in general, the relative performance of a storm-tracking approach can be determined by answering the following three questions: 1) How long do tracks typically last? 2) How variable is the intensity of the affiliated precipitation within the tracks? 3) How linear are the tracks? This approach to evaluating storm-tracking algorithms is suggested over track-by-track verification for large datasets (Lakshmanan and Smith 2010; Fiolleau and Roca 2013; Houston et al. 2015). The goal of using this assessment approach is not to create a general-purpose storm-tracking algorithm but rather to examine the relative performance between the available perturbations.

b. Approach

The tracking procedure uses two open-source packages in the Python programming language: ‘‘pandas/geopandas’’ (0.20.3/0.2.1; McKinney 2010) and ‘‘shapely’’ (1.5.17). First, slice-feature information for all of the slices from 2015 and 2016 is read into a pandas ‘‘Dataframe.’’ For each of the 48 perturbations, a query is used to select only those slices that are associated with each CRSR, SSR, and P_{MCS} value. Saved images associated with the resulting slices are then loaded into memory, and the locations of pixels exceeding 50 dBZ (intense) are used to generate a convex hull. The resulting shapely polygon approximates the MCS core and usually takes much less memory to store than do points for each pixel location. The polygon is then inserted into a geopandas Dataframe and associated with its affiliated slice.

For each 15-min period for May–September in 2015 (and again for the same period in 2016), slices are selected for the current and next time steps. The matching procedure then builds a two-dimensional matrix in which each row represents a slice within an existing track at the current time step and each column is an unmatched slice at the next time step. The similarity between the slices is calculated and is inserted into the affiliated cell. Similarity (normalized difference) is calculated by first dividing the feature values (see Table 3 in Part I) in each slice by the maximum value for each feature and then finding the 14-dimensional Euclidean distance between two slice features. This process is simplified by only calculating the similarity of overlapping slices and assigning a null value to all cells that are affiliated with slices that do not overlap each other. Then, until all values are null, the procedure finds the lowest value (highest similarity) and associates the unmatched slice with the track number affiliated with the slice at the current time step. Cells that represent matches are then set to null. All unmatched slices at the next time step are then considered to be new tracks and are assigned a new storm number. If only one overlap is found, the method behaves as a simple-overlap matching approach. If more than one overlap is found, the most similar slice is chosen to be associated with the existing track. This matching process is called the Hungarian method (Munkres 1957) and has been used in many storm-tracking algorithms (Dixon and Wiener 1993; Han et al. 2009; Lakshmanan et al. 2013; Gagne et al. 2017). A merging event at 0430 UTC 7 June 2015 can be used as an example of the Hungarian method

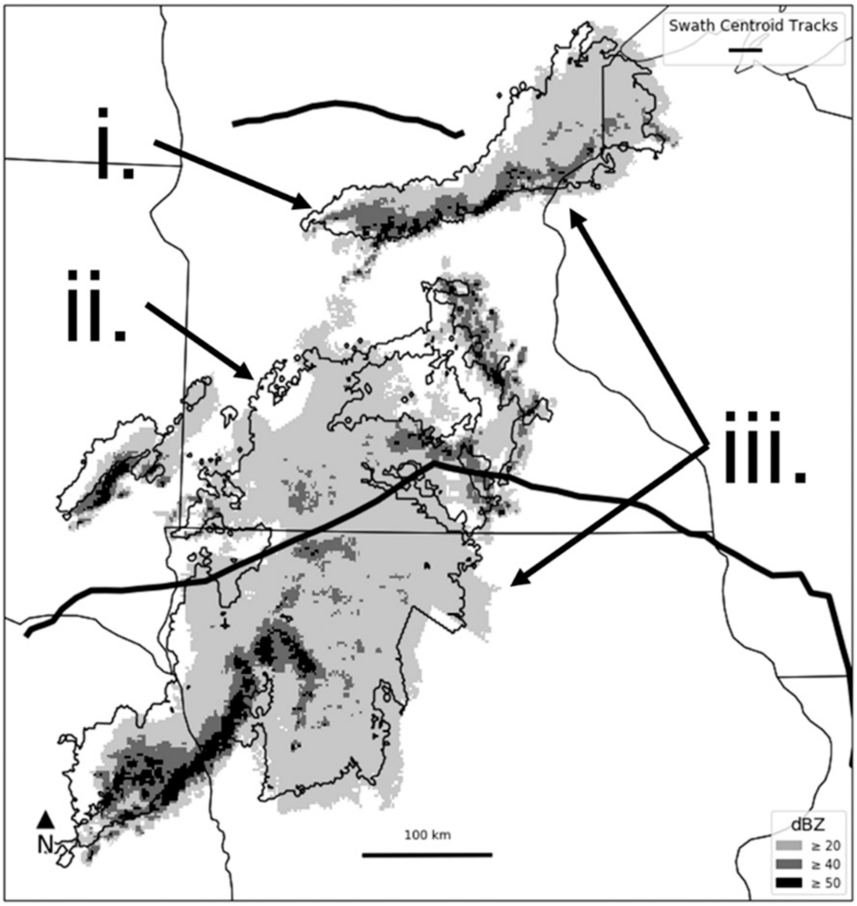

FIG. 2. Example of a merging case at 0430 UTC 7 Jun 2015. The black contours (labeled by i and ii) represent the stratiform precipitation extent of two slices associated with unique swaths at 0415 UTC 7 Jun 2015. Since both overlap with the single slice at 0430 UTC (label iii), the most similar slice retains its track, whereas the track associated with the least similar slice is discontinued. In this case, the southernmost slice (label ii) is most similar to the new, merged, slice (label iii). Centroid paths are included only for visualization purposes.

(Fig. 2). At 0400 UTC 7 June 2015, two distinct MCS swaths are ongoing in the upper-midwestern CONUS. At 0415 UTC 7 June 2015, the segmentation process determines that there are still two unique slices (labels i and ii in Fig. 2), resulting in a straightforward matching decision using spatiotemporal overlap only (Table 1). In the next radar image, however, the segmentation process determines that the two slices have merged (label iii in Fig. 2) and, thus, that only one slice at 0430 UTC overlaps with the two at 0415 UTC. In this case, the matching decision is determined by associating the two most similar slices (Table 2), and the southernmost slice is matched with the merged slice. Although this combination produced the lowest normalized difference of the available choices (Table 2), this value was roughly 10 times as large as the value for its previous, straightforward, match at 0415 UTC (Table 1). This is a result of the merging of the northern slice (Fig. 2, label i) and the southern slice (Fig. 2, label ii) and the affiliated modification of feature values.

TABLE 1. Normalized differences between existing slices at 0400 UTC (S_i and S_{ii}) and new slices at 0415 UTC (N_1–N_3) for 7 Jun 2015 (see Fig. 2). Normalized differences that are denoted with an asterisk are winning matches. For example, slice S_i is matched with new slice N_1. Em dashes in a cell denote that the new slices did not overlap with the corresponding existing slice. Note that N_3 is included as a case in which a new slice does not overlap with any existing slice.

	N_1	N_2	N_3
S_i	0.049*	—	—
S_{ii}	—	0.195*	—

TABLE 2. As in Table 1, but between existing slices at 0415 UTC and new slices at 0430 UTC (see Fig. 2). In this example, slice S_{ii} is matched with new slice N_1. In this case, N_1 is not matched with S_i; this is because the normalized difference between these two slices is greater than that of the normalized difference between S_{ii} and N_1.

	N_1	N_2	N_3
S_i	1.384	—	—
S_{ii}	0.495*	—	—

c. *Performance assessment*

1) SUBJECTIVE ASSESSMENT

Forty-eight different tracking outputs using the same underlying slice data are generated by the process described in section 4b. Subjective assessment of the tracking algorithm is performed on many known MCS cases, of which three cases are included in this paper: 1) a merging event between two MCSs, taking place between 0000 and 1700 UTC 7 June 2015, 2) a back-building (Maddox et al. 1979) MCS that occurred between 2200 UTC 24 June 2015 and 1800 UTC 25 June 2015, and 3) a derecho-producing (Johns and Hirt 1987; Corfidi et al. 2016) MCS that occurred between 0000 and 2300 UTC 22 June 2015. These cases do not represent the entire spectrum of possible morphologies and evolutions, but they are useful to subjectively demonstrate the strengths and weaknesses of the MCS segmentation, classification, and tracking procedure described in this paper and in Part I.

(i) 7 June 2015

At the beginning of this period (0000 UTC), several regions of relatively isolated DMC are ongoing over northeastern Nebraska, eastern South Dakota, and western Minnesota. By 0300 UTC, upscale growth and linear organization of the initially isolated DMC occurs in east-central Minnesota and in eastern Nebraska and western Iowa. As these linearly shaped regions of DMC (and associated stratiform precipitation) propagate eastward, they begin to undergo a merger (e.g., Foilleau and Roca 2013). This process occurs between 0400 and 0800 UTC in eastern Minnesota and Iowa and is denoted by the spatial meshing of two distinct stratiform rainfall shields, eventually followed by the combining of two distinct lines of DMC. At 1000 UTC, the merger is complete and the linear DMC has visual characteristics in radar images that are consistent with a mature MCS (''trailing stratiform'' morphology; PJ00). After 1200 UTC, the linear DMC begins to lose intensity; by 1700 UTC, much of the convective precipitation has dissipated as it moves into eastern Illinois and northwestern Indiana.

The evolution of this MCS is manually tracked by circling MCS slices in composite reflectivity mosaic images every 15 min in a manner consistent with how training and testing samples were gathered for Part I. These MCS slices are automatically combined into MCS swaths, and the output from this procedure is illustrated in Fig. 1. Results from this manual approach reflect what is described in the previous paragraph: 1) swaths begin where the DMC first took on MCS-like characteristics, 2) one track ends in central Minnesota (Fig. 1, label i), whereas another experiences a northward jump but continues eastward (Fig. 1, label ii); and 3) the main MCS swath path ends in eastern Illinois. Output from the automated approaches generally agrees with that generated by the manual approach (Fig. 3). In Fig. 3, the effects of various CRSR and SSR values are illustrated to assess subjectively the general performance of four select perturbations. In addition to illustrating the effect of varied search-radius values, the effect of varied minimum P_{MCS} per swath is demonstrated by using different-shaded centroid paths. These paths are only used for visualization purposes, because the actual ''track'' is the spatial coverage of a slice within the MCS swath at any given time.

In all of the cases, the MCS swaths generated by using all qualifying slices (P_{MCS} of 0.0) produce a centroid path that resembles that of the manual swaths (Fig. 1). When increasing the minimum P_{MCS} per swath, the paths begin to diverge from the manual path. For example, in all of the included cases in Fig. 3, the northern MCS swath (Fig. 1, label i) does not qualify as an MCS when using strict P_{MCS} thresholds of 0.90 and 0.95. This is likely due to the relatively small size of the convective and intense precipitation within the slices belonging to this MCS swath. The termination point for this MCS swath is reasonable, however, because it becomes merged with the southern MCS (Fig. 1, label ii) in eastern Minnesota in all four cases. One discrepancy between the manual swaths and the swaths generated for the northern MCS swath is its premature cessation in the automated approach that limits its southern and eastward extent. For the situation depicted in Fig. 3a, a short-lived swath between the northern and southern MCSs in southeastern Minnesota is identified at 0430 UTC. This swath and the northern MCS swath merge at 0500 UTC, at which point the matching process determines that the merged slice is the continuation of the short-lived swath. This merged swath continues until it merges with the southern MCS at 0545 UTC. For the situations depicted in Figs. 3b–d, the aforementioned short-lived swath depicted in Fig. 3a is instead attached to the southern MCS. This attachment causes a premature merger between the northern and southern MCS

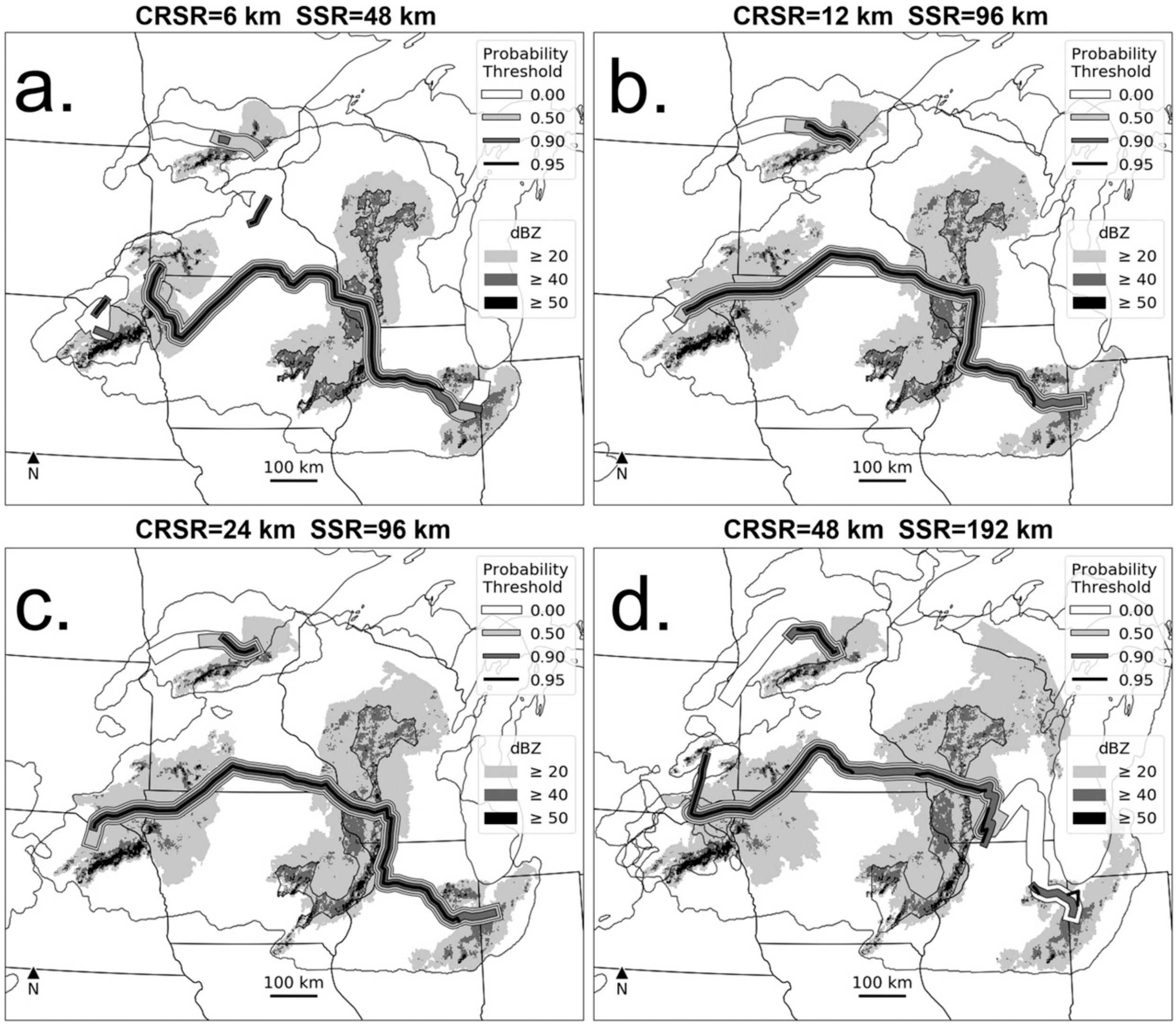

FIG. 3. The effects of modifying CRSR, SSR, and MCS probability thresholds on resulting MCS cores, MCS slices, and MCS swaths (see Fig. 1) and their affiliated centroid paths (2-h mean position) from 0000 to 1700 UTC 7 Jun 2015. Pictured are MCS slices from 0300, 1000, and 1600 UTC for each swath that lasted for at least 3 h. Also shown are tracks composed of slices meeting or exceeding an MCS probability of 0.90 that last at least 0.5 h. The different-shaded centroid tracks represent swaths generated by using only those slices that are assigned an MCS-label probability exceeding 0.00 (white), 0.50 (light gray), 0.90 (dark gray), or 0.95 (black). MCS core boundaries (black outlines) are plotted at 1000 UTC. The CRSR/SSR combinations are as follows, respectively: (a) 6/48, (b) 12/96, (c) 24/96, and d) 48 km/192 km. Centroid paths are included only for visualization purposes.

swaths at 0430 UTC for Figs. 3b and 3c and at 0400 UTC for Fig. 3d. These difficulties demonstrate the trade-offs among different values of SSR, CRSR, and minimum P_{MCS}, which can result in spurious swaths and incorrect disconnects and linkages. For example, the beginning of the track in Fig. 3a is incorrectly split into two tracks at 0200 UTC. This is caused by a small SSR, which allows a gap between two MCS core regions, resulting in two unique slices. In a similar way, around 0400 UTC a swath in eastern Minnesota is incorrectly identified between the northern and southern MCS swaths. In Fig. 3d, the swath for P_{MCS} of 0.95 experiences a spurious disconnect in southeastern Minnesota that is caused by a 30-min period in which the merged swath does not exceed the P_{MCS} threshold. The southwestward direction of the track for P_{MCS} of 0.90 in Fig. 3d reflects the splitting of the northern area of stratiform from the southern area of decaying convection. As the swath moves into eastern Wisconsin and northern Illinois, the higher-P_{MCS} swaths are lost, whereas the path for P_{MCS} of 0.00 is retained. This result is caused by the combination of the decay of the MCS and the inclusion of a broad area of stratiform precipitation that reduces the P_{MCS} values of the slices. The higher-P_{MCS} swaths in Fig. 3d return in eastern Illinois and northwestern Indiana when the northern region of stratiform is ‘‘detached’’ and is no longer associated with the swath. In contrast, for the tracks in Figs. 3a–c there are fewer stratiform pixels included within the slice and thus the premature track cessation is eliminated and the tracks are continuous.

(ii) 24–25 June 2015

During the evening and overnight hours of 24 and 25 June 2015, an MCS developed over Iowa and expanded eastward into northern Illinois [see Peters et al. (2017) for an in-depth discussion of this event]. Key features in the life cycle of this MCS included 1) the upscale growth of a loosely connected line of supercells into a southeastward-propagating bow echo from approximately 2300 to 0200 UTC, 2) back-building (Maddox et al. 1979) convection, resulting in a nearly stationary western flank of the MCS, despite the southeastward propagation of the eastern flank, from approximately 0200 to 1100 UTC, and 3) the splitting (e.g., Fig. 2 in Vila et al. 2008) of the region associated with a propagating linear segment from the region associated with weakening, quasi-stationary, convection around 1100 UTC. The event ends with the dissipation of the western and eastern regions of organized convection at approximately 1300 and 1500 UTC, respectively.

The swaths generated by the automated tracking procedure (Fig. 4a) match up well with the manual track (not shown). As in Fig. 3, the P_{MCS} thresholds of 0.0 and 0.5 produce a track that extends from central Iowa southeastward into central Indiana and western Ohio. The swath for P_{MCS} of 0.95 starts near the swath for P_{MCS} of 0.5 but ends approximately 100 km sooner at the Ohio and Indiana border. A splitting event (not shown) occurs between 1030 and 1045 UTC for P_{MCS} levels of 0.0, 0.5, 0.90 and 0.95, and the Hungarian method selects the slice over Indiana as the continuation of the swath that originated in Iowa. The southwestern slice forms a new swath that persists for approximately 2 h until 1230 UTC, appearing in Fig. 4a as the relatively short centroid paths for P_{MCS} of 0.90 and 0.95 in northeastern Missouri. The continuation of the original swath persists until around 1800 UTC, at which time it dissipates for all P_{MCS} values.

(iii) 22 June 2015

This period begins at 0000 UTC with two areas of linear DMC—a result of upscale growth by isolated DMC that developed during the late afternoon. By 0300 UTC, the northern and southern areas of DMC merged in southeastern North Dakota and exhibited two bowing segments (Przybylinski 1995) within the contiguous stratiform shield. The bowing segments, and their affiliated intense precipitation, dissipate by 0500 UTC as the MCS moves into western Minnesota. Farther west, multiple clusters of DMC develop concurrently in southwestern North Dakota and western and central South Dakota from 0400 to 0700 UTC. After 0800 UTC, these areas of DMC merge into a single MCS in southeastern South Dakota. Between 0800 and 1500 UTC the MCS took on a leading-line, trailing-stratiform (PJ00) appearance on radar before weakening as it approached Lake Michigan by 1700 UTC. The MCS then dissipated around 2100 UTC over eastern Michigan. This MCS produced wind damage and tornadoes from northern Iowa eastward into southern Michigan.

For this event, the tracking procedure produced two MCS swaths (Fig. 4b). The first swath is associated with the initial area of linearly organized DMC in northern and central North Dakota. Two regions of linear DMC merged by 0215 UTC in eastern North Dakota, with the southern, short-lived swath plotted in south-central North Dakota (P_{MCS} of 0.90 and 0.95) between the centroid paths of the two main MCS swaths. The initial MCS swath dissipates and merges with the second MCS swath around 0600 UTC. This swath then moves over Minnesota, Iowa, Wisconsin, Illinois, and Michigan before dissipating around 2300 UTC. Similar to some of the previous examples, the swaths for P_{MCS} of 0.00, 0.50, and 0.90 form a contiguous path from North Dakota to eastern Michigan. The P_{MCS}-0.95 swath initially forms a continuous path from North Dakota to western Michigan by 1800 UTC; after this time, the path only intermittently shows up over Michigan before the system dissipates around 2300 UTC. This lack of swath cohesion coincides with the visual appearance of weakening by the MCS as was previously noted in the subjective assessment of the event.

(iv) Intermittent swaths

In the previous three examples, there were cases in which the spatiotemporal matching procedure failed to create contiguous swaths. This was particularly true when the P_{MCS} threshold exceeded 0.95. Although the goal of using higher probabilistic thresholds is to reduce the inclusion of non-MCS events, these cases suggest that this approach may also be removing, truncating, or splitting legitimate MCS swaths. We hypothesize that this is caused by periodic reductions in the P_{MCS} value for slices that cause them not to be included in a spatiotemporal matching run that only considers slices that exceed, say, P_{MCS} of 0.95. Because the matching procedure only examines the current period and the

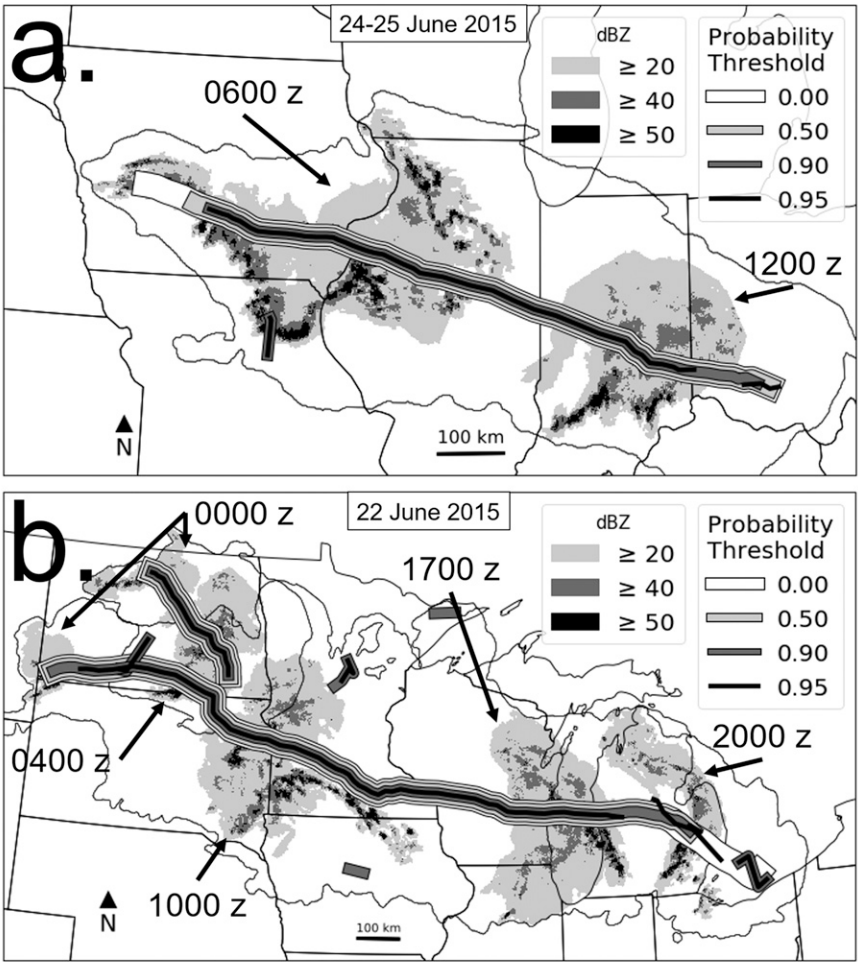

FIG. 4. Example output of slice, swath, and swath-centroid track from select periods and regions during June of 2015: (a) 2300–1800 UTC 24–25 Jun and (b) 0000–2300 UTC 22 Jun. The shades for the swath-centroid track, representing the minimum MCS probability per swath, are the same as described in Fig. 3. The CRSR chosen for these maps is 24 km, and the SSR is 96 km. Swaths are included on the basis of criteria in Fig. 3. Centroid paths are included only for visualization purposes.

next 15-min period, if a slice fails to exceed a threshold value in one radar image then the track is ended.

One way to address this issue is to reanalyze the track database to connect previously unconnected tracks. This method is demonstrated by Lakshmanan et al. (2015), who provided evidence that it resulted in more contiguous tracks (see their Fig. 6). The current study uses a similar approach. Namely, the goal is to attempt to connect the end of swaths that contain at least two slices (30-min duration) to the beginning of swaths with at least two slices. To find suitable matches, the following conditions must be met: 1) the start of the matching candidate swath must not exceed 60 min from the time that the previous swath ended and 2) the first slice of the matching candidate swath must either overlap or be within 100 km of the last slice in the previous swath. This process is illustrated on two previously discussed examples from 7 and 22 June 2015 (Fig. 5). In Fig. 5a, the reanalyzed swaths exhibit a more contiguous track for all P_{MCS} thresholds—at least until the MCS moves into eastern Wisconsin, whereas the original swaths (with the same CRSR and SSR) in Fig. 3d display a swath discontinuity in southeastern Minnesota for P_{MCS} of 0.95. In Fig. 5b, the reanalyzed swaths improve on the original swaths in Fig. 4b by producing one contiguous swath for P_{MCS} of 0.95.

2) OBJECTIVE ASSESSMENT

Objective assessment of the tracking performance was achieved by calculating and comparing select summary statistics for reanalyzed swaths generated by each perturbation (Lakshmanan and Smith 2010). Namely,

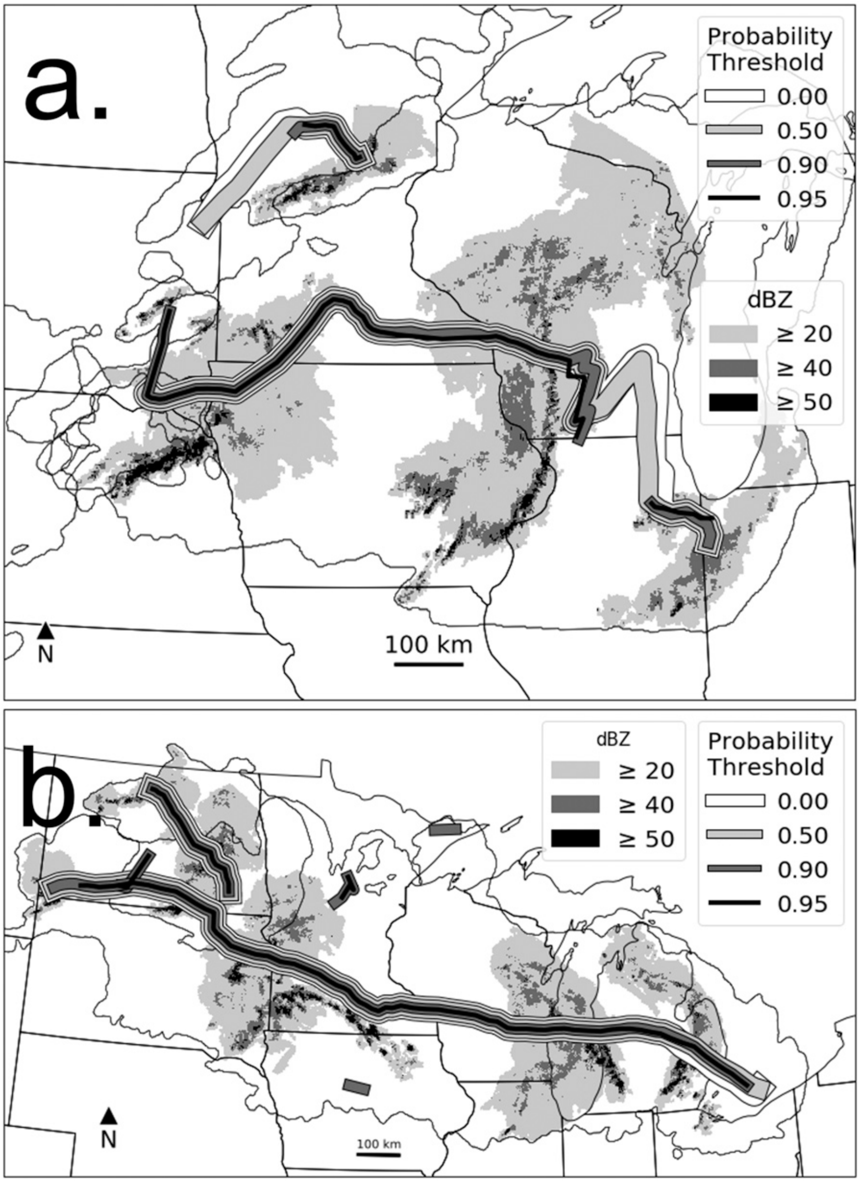

FIG. 5. Example output of slice, swath, and swath-centroid track using reanalyzed tracks from select periods and regions: (a) 0000–1700 UTC 7 Jun and (b) 0000–2300 UTC 22 Jun. The shades for the swath-centroid track, representing the minimum MCS probability per swath, are the same as described in Fig. 3. The CRSR and SSR chosen for these maps are (a) 48 and 192 km, respectively, and (b) 24 and 96 km, respectively. Pictured are MCS slices from 0300, 1000, and 1600 UTC for each swath that lasted for at least 3 h. Swaths are included on the basis of criteria in Fig. 3. Centroid paths are included only for visualization purposes.

duration, standard deviation of reflectivity, and linearity error were calculated for each swath, and the average values for each perturbation are compared. The duration of each swath is calculated by finding the temporal difference between its last slice and its first slice. The standard deviation of reflectivity is calculated by using all nonzero pixel values in each slice within an MCS swath. The linearity error is calculated by first fitting a line to all slice centroids within a swath (using scikit-learn's ''LinearRegression'') and then finding the root-mean-square error between points on that line and observed centroids. The best-performing perturbation, according to Lakshmanan and Smith (2010), is the one with the longest mean duration, lowest mean standard deviation of reflectivity, and lowest mean linearity error. To assess quantitatively the best-performing

TABLE 3. The effects of varying CRSR, SRS, and minimum MCS probability threshold (0.0, 0.5, 0.9, or 0.95) on the count of MCS swaths, count of slices within MCS swaths, and the percentage of slices that are contained within MCS swaths. To qualify, an MCS swath must last at least 3 h.

		≥0.0			≥0.5			≥0.9			≥0.95		
CRSR (km)	SSR (km)	Swaths	Slices	%	Swaths	Slices	%	Swaths	Slices	%	Swaths	Slices	%
6	48	955	23 988	55	520	13 851	66	323	8506	68	277	7084	66
6	96	885	23 773	58	591	16 955	68	463	13 187	72	413	11 583	72
6	192	837	23 580	59	605	17 490	69	501	13 739	71	444	11 865	71
12	48	1182	30 224	56	588	15 978	68	376	10 265	71	329	8693	69
12	96	1071	29 475	57	664	19 497	69	514	14 890	73	448	12 803	73
12	192	1033	29 241	59	677	19 732	69	561	15 239	72	461	12 619	71
24	48	1581	40 552	58	653	18 793	70	442	12 500	73	388	10 687	72
24	96	1469	39 686	59	774	22 886	72	580	17 113	75	503	14 287	74
24	192	1414	39 224	60	793	22 999	72	591	16 697	73	492	13 450	71
48	48	2087	54 239	61	726	21 837	72	537	15 254	75	461	12 315	73
48	96	1946	52 904	62	798	25 688	73	639	18 655	75	533	14 802	75
48	192	1877	52 045	62	841	25 613	73	636	17 746	73	507	13 402	72
Mean		1361	36 578	59	686	20 110	70	514	14 483	73	438	11 966	72
Std dev		417	11 129	2	97	3603	2	94	2899	2	72	2146	2

perturbation, a total normalized error metric is calculated by finding the normalized sum of the means for linearity error, intensity error, and negative duration.

Counts of MCS swaths—swaths that last at least 3 h—vary from 277 to 2087 over the 5-month period in 2015 and from 316 to 2173 in 2016 (Tables 3 and 4). Swath counts decrease as the P_{MCS} threshold increases, with a mean of 1361 swaths for all P_{MCS}-0.00 perturbations and 438 swaths for all P_{MCS}-0.95 perturbations in 2015. Counts in 2016 exhibited a similar decrease, with values of 1423 and 456. As a comparison, Pinto et al. (2015) identified 837 and 929 MCSs during June–August in 2012 and 2013, respectively. During the same months in 2015, this study identified between 202 and 1456 MCS swaths (197 and 1552 in 2016), depending greatly on the P_{MCS} that was used. This disparity is likely caused by the different methodological approaches used, specifically as they relate to segmentation. For example, in Fig. 7 in Pinto et al. (2015), the three outlined clusters in North Dakota, South Dakota, Wyoming, and Montana would be considered part of the same MCS slice in this study (depending on the combination of CRSR and SSR). Further, Pinto et al. (2015) state that their intention was to include "less organized convective areas" that are common in the southeastern United States. As illustrated in Part I, using a higher P_{MCS} results in fewer available slices in this region and, thus, fewer MCS swaths. Also, as the P_{MCS} threshold increases, the percentage of qualifying slices that are part of an MCS swath also increases. This suggests that slices with higher P_{MCS} are more likely to be within a long-lasting swath

TABLE 4. As in Table 3, but for 2016.

		≥0.0			≥0.5			≥0.9			≥0.95		
CRSR (km)	SSR (km)	Swaths	Slices	%	Swaths	Slices	%	Swaths	Slices	%	Swaths	Slices	%
6	48	971	24 719	55	534	14 661	67	349	9528	72	316	8134	71
6	96	884	24 622	58	608	18 037	68	475	14 116	73	434	12 396	73
6	192	834	24 332	59	598	18 335	69	504	14 761	72	470	13 024	73
12	48	1240	31 574	56	590	16 789	68	390	11 109	73	334	9408	71
12	96	1129	30 776	58	686	20 512	70	515	15 833	74	477	13 887	75
12	192	1070	30 347	59	681	20 845	70	553	16 143	72	489	13 782	72
24	48	1657	42 613	58	664	19 676	70	462	13 541	75	412	11 629	74
24	96	1562	41 767	59	775	23 756	72	578	17 803	75	502	15 131	75
24	192	1510	41 089	60	795	24 005	72	628	17 996	74	532	14 803	73
48	48	2173	56 683	61	749	22 668	72	550	16 077	76	457	12 977	74
48	96	2040	55 438	62	841	22 861	73	650	19 461	75	526	15 430	74
48	192	2006	54 914	62	871	27 122	73	640	18 683	72	525	14 390	71
Mean		1423	38 240	59	699	20 772	70	525	15 421	73	456	12 916	73
Std dev		452	11 824	2	103	3 361	2	91	2878	1	68	2156	1

TABLE 5. Select summary statistics for each of the 48 combinations of CRSR (km), SSR (km), and minimum MCS probability threshold (0.0, 0.5, 0.9, or 0.95) per swath for May–September in 2015. Included statistics are mean per swath: duration (labeled Dur; h), 2) standard deviation of reflectivity (StdDev; dBZ), mean linearity error (LinErr; km), and normalized total error (NorErr). The boldface cells denote the lowest values of mean reflectivity standard deviation, linearity error, and normalized error and the highest values of duration.

		≥0.0				≥0.5				≥0.9				≥0.95			
CRSR	SSR	Dur	StdDev	LinErr	NorErr	Dur	StdDev	LinErr	NorErr	Dur	StdDev	LinErr	NorErr	Dur	StdDev	LinErr	NorErr
6	48	2.90	8.77	**21.88**	0.05	3.64	8.22	26.80	−0.03	3.67	8.20	27.67	−0.01	3.53	8.20	27.23	0.01
6	96	3.08	8.65	23.83	0.06	3.89	8.11	28.46	−0.04	4.05	8.08	30.62	−0.02	4.08	8.12	30.08	−0.04
6	192	3.19	8.60	25.58	0.08	3.95	8.05	30.14	−0.01	3.96	7.99	29.30	−0.05	3.89	8.04	29.43	−0.02
12	48	2.89	8.83	21.94	0.07	3.73	8.26	27.60	−0.02	3.89	8.23	28.74	−0.02	3.71	8.23	28.48	0.01
12	96	3.02	8.73	23.53	0.07	3.93	8.10	29.04	−0.04	4.17	8.06	30.91	−0.04	4.09	8.10	31.12	−0.01
12	192	3.10	8.68	24.81	0.09	3.95	7.99	29.97	−0.03	3.97	7.95	29.94	−0.04	3.88	8.02	29.76	−0.01
24	48	2.99	8.88	22.18	0.06	3.95	8.25	28.26	−0.05	4.05	8.23	29.45	−0.04	3.91	8.24	29.29	−0.01
24	96	3.10	8.79	23.46	0.06	4.19	8.03	30.36	−0.06	4.36	8.00	31.66	**−0.07**	4.25	8.06	31.10	−0.05
24	192	3.15	8.76	24.63	0.08	4.15	7.91	31.11	−0.05	4.08	7.91	30.41	−0.05	3.93	8.00	29.17	−0.04
48	48	3.19	8.84	25.45	0.10	4.20	8.21	32.38	0.02	4.22	8.22	33.48	0.04	3.94	8.22	31.51	0.05
48	96	3.30	8.76	26.18	0.09	**4.49**	8.00	32.73	−0.06	4.46	8.01	32.68	−0.06	4.19	8.05	31.62	−0.02
48	192	3.33	8.73	27.02	0.11	4.37	**7.89**	32.85	−0.05	4.18	7.94	32.04	−0.02	3.91	8.02	30.49	0.00
Mean		3.10	8.75	**24.21**	0.08	4.04	8.09	29.98	**−0.03**	**4.09**	**8.07**	30.57	−0.03	3.94	8.11	29.94	−0.01
Std dev		0.14	0.08	1.62	0.02	0.24	0.12	1.94	0.02	0.20	0.12	1.62	0.03	0.19	0.09	1.26	0.03

and that not all slices that meet the objective PJ00 criteria belong to an MCS swath.

In general, swaths that were generated using only those slices that had a P_{MCS} of at least 0.50 and that lasted at least 0.5 h had longer durations than swaths that used all qualifying slices (Tables 5 and 6; Fig. 6). Swaths from 2015 built using all qualifying slices had mean durations ranging from 2.90 to 3.33 h (from 2.84 to 3.36 h in 2016), which increased as the CRSR and SSR values increased. In comparison, swaths built using only slices with a P_{MCS} of 0.50 or greater had mean durations ranging from 3.64 to 4.49 h (from 3.77 to 4.44 h in 2016), with a mean increase of around 55 min (61 min in 2016) over the P_{MCS}-0.00 swaths. Swath durations were maximized when using a $\mathrm{P_{MCS}}$ threshold of 0.90, with a mean duration exceeding 4 h for both years. Because of the relatively low probability of false detection enforced by the minimum P_{MCS} of 0.95, slices with attributes that deviate slightly from those of slices used to train the classifiers are disqualified from the matching process, resulting in a slight decrease in duration from a P_{MCS} of 0.90 for 2015 and 2016. As a result, it is more likely that the spatiotemporal overlap procedure will fail to produce a match. Despite this, results from two-sample Kolmogorov–Smirnov (KS) tests (Kolmogorov 1933) suggest that the swath reanalysis led to significantly different duration distributions for all perturbations (significance level $p < 0.001$). In addition to the increases in mean and median values for all perturbations, these differences in distribution characteristics are likely due to reanalyzed swaths with longer durations. Mean duration increases after the reanalysis range from 0.91 to 1.69 h.

The per-swath standard deviation of reflectivity (intensity error) for swaths lasting at least 1 h ranged from 7.89 to 8.88 dBZ in 2015 and from 7.97 to 8.97 km in 2016 (Tables 5 and 6; Fig. 7). There was not much variation in the means or medians of this metric among the various perturbations, but there was a marked difference in the variability among the four P_{MCS} thresholds. For a P_{MCS} threshold of 0.00, the range from the 5th to 95th percentile was from 6 to 12 dBZ for both 2015 and 2016. In contrast, swaths using a P_{MCS} threshold of 0.95 had a range as small as 7–10 dBZ for both years. This suggests that the lower P_{MCS} thresholds may be capturing more events with unusually high and unusually low variability in reflectivity. This could be explained by the disqualification of small convective clusters (high reflectivity variability) and larger, more synoptic, rainfall clusters (low reflectivity variability) with an increasing P_{MCS} threshold. When comparing the distribution of intensity error in preanalysis and postreanalysis swaths, results from two-sample KS tests suggested that there were no significant ($p < 0.001$) differences for any of the perturbations.

Mean linear error for swaths lasting at least 1 h ranged from 21.88 to 33.48 km in 2015 and 21.62 to 34.72 km in 2016 (Tables 5 and 6; Fig. 8). In general, these values increased as CRSR, SSR, and P_{MCS} increased, with the lowest mean linear error belonging to swaths with slices built using a CRSR of 6 km, an SSR of 48 km, and a P_{MCS} threshold of 0.00. For the CRSR and SSR, the chaotic nature of stratiform and convective precipitation can result in unpredictable "chaining" (Chang et al. 2016) between radar images (Houston et al. 2015). A merging event, for example, can shift the swath centroid

TABLE 6. As in Table 5, but for 2016.

		≥0.0				≥0.5				≥0.9				≥0.95			
CRSR	SSR	Dur	StdDev	LinErr	NorErr	Dur	StdDev	LinErr	NorErr	Dur	StdDev	LinErr	NorErr	Dur	StdDev	LinErr	NorErr
6	48	2.84	8.87	**21.62**	0.09	3.77	8.33	27.49	−0.01	3.99	8.27	30.03	0.00	3.89	8.28	29.60	0.01
6	96	3.04	8.74	23.87	0.09	3.98	8.19	28.79	−0.04	4.16	8.14	31.21	−0.01	4.15	8.18	30.92	−0.02
6	192	3.17	8.67	24.88	0.08	4.04	8.09	28.79	−0.06	4.07	8.06	30.08	−0.04	4.13	8.10	30.36	−0.04
12	48	2.92	8.93	22.21	0.09	3.88	8.34	28.81	−0.01	4.08	8.30	29.97	−0.01	3.99	8.32	30.15	0.01
12	96	3.06	8.80	23.25	0.08	4.10	8.17	28.42	−0.06	4.34	8.11	31.41	−0.05	4.32	8.14	31.31	−0.05
12	192	3.14	8.75	24.17	0.08	4.13	8.07	28.93	−0.07	4.11	8.04	29.32	−0.07	4.03	8.08	28.75	−0.06
24	48	3.04	8.97	22.28	0.07	4.05	8.35	29.27	−0.03	4.34	8.32	31.62	−0.02	4.27	8.31	31.28	−0.02
24	96	3.14	8.85	23.27	0.06	4.29	8.12	28.96	−0.09	**4.47**	8.09	31.10	**−0.09**	4.39	8.15	31.26	−0.06
24	192	3.17	8.80	24.08	0.08	4.26	8.01	29.62	−0.07	4.19	8.02	29.87	−0.07	4.05	8.08	28.93	−0.06
48	48	3.23	8.91	25.15	0.11	4.25	8.34	33.53	0.05	4.35	8.35	34.72	0.07	4.15	8.32	32.63	0.05
48	96	3.33	8.81	25.96	0.10	4.44	8.08	32.42	−0.05	4.28	8.12	33.23	0.01	4.14	8.17	31.92	0.01
48	192	3.36	8.77	26.70	0.10	4.42	**7.97**	33.28	−0.03	4.02	8.07	30.28	−0.02	3.87	8.11	29.35	−0.01
Mean		3.12	8.82	**23.95**	0.09	4.13	8.17	29.86	**−0.04**	**4.20**	**8.16**	31.07	−0.03	4.11	8.19	30.54	−0.02
Std dev		0.14	0.08	1.48	0.01	0.20	0.13	1.93	0.04	0.15	0.11	1.49	0.04	0.15	0.09	1.17	0.04

several dozen kilometers between scans. For the case in Fig. 1, the mid-life-cycle merging event, followed by an end-of-life-cycle splitting event, produces a long-lasting, northward-bulging, arc away from a best-fit line. This results in a linear error of 87 km for the manually generated southern MCS swath (label ii in Fig. 1). For larger P_{MCS} thresholds, the swath centroid will be more chaotic, because there are many cases in which the swath-centroid track will jump more than 15 min ahead (e.g., Fig. 4b). The values produced by this study are

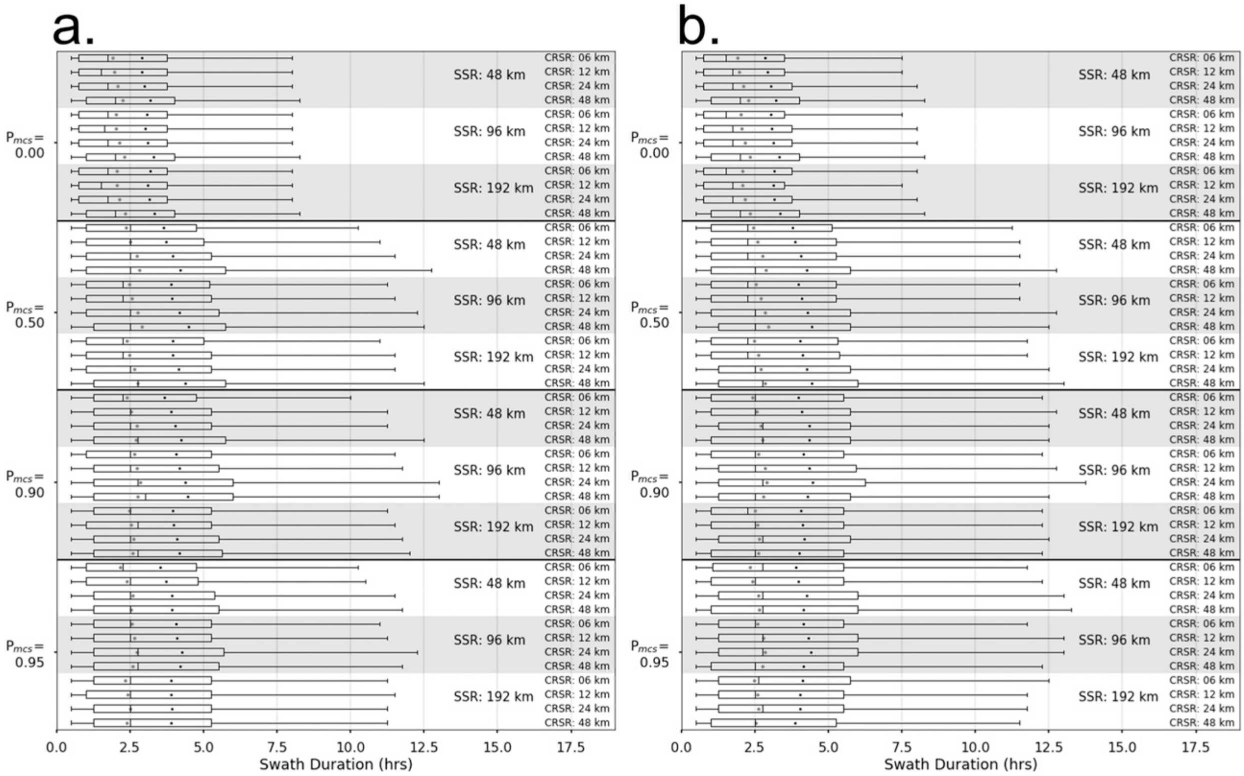

FIG. 6. Distribution of reanalyzed swath durations in hours for combinations of CRSR, SSR, and P_{MCS} for (a) 2015 and (b) 2016. The duration is calculated by finding the time-stamp difference between the last slice and first slice in a swath. The distribution medians and means are denoted with black vertical lines and black dots, respectively. The gray dots are the mean duration values for swaths before the reanalysis process. The box represents the interquartile range. The whiskers represent values between the 5th and 95th percentiles.

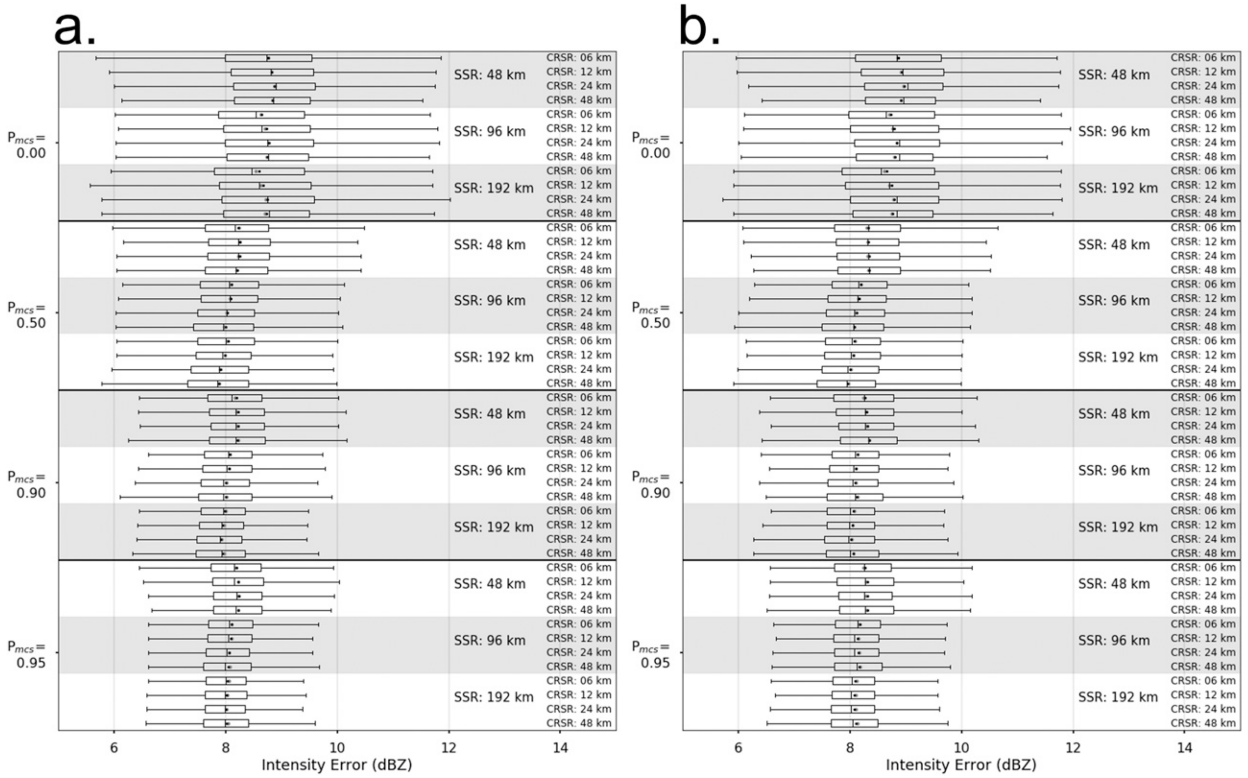

FIG. 7. As in Fig. 6, but for the standard deviation of reflectivity values (for pixels greater than 0 dBZ).

much larger than those presented by Lakshmanan and Smith (2010) and Houston et al. (2015). Although this could be partly due to tracking deficiencies, one major contributor to linearity error is the size of the storm cluster (Fig. 8). Houston et al. (2015) state that one of the goals for their tracking algorithm is to be sensitive to detecting ''reasonably small-scale storms,'' and Lakshmanan and Smith (2010) use a minimum storm size of 20 km^2. In comparison, MCS slices analyzed in Part I typically range in size from 10 000 to 100 000 km^2. As was the case with mean duration, the distributions of linearity error were significantly different between pre- and postreanalysis swaths for many of the perturbations, on the basis of results from a two-sample KS test. Because all of the mean values of linearity increased, this result suggests that the reanalysis step generally introduces more linearity error. This could also be a by-product of significantly longer tracks after reanalysis (Houston et al. 2015).

Relative performance can be quantitatively measured for each perturbation by combining duration, intensity error, and linearity error into a single error metric. This is performed by finding the sum of the negative normalized duration, normalized intensity error, and normalized linearity error. For this study, negative duration is used because a longer track suggests better tracking performance, whereas increases in intensity error and linearity error suggest worse performance (Lakshmanan and Smith 2010). To assess each perturbation's performance relative to the mean, the sum of errors is subtracted from the mean sum of errors across all perturbations (Tables 5 and 6; Fig. 9). Swaths generated using all qualifying slices (P_{MCS} of 0.00) have the worst collective performance of the four reported P_{MCS} thresholds. Swaths using P_{MCS} thresholds of 0.50 had the best performance in 2015 and 2016, although values for P_{MCS} of 0.90 were similar for both years. For both years, P_{MCS} of 0.95 performed better than P_{MCS} of 0.00. In 2015 and 2016, the best-performing perturbation used a CRSR of 24 km, an SSR of 96 km, and a P_{MCS} threshold of 0.90, whereas the worst-performing perturbation used a CRSR of 48 km, an SSR of 192 km, and a P_{MCS} threshold of 0.00. A major caveat of these results, in the context of general-purpose storm tracking, is that they are describing the *relative* performance of the 48 perturbations; that is, these results are only meaningful when considering tracking MCSs.

5. 2015 and 2016 warm-season case studies

To demonstrate the utility of the method described in this paper, we examine the spatiotemporal frequency of

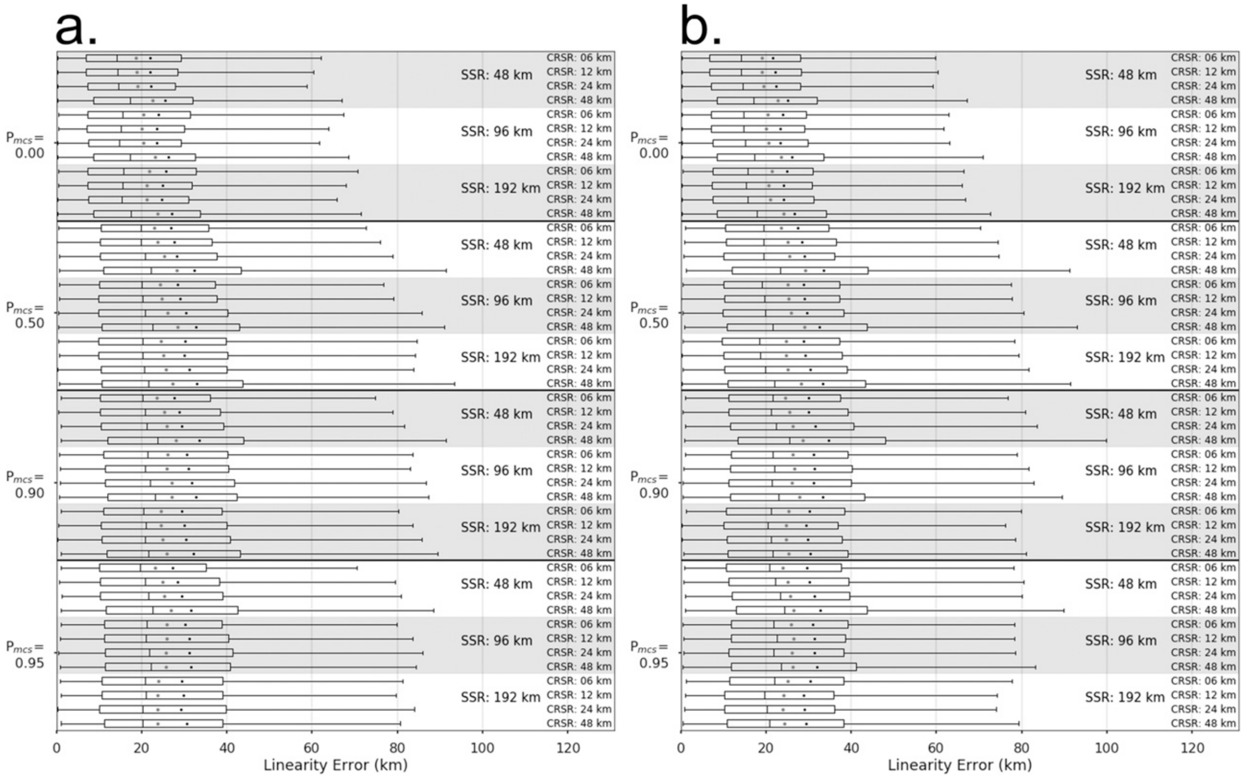

FIG. 8. As in Fig. 6, but for linearity error (km).

generated MCS swaths (Figs. 10 and 11). This analysis also serves as a subjective validation of the method; namely, the spatial patterns of MCS activity are compared with applicable studies and climatological expectations. The data in Figs. 10 and 11 are generated by selecting only those events that lasted for 3 h or more (MCS swaths) from each of the 48 reanalyzed track perturbations for 2015 and 2016. In general, the area covered by MCS swaths increases as CRSR and SSR increase. This is not surprising, because larger values of CRSR allow for more nearby cells to be combined into one larger MCS core, which, in turn, permits more area to be searched for affiliated stratiform regions. As was illustrated in Part I, MCS swaths generated using a P_{MCS} value of 0.95 result in the retraction of relatively high MCS swath frequency to the east-central Great Plains. In contrast, MCS swaths using a P_{MCS} value of 0.00 extend the same 40-h isopleth to most of the Gulf and Atlantic Coasts for some perturbations. For all of the perturbations, the maximum MCS swath occurrence lines up well with comparable climatologies (Rodgers et al. 1985; Augustine and Howard 1988, 1991; Anderson and Arritt 1998, 2001; Ashley et al. 2003). These studies found that mesoscale convective complexes and other MCS subtypes occurred most often in the central and eastern plains during the warm season. On the other hand, studies such as those by Geerts (1998) and Pinto et al. (2015) have included "less organized convective areas" in the southeastern United States in their MCS analyses, resulting in a frequency maximum along the Gulf Coast. The use of a 50-dBZ threshold to generate MCS cores could exclude many of these events from the dataset generated by this study.

Next, a subjective comparison between MCS swath occurrence generated by this study and an external source is performed. To be specific, the results of this study are compared with those presented by Geerts et al. (2017). That study objectively required an MCS to have the following properties: 1) the maximum precipitation intensity is greater than or equal to 35 dBZ, 2) the horizontal extent of intensity of at least 35 dBZ is greater than or equal to 100 km, and 3) the precipitation cluster lasts at least 1 h. For their purposes, they only examined precipitation clusters that occurred between 0200 and 1100 UTC (i.e., "nocturnal") from 1 June to 15 July 2015. They found that the greatest nocturnal MCS activity occurred in southern Iowa, southeastern Nebraska, northeastern Kansas, northern Missouri, and southern Illinois (see Fig. 1 in Geerts et al. 2017). To generate comparable frequency maps for the current study (Fig. 12), reanalyzed MCS swaths (CRSR = 24 km and SSR = 96 km) are selected for the same dates and times and the following P_{MCS} thresholds are used: 0.00

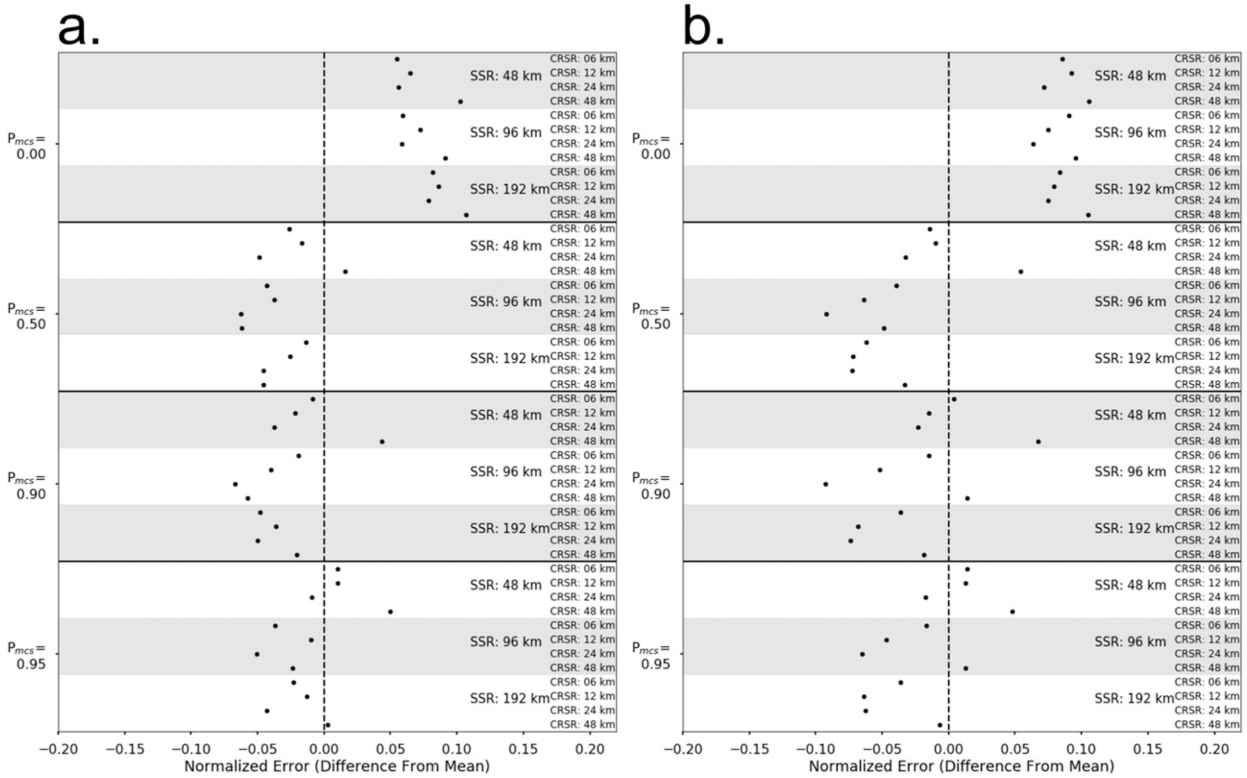

FIG. 9. Total normalized error for combinations of CRSR, SSR, and P_{MCS}, denoted by black dots, for (a) 2015 and (b) 2016. The sum of normalized linearity error and intensity error is subtracted from the normalized duration value. The difference between the sum for each perturbation and the average sum for all perturbations is the reported normalized error metric. A negative normalized error suggests better-than-average performance.

(Fig. 12a), 0.50 (Fig. 12b), 0.90 (Fig. 12c), and 0.95 (Fig. 12d). The resulting frequency maps agree reasonably with the map presented by Geerts et al. (2017). The following spatial features exist in both datasets: 1) the placement and shape of the relative MCS activity maximum extending from western Nebraska southeastward to central Tennessee, 2) the placement of the overall maximum in southeastern Nebraska and northwestern Missouri (particularly for swaths for P_{MCS} of 0.90 and 0.95), 3) the location of regional MCS activity minima in southern Wisconsin, northern Arkansas, and southwestern Kansas, and 4) the location of a regional MCS activity maximum in northern Texas. Although the MCS qualification criteria vary between the two studies, similarities in the spatial structure of MCS activity for this period are encouraging.

Last, we demonstrate the use of the dataset to generate time series analyses of the spatial coverage of convective pixels associated with MCS swaths over the CONUS for June of 2015 (Fig. 13). The darkened areas in Fig. 13 represent nocturnal hours (Geerts et al. 2017) to illustrate the diurnal cycle of MCS activity. For comparative purposes, the spatial coverage of all convective pixels in each image is calculated, as well as a differentiation in the area covered by swaths for P_{MCS} thresholds of 0.00 and 0.95. This map effectively shows that in many cases the timing of the maximum diurnal convective coverage over the CONUS does not match up with the maximum in the areal coverage of convection within MCS swaths. This result is expected, because MCSs are largely a late-evening and overnight phenomenon for many parts of the CONUS (Carbone et al. 2002), whereas smaller-scale DMC frequency is largely controlled by the diurnal cycle of instability (Carbone and Tuttle 2008; Haberlie et al. 2015). Further, using a P_{MCS} threshold of 0.95 instead of 0.00 appears to strengthen this diurnal disparity. One example occurred on 11 June 2015; in this example, the spatial coverage of convection associated with P_{MCS}-0.00 swaths is strongly tied to the overall convective coverage and peaks in the late afternoon. In contrast, convective coverage associated with P_{MCS}-0.95 swaths peaks overnight.

6. Discussion and conclusions

This paper is the second of two related papers that describe, verify, and utilize an MCS segmentation,

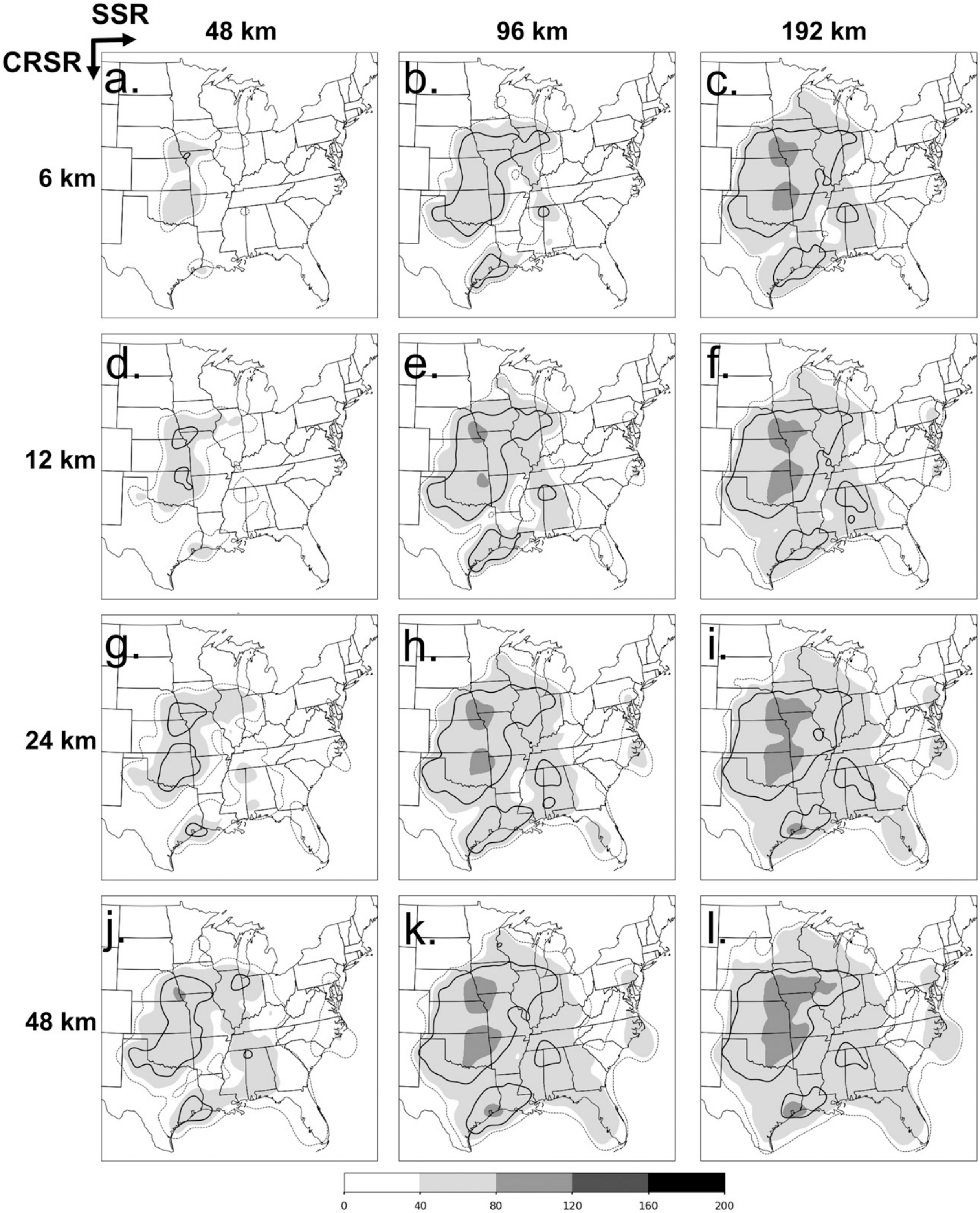

FIG. 10. Spatial occurrence (h; shaded) of MCS swaths (minimum of 3 h) with a P_{MCS} of 0.5 or higher in 2015 during May–September for varying CRSR and SSR. The solid line denotes the 40-h isopleth for slices with a P_{MCS} of 0.95 or higher, and the dotted line denotes the 40-h isopleth for all qualifying slices. The CRSR values are (a)–(c) 6, (d)–(f) 12, (g)–(i) 24, and (j)–(l) 48 km. The SSR values are (left) 48, (center) 96, and (right) 192 km.

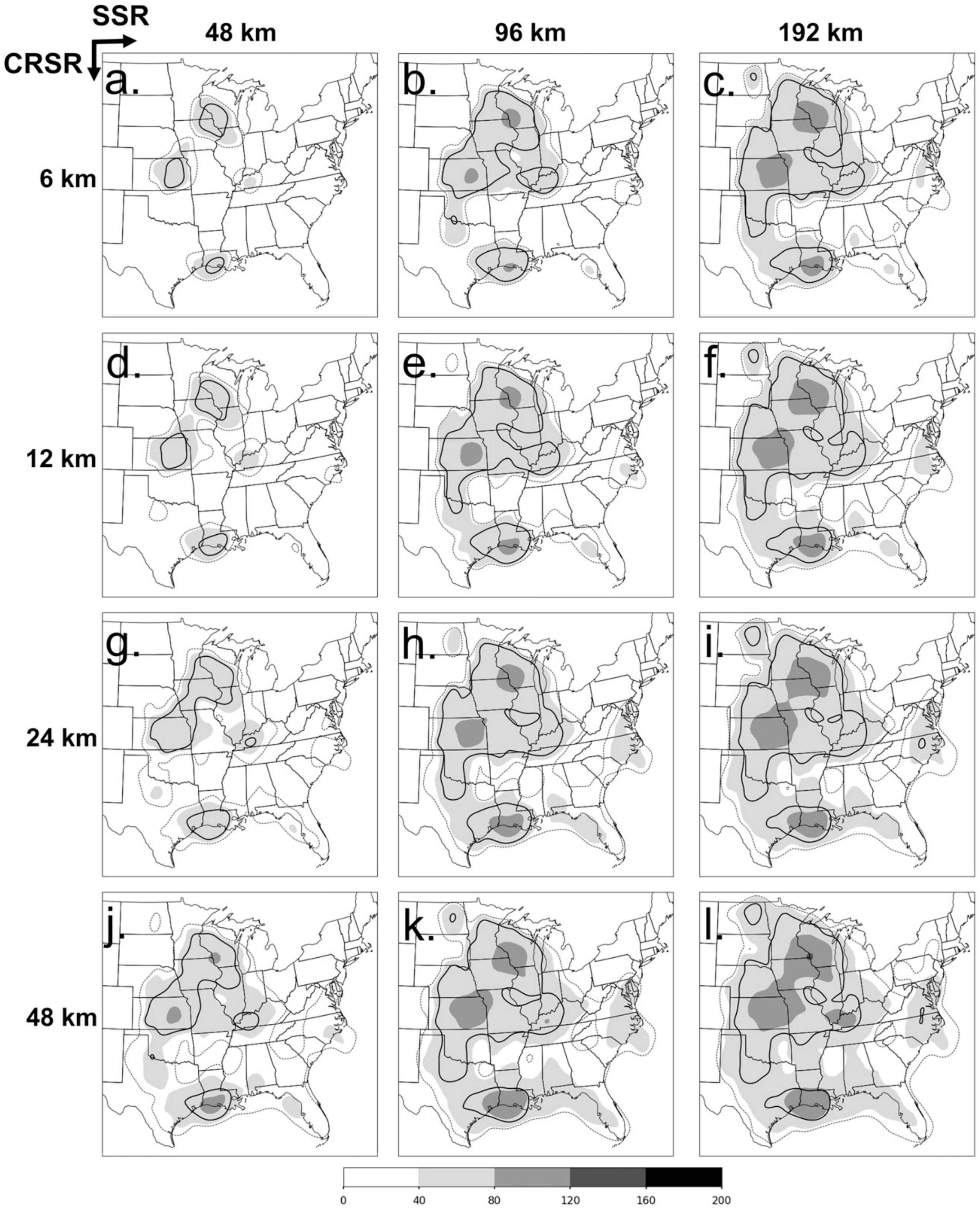

FIG. 11. As in Fig. 10, but for 2016.

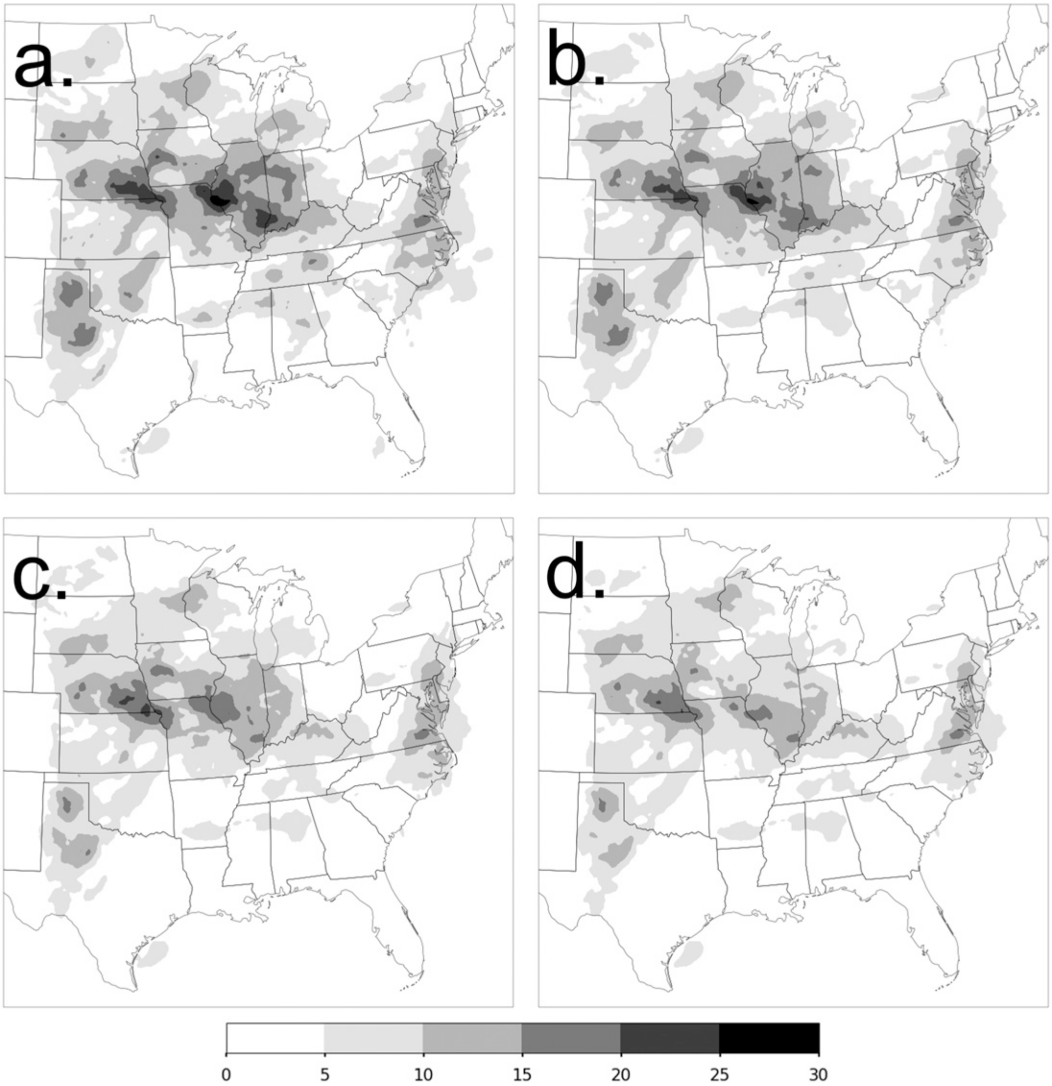

FIG. 12. Spatial occurrence (h) of nocturnal (0200–1100 UTC; Geerts et al. 2017) MCS swaths (minimum of 3 h) with a P_{MCS} of (a) 0.00, (b) 0.50, (c) 0.90, or (d) 0.95, occurring between 1 Jun and 15 Jul 2015 (CRSR is 24 km and SSR is 96 km).

classification, and tracking framework introduced in Part I. This second paper specifically focuses on the tracking portion of this framework. The specific goal of this work is to use the MCS slices generated in Part I to produce MCS swaths. MCS slices are associated with a number of attributes and are assigned a probabilistic classification value P_{MCS}, for which a value of 1 suggests that the slice is very likely to be an MCS slice and a value of 0 suggests that the slice is not likely to be an MCS slice. Using four probability thresholds (0.00, 0.50, 0.90, and 0.95), four CRSR values (6, 12, 24, and 48 km), and three SSR values (48, 96, and 192 km), a total of 48 perturbations are used to generate MCS swaths for the purposes of testing the sensitivity of the tracking procedure to these values (see Part I for more information on these values).

MCS swaths are generated through a two-step procedure. First, slices are matched using the spatiotemporal overlap technique. If more than one match was found, the Hungarian method (Munkres 1957; Dixon and Wiener 1993) is used to associate the most similar slices. Second, swaths that last at least 0.5 h are reanalyzed for the purposes of connecting multiple swaths together that are separated by brief (1 h or less) discontinuities. Subjective and objective assessments of tracking performance for each of the 48 perturbations are carried out to determine the optimal combination of the available parameter values. Performance is

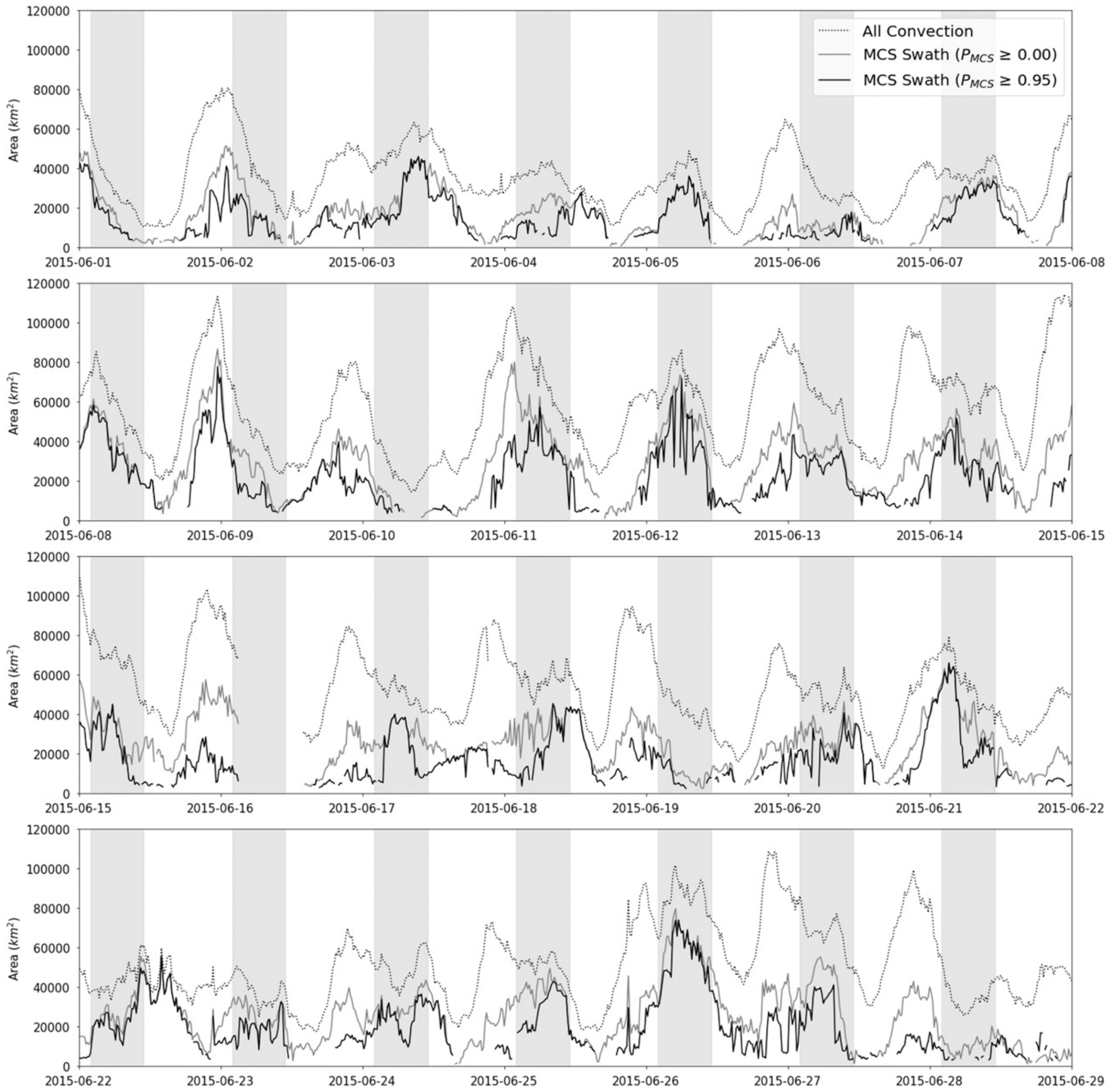

FIG. 13. The areal coverage of all convective pixels (dashed gray line), convective pixels within P_{MCS}-0.00 swaths (solid gray line), and convective pixels within P_{MCS}-0.95 swaths (solid black line) during June 2015. The darkened areas are times from 0200 to 1200 UTC (nocturnal; Geerts et al. 2017). Convective pixels are defined as pixels with intensities of ≥40 dBZ.

determined on the basis of three metrics: 1) mean swath duration, 2) mean standard deviation of reflectivity per swath (intensity error), and 3) root-mean-square error between centroid positions and a linear regression fit to all centroid positions in the swath (linearity error). The swaths are then used to generate a climatology for the 2015 and 2016 warm season. These results are then compared with external MCS frequency data to assess the level of agreement with existing research.

Subjective MCS swath accuracy varied among perturbations for the three cases examined. Overall, there was agreement between manual tracks and automatically generated tracks. One issue illustrated by the subjective assessment was that swaths generated using larger P_{MCS} values (i.e., 0.90 and 0.95) are sometimes incorrectly truncated (Figs. 3d and 4b). This is caused by the stricter thresholds removing all slices from spatiotemporal overlap consideration for a couple of radar images. When the slices regain the higher P_{MCS} values, the previous swath is already terminated, and therefore a new track is created. To address this issue, swaths lasting at least 0.5 h are reanalyzed to connect swaths together, as long as no more than 60 min have elapsed since swath termination. This step resulted in statistically significantly longer tracks ($p < 0.001$), which is considered to be a positive outcome in the context of storm-tracking performance (Lakshmanan and Smith 2010).

The objective assessment of tracking performance was completed by computing three key metrics outlined as important in previous work (Lakshmanan and Smith 2010): 1) mean swath duration (Fig. 6), 2) intensity error (Fig. 7), and 3) linearity error (Fig. 8). These values are calculated using reanalyzed swaths for all 48 perturbations. Mean swath durations for all tracks lasting at least 0.5 h are increased by using a P_{MCS} of 0.50 as compared with a P_{MCS} of 0.00. The reanalyzed swaths increased linearity error, which is likely due to abrupt changes in centroid location caused by associating slices that have been moving away from the location of terminated swaths for up to 60 min. Overall, the best-performing perturbation used a CRSR of 24 km and an SSR of 96 km (Tables 5 and 6; Fig. 9). As a group, the swaths for P_{MCS} of 0.50 and 0.90 had better performance metrics than the P_{MCS}-0.95 and P_{MCS}-0.00 swaths.

In general, the spatial frequency of MCSs that is presented in Figs. 10 and 11 agreed with previous work (e.g., Ashley et al. 2003; Anderson and Arritt 2001). Using a P_{MCS} threshold of 0.95 to generate swaths limits the area of relatively high MCS activity to the central and eastern Great Plains. Output from a 6-week period that overlapped with a field campaign described in Geerts et al. (2017) matched up well with their automated MCS frequency map (Fig. 12). The spatial structure of occurrence, as well as regional maxima and minima in MCS occurrence, is similar in the two datasets. These results are encouraging and suggest that the segmentation, classification, and tracking framework would be able to generate an accurate long-term, automated, climatology of CONUS MCSs.

Within the MCS literature, it is clear that, once convective clusters meet the objective PJ00 criteria, the subjective inclusion or exclusion of events is largely an ad hoc endeavor. The results presented by this study are based on the subjective assessment of convective clusters performed by the authors. To be specific, we do not claim that MCS swaths generated using higher P_{MCS} thresholds are more ''MCS like'' than other events—this designation only suggests that these events adhere more strongly to our mental schema of what constitutes an MCS. The ultimate goal of this work is not to provide a definitive definition of an MCS but rather to propose a framework for exploring an acceptable balance between probability of detection and probability of false detection for the particular task in which the data are being used. Future work should focus on improving the ability of computers to translate subjective expert classifications into accurate and reliable predictions on previously unseen data. Future gains in accuracy will likely require new image-classification techniques, such as convolutional neural networks (LeCun and Bengio 1995; Krizhevsky et al. 2012; Dieleman et al. 2015), that retain the spatial relationships of varied intensity within MCS slices. For this study, those relationships are largely lost when the MCS slice is reduced to 14 features (see Table 3 in Part I). Further, more-exotic tracking methods should be explored to improve tracking performance. For example, multiple hypothesis testing would be useful for determining the best spatiotemporal association to perform during merging or splitting events (Lakshmanan et al. 2013).

Acknowledgments. We thank Drs. Russ Schumacher (Colorado State University), Victor Gensini [Northern Illinois University (NIU)], David Changnon (NIU), Thomas Pingel (NIU), and Jie Zhou (NIU) for their suggestions and insight that improved the research and paper. In addition, we thank Dr. Wen-Chau Lee and three anonymous reviewers for their suggestions that greatly improved this paper. We also thank Arthur Person (senior research assistant in the Department of Meteorology at The Pennsylvania State University) for providing computational resources. This research was supported by National Science Foundation Grant ATM-1637225, an NIU Division of Research and Innovation Partnerships Research and Artistry Grant, and an NIU Graduate School Dissertation Completion Fellowship. This work used resources of the Center for Research Computing and Data at NIU.

REFERENCES

Anderson, C. J., and R. W. Arritt, 1998: Mesoscale convective complexes and persistent elongated convective systems over the United States during 1992 and 1993. *Mon. Wea. Rev.*, **126**, 578–599, https://doi.org/10.1175/1520-0493(1998)126<0578:MCCAPE>2.0.CO;2.

——, and ——, 2001: Mesoscale convective systems over the United States during the 1997–98 El Niño. *Mon. Wea. Rev.*, **129**, 2443–2457, https://doi.org/10.1175/1520-0493(2001)129<2443:MCSOTU>2.0.CO;2.

Ashley, W. S., T. L. Mote, P. G. Dixon, S. L. Trotter, E. J. Powell, J. D. Durkee, and A. J. Grundstein, 2003: Distribution of mesoscale convective complex rainfall in the United States. *Mon. Wea. Rev.*, **131**, 3003–3017, https://doi.org/10.1175/1520-0493(2003)131<3003:DOMCCR>2.0.CO;2.

Augustine, J. A., and K. W. Howard, 1988: Mesoscale convective complexes over the United States during 1985. *Mon. Wea.*

Rev., **116**, 685–701, https://doi.org/10.1175/1520-0493(1988)116<0685:MCCOTU>2.0.CO;2.

——, and ——, 1991: Mesoscale convective complexes over the United States during 1986 and 1987. *Mon. Wea. Rev.*, **119**, 1575–1589, https://doi.org/10.1175/1520-0493(1991)119<1575:MCCOTU>2.0.CO;2.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

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

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

Chang, W., M. L. Stein, J. Wang, V. R. Kotamarthi, and E. J. Moyer, 2016: Changes in spatiotemporal precipitation patterns in changing climate conditions. *J. Climate*, **29**, 8355–8376, https://doi.org/10.1175/JCLI-D-15-0844.1.

Chen, T., and C. Guestrin, 2016: XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, Association for Computing Machinery, 785–794, https://dl.acm.org/citation.cfm?id=2939785.

Clark, A. J., R. G. Bullock, T. L. Jensen, M. Xue, and F. Kong, 2014: Application of object-based time-domain diagnostics for tracking precipitation systems in convection-allowing models. *Wea. Forecasting*, **29**, 517–542, https://doi.org/10.1175/WAF-D-13-00098.1.

Corfidi, S. F., M. C. Coniglio, A. E. Cohen, and C. M. Mead, 2016: A proposed revision to the definition of "derecho." *Bull. Amer. Meteor. Soc.*, **97**, 935–949, https://doi.org/10.1175/BAMS-D-14-00254.1.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

Dieleman, S., K. W. Willett, and J. Dambre, 2015: Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Mon. Not. Roy. Astron. Soc.*, **450**, 1441–1459, https://doi.org/10.1093/mnras/stv632.

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.

Fabry, F., V. Meunier, B. P. Treserras, A. Cournoyer, and B. Nelson, 2017: On the climatological use of radar data mosaics: Possibilities and challenges. *Bull. Amer. Meteor. Soc.*, **98**, 2135–2148, https://doi.org/10.1175/BAMS-D-15-00256.1.

Fiolleau, T., and R. Roca, 2013: An algorithm for the detection and tracking of tropical mesoscale convective systems using infrared images from geostationary satellite. *IEEE Trans. Geosci. Remote Sens.*, **51**, 4302–4315, https://doi.org/10.1109/TGRS.2012.2227762.

Fritsch, J. M., and G. S. Forbes, 2001: Mesoscale convective systems. *Severe Convective Storms*, *Meteor. Monogr.*, No. 50, Amer. Meteor. Soc., 323–358, https://doi.org/10.1175/0065-9401-28.50.323.

Gagne, D. J., A. McGovern, 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.

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

——, and Coauthors, 2017: The 2015 Plains Elevated Convection at Night field project. *Bull. Amer. Meteor. Soc.*, **98**, 767–786, https://doi.org/10.1175/BAMS-D-15-00257.1.

Haberlie, A. M., and W. S. Ashley, 2018: Identifying mesoscale convective systems in radar mosaics. Part I: Segmentation and classification. *J. Appl. Meteor. Climatol.*, **57**, 1575–1598, https://doi.org/10.1175/JAMC-D-17-0293.1.

——, ——, and T. Pingel, 2015: The effect of urbanization on the climatology of thunderstorm initiation. *Quart. J. Roy. Meteor. Soc.*, **141**, 663–675, https://doi.org/10.1002/qj.2499.

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.

Hitchens, N. M., M. E. Baldwin, and R. J. Trapp, 2012: An object-oriented characterization of extreme precipitation-producing convective systems in the midwestern United States. *Mon. Wea. Rev.*, **140**, 1356–1366, https://doi.org/10.1175/MWR-D-11-00153.1.

Houston, A. L., N. A. Lock, J. Lahowetz, B. L. Barjenbruch, G. Limpert, and C. Oppermann, 2015: Thunderstorm Observation by Radar (ThOR): An algorithm to develop a climatology of thunderstorms. *J. Atmos. Oceanic Technol.*, **32**, 961–981, https://doi.org/10.1175/JTECH-D-14-00118.1.

Houze, R. A., Jr., 2004: Mesoscale convective systems. *Rev. Geophys.*, **42**, RG4003, https://doi.org/10.1029/2004RG000150.

Johns, R. H., and W. D. Hirt, 1987: Derechos: Widespread convectively induced windstorms. *Wea. Forecasting*, **2**, 32–49, https://doi.org/10.1175/1520-0434(1987)002<0032:DWCIW>2.0.CO;2.

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

Kolmogorov, A., 1933: Sulla determinazione empirica di una legge di distribuzione (On the empirical determination of a distribution law). *G. Ist. Ital. Attuari*, **4**, 83–91.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012: Imagenet classification with deep convolutional neural networks. *Proc. 25th Conf. on Advances in Neural Information Processing Systems*, Lake Tahoe, NV, Neural Information Processing Systems Foundation, 1097–1105, https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.

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

——, 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.

——, M. Miller, and T. Smith, 2013: Quality control of accumulated fields by applying spatial and temporal constraints. *J. Atmos. Oceanic Technol.*, **30**, 745–758, https://doi.org/10.1175/JTECH-D-12-00128.1.

——, B. Herzog, and D. Kingfield, 2015: A method for extracting postevent storm tracks. *J. Appl. Meteor. Climatol.*, **54**, 451–462, https://doi.org/10.1175/JAMC-D-14-0132.1.

LeCun, Y., and Y. Bengio, 1995: Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed., MIT Press, 255–258.

Maddox, R. A., C. F. Chappell, and L. R. Hoxit, 1979: Synoptic and meso-α scale aspects of flash flood events. *Bull. Amer. Meteor. Soc.*, **60**, 115–123, https://doi.org/10.1175/1520-0477-60.2.115.

McKinney, W., 2010: Data structures for statistical computing in Python. *Proc. Ninth Python in Science Conf.*, Austin, TX, SciPy, 51–56, https://pdfs.semanticscholar.org/f6da/c1c52d3b07c993fe52513b8964f86e8fe381.pdf.

Munkres, J., 1957: Algorithms for the assignment and transportation problems. *J. Soc. Ind. Appl. Math.*, **5**, 32–38, https://doi.org/10.1137/0105003.

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.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830, http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

Peters, J. M., E. R. Nielsen, M. D. Parker, S. M. Hitchcock, and R. S. Schumacher, 2017: The impact of low-level moisture errors on model forecasts of an MCS observed during PECAN. *Mon. Wea. Rev.*, **145**, 3599–3624, https://doi.org/10.1175/MWR-D-16-0296.1.

Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the High-Resolution Rapid Refresh Model's ability to predict mesoscale convective systems using object-based evaluation. *Wea. Forecasting*, **30**, 892–913, https://doi.org/10.1175/WAF-D-14-00118.1.

Przybylinski, R. W., 1995: The bow echo: Observations, numerical simulations, and severe weather detection methods. *Wea. Forecasting*, **10**, 203–218, https://doi.org/10.1175/1520-0434(1995)010<0203:TBEONS>2.0.CO;2.

Rodgers, D. M., M. J. Magnano, and J. H. Arns, 1985: Mesoscale convective complexes over the United States during 1983. *Mon. Wea. Rev.*, **113**, 888–901, https://doi.org/10.1175/1520-0493(1985)113<0888:MCCOTU>2.0.CO;2.

Skok, G., J. Tribbia, J. Rakovec, and B. Brown, 2009: Object-based analysis of satellite-derived precipitation systems over the low- and midlatitude Pacific Ocean. *Mon. Wea. Rev.*, **137**, 3196–3218, https://doi.org/10.1175/2009MWR2900.1.

Smith, J. A., D. J. Seo, M. L. Baeck, and M. D. Hudlow, 1996: An intercomparison study of NEXRAD precipitation estimates. *Water Resour. Res.*, **32**, 2035–2045, https://doi.org/10.1029/96WR00270.

Vila, D. A., L. A. T. Machado, H. Laurent, and I. Velasco, 2008: Forecast and Tracking the Evolution of Cloud Clusters (ForTraCC) using satellite infrared imagery: Methodology and validation. *Wea. Forecasting*, **23**, 233–245, https://doi.org/10.1175/2007WAF2006121.1.

Zipser, E. J., 1982: Use of a conceptual model of the life-cycle of mesoscale convective systems to improve very-short-range forecasts. *Nowcasting*, K. A. Browning, Ed., Academic Press, 191–204.