UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

MULTIPLE HYPOTHESIS TRACKING AND STRONG POINT ANALYSIS FOR STORM TRACKING WITH WEATHER RADAR

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
MASTER OF SCIENCE

By

MATTHEW JUDY
Norman, Oklahoma
2020

MULTIPLE HYPOTHESIS TRACKING AND STRONG POINT ANALYSIS FOR STORM TRACKING WITH WEATHER RADAR

A THESIS APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Tian-You Yu, Chair

Dr. Boon Leng Cheong

Dr. Mark Yeary

Table of Contents

List of Tables

List of Figures

Abstract

Storm tracking using radar observations of weather is a valuable tool both for the study of weather conditions and for the issuance of advance warning when such conditions are severe. Centroid-based storm tracking algorithms are a common tool used to provide short-term forecasting of storms. In these algorithms, cells representing intense areas of storms are identified. The cells are compared to cells identified at regular intervals, and similar cells are assigned to tracks which indicate the history of the underlying storm. These tracks may then be used to generate forecasts. The Strong Point Analysis, Multiple Hypothesis Tracking (SPA-MHT) technique is a centroid-based algorithm which draws on methods effective in storm tracking and in the more general field of object tracking. Strong Point Analysis identifies storm cells using the erosion and dilation operations from the field of mathematical morphology. Multiple Hypothesis Tracking assigns these cells to tracks, resolving ambiguities by incorporating data from multiple sequential sets of observations.

In this thesis, the motivation for the design of SPA-MHT is presented in the context of previous storm tracking algorithms. The suitability of Strong Point Analysis for storm identification and Multiple Hypothesis Tracking for track association is considered. An implementation of each technique is also presented. The combined implementation is evaluated on radar data from the NEXRAD network of Doppler weather radars. This constitutes the first evaluation of Multiple Hypothesis Tracking on actual weather radar data. Both qualitative visual comparisons and quantitative comparisons are considered. SPA-MHT is found to outperform the current NEXRAD tracking algorithm when tracking storms that are isolated or forming clusters, but performance between the two methods is more comparable in challenging situations such as squall lines.

Chapter 1

Introduction

Nowcasting, or short-term forecasting of weather phenomena, is a valuable tool in the study of weather and in issuing warnings of severe and hazardous events. Accurate nowcasting can be combined with rainfall estimation to generate advance warning for flooding [1]. The ability to identify the motion of storms is a prerequisite to effective nowcasting, yet many challenges are still unaddressed by modern storm tracking techniques. Variation in the shape and motion of tracked weather limits the effectiveness of tracking techniques. By integrating advances in the image processing and hard target tracking domains, the main goal of this work is to improve the performance of existing storm trackers.

The basic unit of data required by storm tracking algorithms is a scan: a set of observations from a radar. An algorithm may require scans in different formats, such as the plan position indicator, which traverses range and azimuth dimensions, or the volume scan, which collects data from sweeps of range, azimuth and elevation. Regardless of format, each scan is associated with a single time.

The specific definition of a storm varies between storm tracking techniques. Considering the stated intent to provide useful nowcasting of weather patterns, this thesis will refer to any cohesive area of moderate to high reflectivity as a storm; the entities defined by different techniques will be referred to as storm cells. Generally, the storm cells identified will provide different interpretations of the same underlying storm.

1.1 Review of Existing Storm Tracking Algorithms

Approaches to the storm tracking problem may be grouped into two main categories [2]. The first category consists of storm trackers which cross-correlate data from two scans to estimate the motion between the scans. This procedure estimates the general motion of all regions of high reflectivity in an image rather than the motion of any individual cell [3]. Generally speaking, correlation-based methods are vulnerable to storm growth and decay, especially at smaller scales, when changes in storm intensity between scans reduce correlation. The Tracking Radar Echoes by Correlation (TREC) algorithm illustrates this category of trackers. TREC generates motion vectors by searching for regions in a new reflectivity image which are strongly correlated with a region in the previous image [4]. The search covers an area based on known wind speeds, which provide an upper bound on the amount of motion expected. A motion vector is obtained for each point in a grid. No attempt is made to identify distinct cells. TREC was later applied to hurricanes, which feature well-defined structure, in order to provide continuous estimates of peak and sustained wind speeds [5]. The movement of regions, as detected by correlation, traced wind fields within the hurricane wall.

Algorithms in the second and more broad category, the centroid-based storm trackers, operate on sets of point observations, or centroids. A feature extraction process is required to translate the widely distributed reflectivity fields observed by a weather radar into these centroids before they are assigned to tracks. Among the centroid trackers, the average position $(\bar{x}, \bar{y})$ of reflectivity values included in a

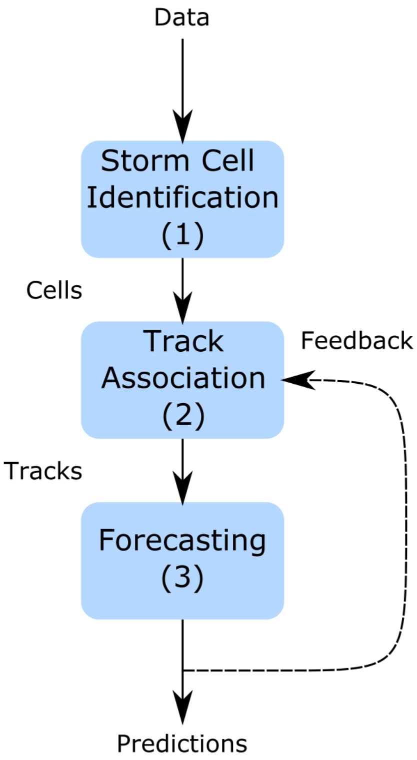


Figure 1.1: Flowchart illustrating three stages of centroid-based storm trackers. Storm cells are identified from incoming data (1) and assigned to new or existing tracks (2), which are then used to generate forecasts (3). Some techniques use forecasts as feedback to improve the track association stage.

storm cell is a commonly used centroid. Alternatives such as the coordinates of the reflectivity maximum within the cell have been used. A centroid-based algorithm may be divided into three stages beginning with the arrival of a scan. First, storm cells are identified from the input data. Each cell is then represented by a single set of observations: the centroid. The second stage associates storm cells between scans, forming tracks. Association is based on some combination of the properties of the cell. In the third stage, each cell in a track is used to forecast positions at which storm cells are expected in future scans. These forecasts may be used as a feedback input into the track association stage. Figure 1.1 illustrates the three stages of centroid trackers.

In [3], centroids are extracted by identifying contours of consistent reflectivity in radar images. To reduce the amount of memory used, contours are extracted as each radial is recorded, and attributes such as area, average and peak reflectivity, and centroid location are accumulated. A secondary, adaptive threshold extracts cell contours from these outer contours set 3 dB below the peak value within the outer contour. The resulting cells are compact, only including measurements that are spatially very close to the maximum. Track assignment is managed by comparing cells from the incoming scan with the extrapolated position of previous tracks. Cells in suitable positions, but with significant differences in height, area, or intensity, are rejected. The forecasting stage of this algorithm is limited to application of a wind estimate to previous centroid positions as feedback to the track association stage; nowcasting is not attempted.

A later technique applied pattern recognition to track a larger portion of the volume of convective cells [6]. Cells are defined as any region of nonzero reflectivity with only a single maximum. The coordinates of this maximum are considered the centroid of the cell. Two cells are separated by the minimum reflectivity values between them, resulting in wide cells. By including reflectivity beyond only the centroid, this technique supports secondary products of the cells such as total rainfall yield. Cells in the new scan are translated to estimate their historical position. All cells are translated according to a mean motion estimate. Initially, this estimate is based off of radiosonde estimates of wind at the steering level. A new cell is assigned to a previous cell if the historical position estimate has significant overlap with the previous cell. Merges and splits are considered if multiple cells overlap. After each cell is matched or considered a new track, the average distance between the centroids in the old scan and the estimated previous centroids of the new scan is used to update the mean motion estimate. This technique was considered unsuitable for stratiform precipitation, decaying convective systems, and rapidly changing storm cell shapes; each of these cases violates the algorithm's expectations of cells with a clear central reflectivity peak. Computer memory limitations were also noted as a weakness of this technique.

The Thunderstorm Identification, Tracking and Nowcasting (TITAN) algorithm also tracks centroids derived from reflectivity images [7]. Any pixels with reflectivity values above 35 dBZ are considered to be members of some storm cell. Adjacent storm pixels are grouped into storm cells, and for each distinct storm cell, the ellipse which best fits its member pixels, weighted by reflectivity, is calculated. Each ellipse is considered to represent a new observation. The center coordinates of the best fit ellipses are the centroids. A cost function measures the similarity of each observation to the last observation in each existing track in terms of difference in centroid position and difference in volume. The total cost of assigning each observation to one and only one existing track is minimized to determine the best extension of each track. Forecasting is based on a linear regression of positions in a track, with the most recent positions receiving the most weight. While not used as feedback for the main track association process, track forecasts are an input to an additional matching process which attempts to identify storm merges and splits.

The Storm Cell Identification and Tracking (SCIT) algorithm is a centroid-based tracker designed to support tracking of storm lines and clusters [2]. Cells are defined as regions of high reflectivity across several range bins, scan angles, and scan elevations. SCIT identifies cells at multiple fixed thresholds of reflectivity. When a cell is identified using a higher threshold within a cell found at a lower threshold, only the cell at the higher threshold is retained. Track association is based on agreement between new cells and extrapolated positions of previous tracks. A least-squares regression of previous locations in a track provides its motion estimate for extrapolation. SCIT is vulnerable to storm growth and decay and is ineffective at tracking stratiform precipitation, a weakness common to the cell-based trackers reviewed. However, the performance of SCIT on storm cell centroids proved strong enough for adoption by the National Weather Service for the Next-Generation Radar (NEXRAD) system. SCIT is currently used with the data provided by the WSR-88D radar network to generate NEXRAD Product 58 - Storm Tracking Information [8].

Related to the centroid trackers are expert systems which make use of storm evolution models in order to assign tracks and generate forecasts. The Generating Advanced Nowcasts for Deployment in Operational Land Surface Flood Forecast (GANDOLF) system, developed to provide flood hazard warnings, maintains an object-oriented model for each storm cell detected [9]. Like the centroid trackers, GANDOLF is divided into cell detection and forecasting stages, but neglects to associate cells to tracks. Forecasts are instead derived from the state of each storm cell combined with an external source of wind estimates. While centroid trackers like TITAN represent cells with little information other than a centroid, GANDOLF uses a conceptual model of the cloud which evolves over time. Initiating such storm models requires an extensive set of inputs including satellite imagery and mesoscale weather modeling data. The added information enables nonlinear growth and decay forecasting unavailable in algorithms such as SCIT or TITAN.

Another expert system, the Auto-Nowcast System (ANC), combines the approaches of correlation and centroid trackers with additional boundary layer identification [10]. ANC identifies and forecasts storm cells according to the TITAN algorithm, but extracts a motion field by applying the TREC procedure to the data rather than by associating cells. Boundaries between regions of the atmosphere are identified, mainly using cloud information from satellite imagery. The identified boundaries modify the predicted behavior of cells: cells moving with a boundary are considered likely to be sustained, while boundaries leaving storm cells behind indicates that the cells are predicted to decay. These forecasts and further data sources, such as radiosonde measurements and lightning data, are integrated into likelihood fields, which indicate where storms are expected to be present, using fuzzy logic. ANC was evaluated alongside GANDOLF, TITAN, and other techniques on data from the Sydney, Australia region [1]. The study concluded that forecasts based only on extrapolation could be outperformed with consideration of boundary characteristics. The complexity and intensive data source requirements of effective expert systems, however, are significant.

1.2 Addressing Challenges to Storm Tracking

In the general case of object tracking from a sequence of images, such as video, factors which may challenge a tracking algorithm include the loss of depth information when capturing images, noise, irregular motion, nonrigid objects, occlusion by other objects, complex object shapes, and scene illumination changes [11]. Many of these apply to storm tracking. Storm cells may move irregularly and change shape as they grow, decay, and merge or split. Cells are nonrigid objects and, depending on the definition used, may have complex shapes. The loss of depth information is not implicit in storm tracking, as storm trackers may make use of the full three dimensions of radar data, but height information is often discarded during the tracking process, as is the case in SCIT and TITAN. In video data, changes in external illumination may correspond to changes in the recorded image even when the objects observed remain constant. A radar, which controls the source of illumination by the same transmitted energy and scan strategies from scan to scan, has better

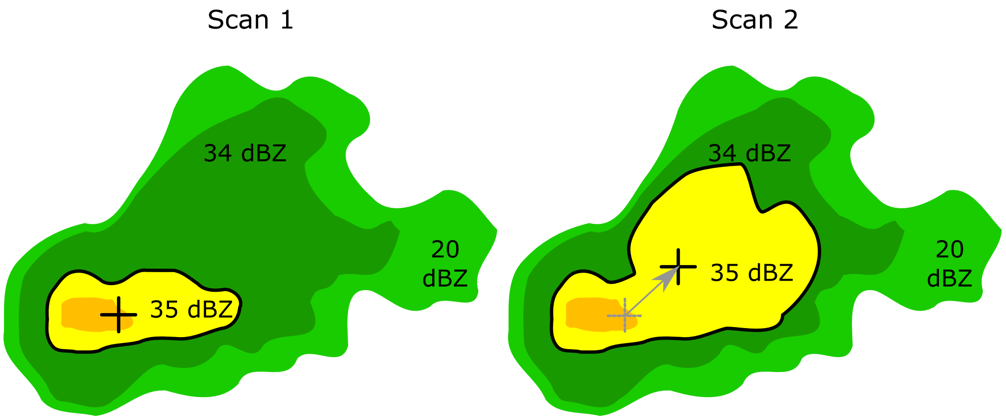


Figure 1.2: Two illustrations of radar reflectivity scans representing a centroid motion artifact using an inconsistent storm identification technique. A region changes from 34 to 35 dBZ between the scans. The simple threshold, drawn in black, expands greatly as a result. A small change in reflectivity values moves the centroid along the grey arrow, even though the rest of the cell has not moved.

assurance that changes in the received data are due to changes in the targets.

Changes in the shape and intensity of storm cells are a substantial source of artifacts interfering with effective storm tracking [12]. When shape and intensity changes are interpreted as storm motion, the assumptions of centroid-based tracking are violated. For example, the simple 35 dBZ threshold used in TITAN leaves the cell identification stage vulnerable to variation in the reflectivity field. As shown in Figure 1.2, a minimal growth in reflectivity can shift a storm centroid even when most of the reflectivity included in the cell has not moved. The change in cell shape, and thus the change in centroid, is disproportionate to the slight change in reflectivity–the cell is inconsistent. Tracking based on storm centroids assumes that centroid motion is representative of storm motion, not storm growth and decay. Consistent cells, for which this assumption holds, are required.

Strong Point Analysis (SPA) is an image segmentation algorithm designed to address this problem by identifying consistent cells for storms [13]. SPA utilizes erosion and dilation techniques to extract clusters which represent the most intense and spatially dense regions within the image. By extracting areas of locally high intensity, SPA confers resistance to artifacts caused by minor variation in reflectivity. Configuration options for the algorithm also support the splitting of complex, irregular cell shapes into smaller, more round clusters better suited to representation by a centroid. The consistency of SPA cells leads to consistent centroids which lead in turn to tracks which capture the motion of the underlying storm rather than minor variation in reflectivity.

The track assignment and forecasting stages of centroid trackers are also ripe for improvement. Linear regression forecasting, as used by SCIT, may be outperformed by forecasting based on Kalman filters [14]. In algorithms where the forecasting stage provides feedback into the track association stage, improvement in forecasting is further incentivized.

A candidate for implementing these improvements in storm tracking is Multiple Hypothesis Tracking (MHT), a multiple target tracking technique dominant in the defense and surveillance fields [15]. Tracks are assigned based on agreement with a Kalman filter derived estimate of the track's next position. The state estimate is then updated according to the new position assigned to the track. Whenever the track to which an incoming observation should be assigned is ambiguous, rather than select one track, MHT records each plausible track assignment as a separate hypothesis. The plausibility of each hypothesis is estimated and, as more data is collected, the most plausible hypotheses are each considered as previous states for child hypotheses. This allows uncertainties in track assignment to be resolved by future data. MHT is presented in greater detail in Chapter 3.

In this thesis, SPA and MHT are combined to create a storm tracker which mitigates the drawbacks of existing trackers. The performance of this tracker is evaluated on archive data from the NEXRAD network of Doppler weather radars.

This thesis begins with examination of the SPA storm identification algorithm in Chapter 2 and the MHT algorithm for track assignment in Chapter 3. The structure and performance of the combined SPA-MHT tracker is evaluated in Chapter 4. Finally, Chapter 5 presents conclusions drawn from the completed work as well as recommendations for future work.

Chapter 2

Strong Point Analysis

Storm identification is the first stage in tracking storm cells by their centroids. Starting from a scan arriving from the radar, this process provides a set of observations representative of the entities to be tracked. Strong Point Analysis (SPA) incorporates image processing techniques from mathematical morphology in order to more consistently represent patterns in the radar reflectivity as storm cell observations. This chapter addresses the operation of SPA in the context of other storm identification stages.

2.1 Background of Storm Identification

A selection of historical approaches to centroid-based storm tracking were presented in Chapter 1. Further detail on the identification stages used by these approaches is collected in this section. Early algorithms emphasize the definition of centroids as local reflectivity maxima rather than average positions. This is the approach taken in [3], which identifies contours at a fixed reflectivity thresholds below each local maximum. Contours which would enclose an existing cell found from a greater local maximum are rejected. With an average diameter of 3 km, such narrowly defined cells are well represented by a coordinate pair, as most nonzero reflectivity in the margins of a storm is not included within any cell.

Such narrowly defined cells are insufficient, however, to describe the motion and properties of the storms they represent [6]. An alternative identification stage also begins with local reflectivity maxima, but divides all nonzero measurements into the cells. Each cell contains one local maximum and all measurements which are not separated from that maximum by any local minima. This wider inclusion of reflectivity enables the tracking of additional statistics beyond centroid position, including total rain rate of the cell, area of the cell's cloud base, and parameters describing the cell's echo top. Rapidly changing storm cells, including those undergoing growth or decay, pose problems for this technique, as they are less likely to maintain a single peak value. Stratiform precipitation, which is widely distributed around inconsistent reflectivity maxima, also violates the assumptions of this identification algorithm.

The extended watershed algorithm is a more recent approach to this definition of storm cells [16]. A cell identified by this method consists of a basin, pixels in the immediate area around a reflectivity maximum. Each cell also is also assigned foothills, pixels more distant from the maximum which are excluded from membership in other cells. The resulting cells are more inclusive than those identified by the fixed threshold method, but are not required to include every nonzero reflectivity value. Instead, basins are extended by adding pixels with progressively lower reflectivity until a sufficient area is included in the cell. This algorithm also begins with a smoothing step to increase cell consistency by reducing the sensitivity of the algorithm to spurious peaks in reflectivity.

SCIT defines cells as regions of high reflectivity across several range bins, scan angles, and scan elevations [2]. The reflectivity thresholds used by SCIT are not adaptive, but seven are applied to identify storm components. The strongest components are combined into cells, and the average component positions, weighted by reflectivity, are recorded as centroids. A single isolated component is insufficient to form a cell, filtering out particularly weak regions. Peak reflectivity within each cell is tracked, but does not restrict identification of the cell. Other parameters such as vertically integrated liquid are recorded as well for use in forecasting. SCIT is vulnerable to track growth and decay and is ineffective at tracking stratiform precipitation. When compared to previous algorithms used on the same radar system, SCIT demonstrated improvement in storm cell identification, a result attributed to the use of multiple thresholds.

GANDOLF is unique among the surveyed methods for identifying cells differently according to their estimated stage of development [9]. For example, the presence of a strong reflectivity maximum low in the cloud is likely to result in an identified mature cumulonimbus cell, while a strong vertical reflectivity gradient within the lower half of the storm suggests a younger cell which is expected to grow [17]. The other cell stages identified are towering cumulus clouds, weakening cumulonimbus, and dissipating storms with downdrafts. This detailed identification process requires extensive data inputs. Determination of the cloud base relies on an external mesoscale weather model. Visible and infrared satellite data is also incorporated into the identification process, and to align with these images, the radar data is interpolated to a Cartesian grid. Each pixel is labeled as a likely candidate for one of the five stages of storm development based on the radar, satellite, and weather model inputs. Square templates are then selected from the resulting image; if the sum of pixel values within a template falls within the expected range for a development stage, that template is considered an identified cell of that development stage. Cells according to this definition are always square and may contain pixels with no reflectivity.

TITAN is also applied to radar data, converted to a Cartesian grid. It also uses one of the simplest identification stages considered: any reflectivity values above a fixed threshold are considered part of some storm cell, with no secondary processing to introduce relative or multiple thresholds [7]. Adjacent pixels are grouped into separate storm cells. For each distinct cell, an ellipse of best fit is constructed. The center points of the best fit ellipses are the centroids to be tracked. The same technique used in TITAN identifies storm cells in the ANC system [10].

A key weakness of the TITAN algorithm is vulnerability to shape changes and merges in the cells identified [12]. Figure 2.1 illustrates a ”false merger” resulting from minor changes which raise pixel values above the constant 35 dBZ threshold. Two distinct regions of high reflectivity are linked by a weak or narrow region which satisfies the conditions required to identify the two storms as one cell. The identification of two distinct cells is suppressed by the single-threshold limitation. This problem was addressed in the Enhanced TITAN (ETITAN) algorithm by a modified identification stage utilizing the erosion and dilation operations from mathematical morphology. Erosion is defined as the removal of pixels from a set according to the values of neighboring pixels; dilation is the addition of pixels in a similar manner [18]. In ETITAN, these operations change the number of pixels included in an identified storm cell while preserving information about the shape of the cell. The storms detected according to the normal identification of TITAN are eroded to remove pixels bridging storms in false merger cases. After stronger storms are identified within the eroded storms, dilation and a second application of erosion further refine the storm cells [12]. The pixels which are included in these final cells are considered more essential to the underlying storm and less likely to be spurious.

Several of the surveyed techniques included an adaptive threshold, either explic-

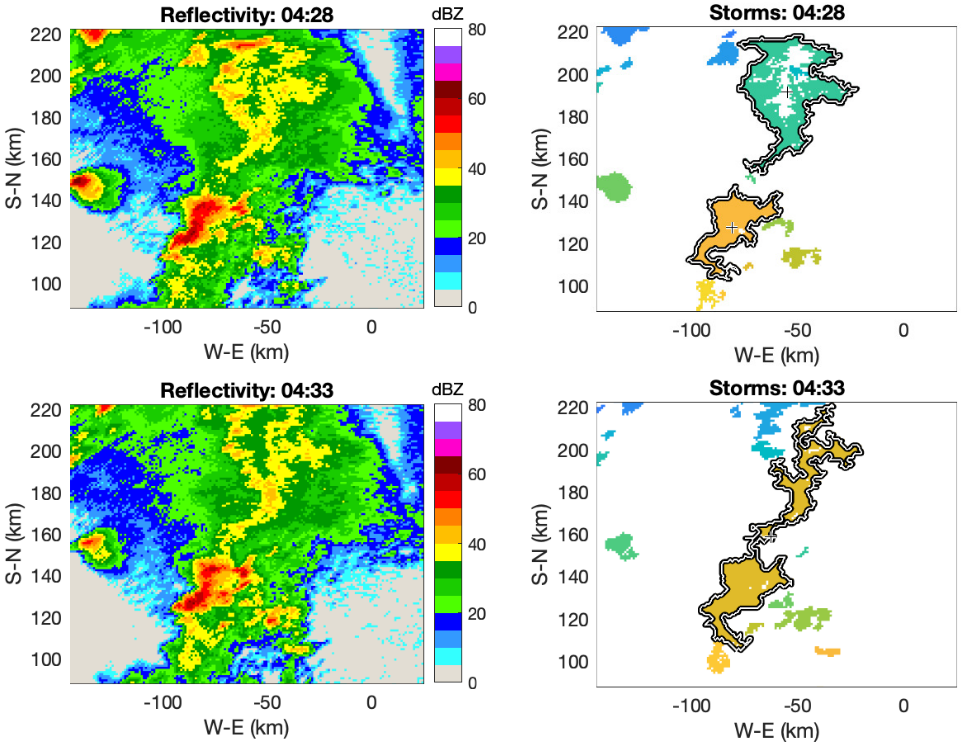


Figure 2.1: A false merger occurs between two scans from the KICT radar in Wichita, KS. Top left: Reflectivity from the scan beginning at 2008/05/24 04:28 UTC. Two large regions exceeding 35 dBZ are visible: an intense storm in the bottom center and a weaker but more widespread storm in the top center. Top right: cells identified by an implementation of TITAN's 35 dBZ threshold. The two storms are identified separately. Each centroid is within the bounds of its cell. Bottom left: Reflectivity from the subsequent scan at 2008/05/24 04:33 UTC. Minimal change is visible in either storm. Bottom right: cells identified by the same technique. A few pixels now bridge the gap between the two storms, allowing TITAN to identify one elongated cell. The resulting single centroid is over 30 km from either previous centroid.

itly or by use of erosion and dilation. The extended watershed algorithm identified cells with consistent area by dilating basins. SCIT, using multiple fixed thresholds and retaining the highest, incorporated erosion. TITAN does not, but ETITAN added erosion and dilation to avoid errors. Adaptive techniques appear advantageous for cell identification due to improvement in cell consistency. This is supported by the importance of consistent features in the general case of tracking objects through video data [11].

In an attempt to provide consistent storm cell identification for storm tracking, SPA was developed [13]. SPA, like TITAN, operates on a Cartesian coordinate system and requires no input products other than reflectivity. The use of an adaptive threshold through erosion and dilation operations addresses weaknesses of TITAN. The structure of SPA incorporates both the strong, high-intensity pixels surrounding a reflectivity maximum and the weaker pixels which provide consistency to the resulting clusters.

2.2 Algorithm

SPA is a recursive application of erosion and dilation procedures. The inputs to SPA are a set of pixels forming an image. Each pixel is defined by a pair of coordinates (x, y) and the intensity value corresponding to that location in the image.

The behavior of SPA is controlled by five parameters. U, the upper sensitivity parameter, and L, the lower sensitivity parameter, affect both procedures. Dilation additionally depends on the pixel reach parameter r and, in some cases, the pixel promotion parameter p. The maximum number of recursion levels at which the procedures are applied is limited by n. The maximum number of recursion levels n is used to limit applications of the recursive procedures.

The erosion procedure is applied to a set S of image pixels as defined above. First, the mean μ and standard deviation σ of pixel intensities in S are calculated. Using these values, pixels are tested for inclusion into two categories. Those with intensities greater than or equal to an upper threshold

$$Z_{\text{upper}} = \mu + U\sigma \tag{2.1}$$

are flagged as strong points. If Z_{upper} would exceed the intensity of every pixel in the set, as may occur in an extremely smooth sample, Z_{upper} is instead selected to include the maximum intensity in the set as a strong point. At least one point is always considered strong. Of the remaining points, those with intensities greater than or equal to a lower threshold

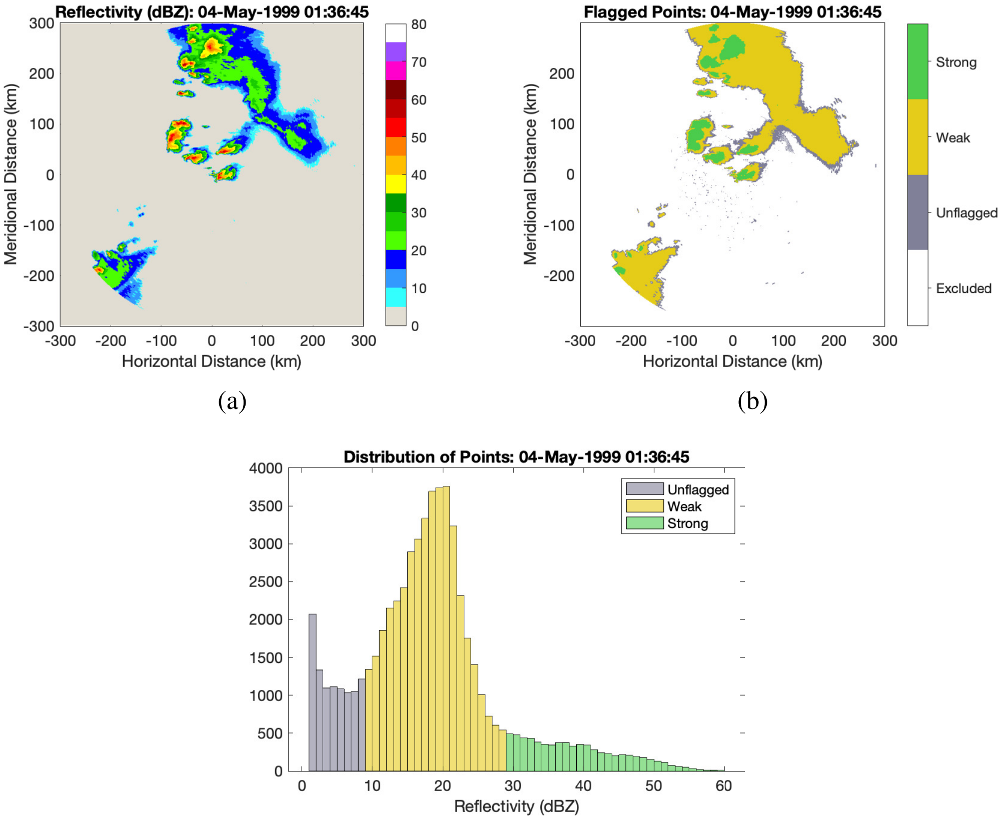
$$Z_{\text{lower}} = \mu - L\sigma \tag{2.2}$$

are considered weak points. Points with intensities below Z_{lower} are not assigned either flag. A sample distribution of values divided into strong, weak and unflagged points, is provided in Figure 2.2.

In the promotion step, unflagged points which are sufficiently surrounded by many strong points are flagged as weak. Such a point is promoted if its enhanced value Z_{en} exceeds Z_{lower}. For a point with actual intensity Z_{actual} and k strong points located $d_{\text{k}} < r$ pixels from the point,

$$Z_{\text{en}} = Z_{\text{actual}} + p \sum_{k=1}^{n} \frac{1}{d_k}. \tag{2.3}$$

This process is related to dilation and allows the subsequent clustering step to fill in very small gaps in the image. The enhanced value is only used for considering



(a)

(b)

(c)

Figure 2.2: (a): Reflectivity observed within 300 km of the KTLX radar on May 4, 1999 at 01:36 UTC. (b): Nonzero points from the reflectivity image are colored according to the flag assigned during the application of SPA. Selection of the intense, high-reflectivity regions as strong points is visible. Strong regions are surrounded by many more weak points. Weak returns, left unflagged, are eroded. (c): Histogram showing the nonzero reflectivity values. Points are marked as unflagged, weak, or strong according to their value relative to Z_{lower} and Z_{upper}. Thresholds were determined with parameters $U = 1.1$ and $L = 0.9$.

promotion and has no meaning in terms of the original image intensity, as it is based on inverse distance. An unflagged point may be promoted to a weak point but never a strong point.

Once points are flagged, clusters are formed by connected-components labeling. Pixels are considered connected if the distance between their measurements in units of pixels is less than r. Any strong point may be included in a cluster. The resulting clusters are eroded representations of storm cells, containing only the strongest points in the set. Strong points then propagate their label to any connected weak point. Weak points cannot further propagate the cluster; they may only be used to fill out clusters seeded by at least one strong point. This step is analogous to a dilation process and supports the formation of more regular clusters. Some amount of resistance to noise is also conferred by this step. Even if a given pixel appears weak instead of strong, as long as nearby pixels are still considered strong, it will be included in the appropriate cluster.

Erosion and dilation are applied recursively to each cluster of sufficient size found by the connected-components method. The pixels labeled in a single cluster are considered the new set of pixels from which μ and σ, and thus Z_{lower} and Z_{upper}, are calculated. Recursion is not applied to clusters which contain five or fewer pixels or clusters which were found at a recursion depth equal to n. In either case, the cluster found after dilation is reported in the final output of the algorithm. The final product returned by SPA is a distinct label value assigned to each continuous region identified within the input set of pixels after $n + 1$ applications of the erosion and dilation procedures. Pixels left unflagged at any recursion depth are labeled with a zero to indicate that they belong to no storm cell.

2.3 Implementation

The SPA technique was implemented from scratch in MATLAB. The initial set of pixels considered is all non-zero reflectivity values in a single scan of radar reflectivity data. Because pixel distances are important to the operation of the algorithm, the radar data is first converted from spherical to Cartesian coordinates by inverse distance weighting interpolation via a Python script. Interpolated values more than 300 km from the radar are not calculated. The selected resolution for the Cartesian coordinate grid is 1.0 km in both the horizontal and meridional directions. An example of spherical and corresponding Cartesian data is shown in Figure 2.3.

The (x, y) coordinates of each nonzero reflectivity value in the image are used as inputs to the algorithm. A list of the corresponding reflectivity values are provided as the intensity input. The output is a label for each value or zero for values which were determined to belong to no storm. Pixels with an reflectivity of zero are not included in the input set in order to reduce the computational load of the algorithm; these pixels are automatically removed from consideration as storm cells.

Values for the five SPA parameters were initially selected according to the conclusions of Root *et al.* [13]. Common weather reflectivity ranges are captured with $U = 1.1$ and $L = 0.9$. A moderate level of smoothing of the clusters is provided by $r = 5.1$ and $p = 2.5$. The number of recursion levels was increased from 1 to 2 based on a preference for more compact storm cells. Figure 2.4 illustrates the decision. Several of the larger cells detected at $n = 1$ are divided into more regular cells when n is increased.

Separately, alternate values of U were evaluated in order to improve compatibility with MHT. The initial value $U = 1.1$ in was unsatisfactory to visual inspection, failing to split some large or irregular storms into multiple cells. The intense area

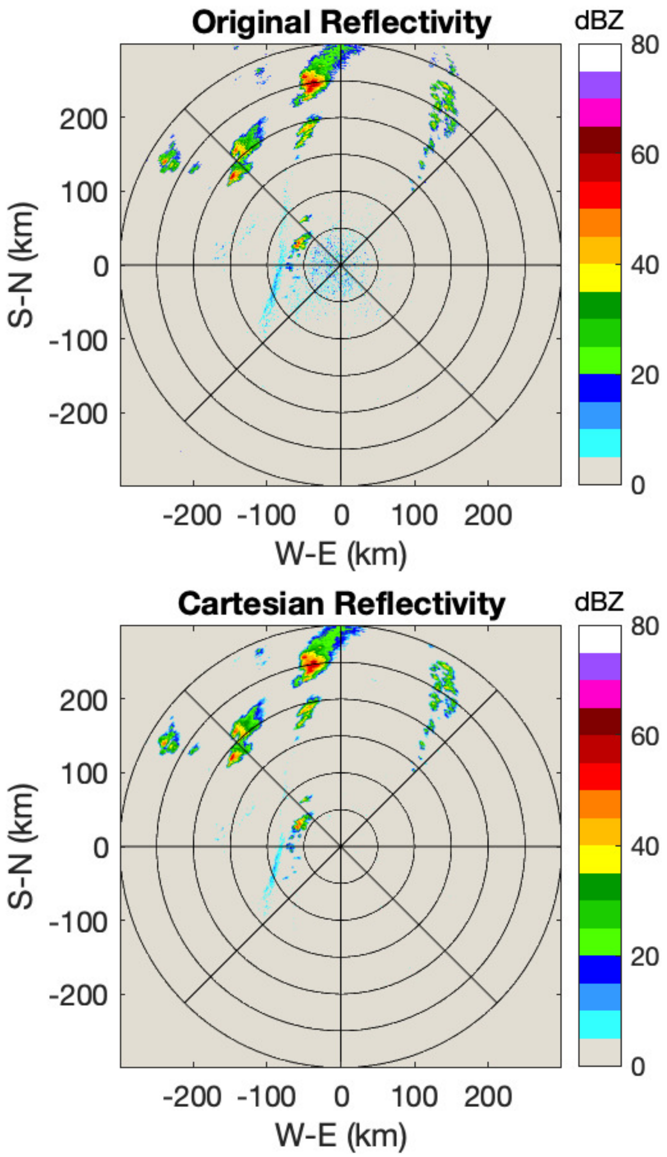
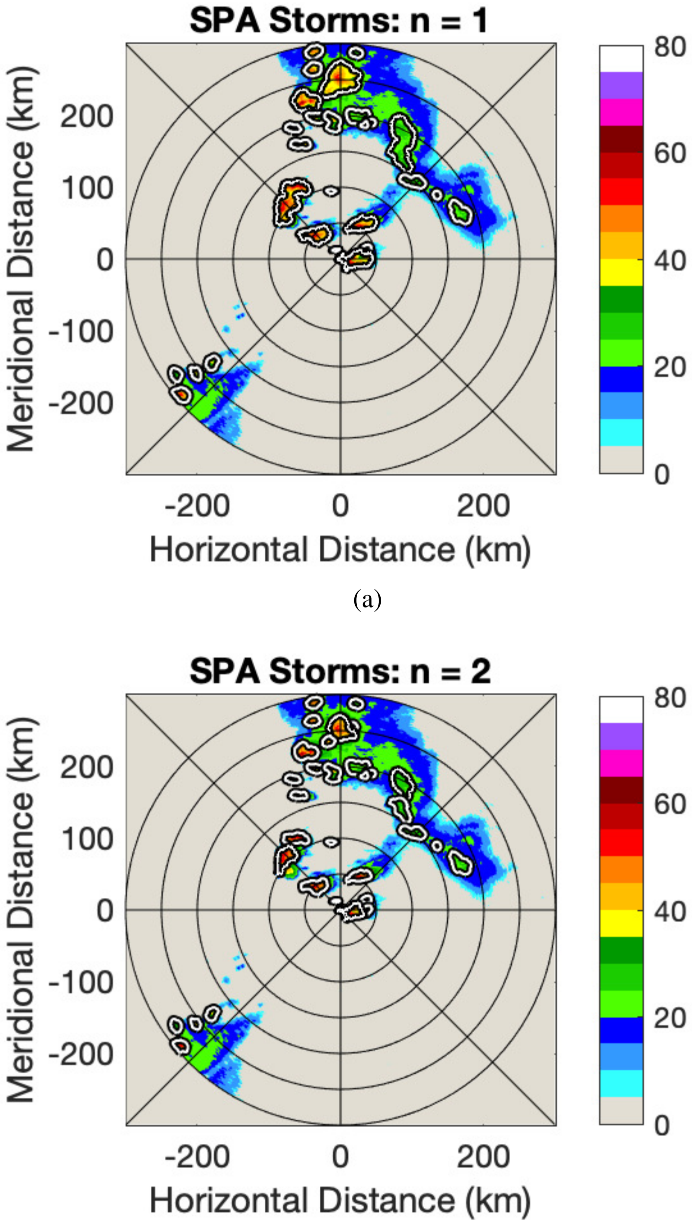


Figure 2.3: Comparison of original spherical and interpolated Cartesian radar data. Top: The 0.5 degree elevation scan from the KTLX radar in Oklahoma City, OK at 20:02 UTC on 2013/05/19. Bottom: Interpolation of the same data to a plane 1 km above the radar. Rings are spaced every 50 km from the radar. The images are very similar. The center clutter region is missing in the Cartesian data because it is below the plane.



(a)

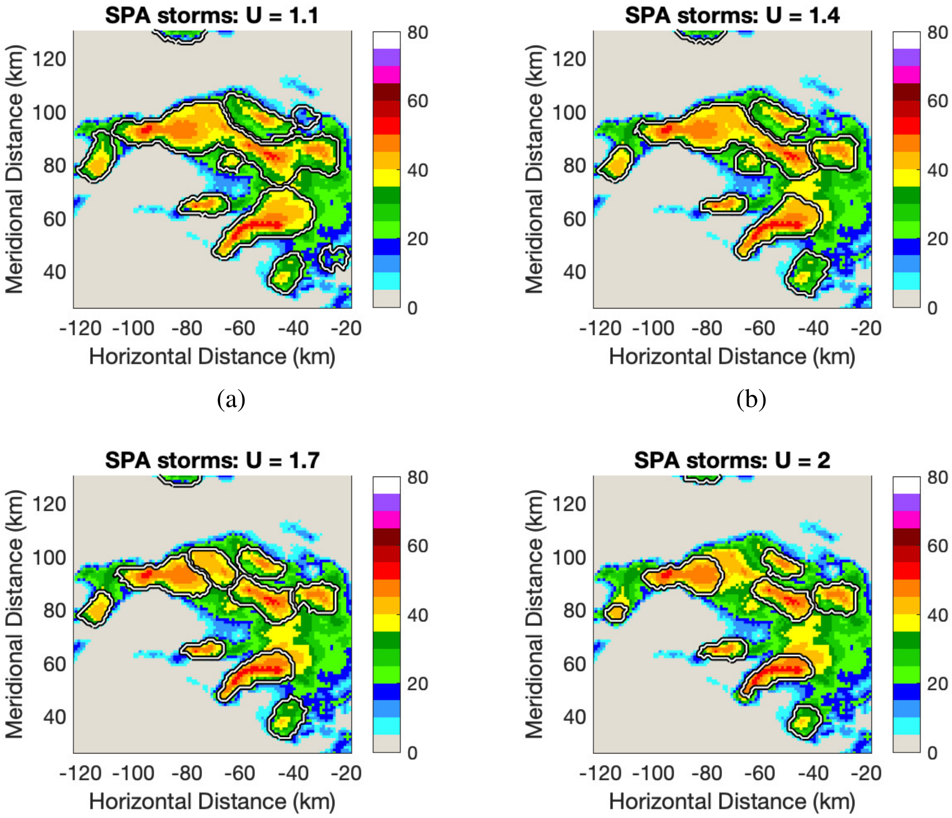
(b)

Figure 2.4: Reflectivity observed within 300 km of the KTLX radar on May 4, 1999 at 01:36 UTC, with detected SPA storm cells outlined in white. Cells detected using one level of recursion are shown in (a), while those detected after two levels are shown in (b). Both sets cover similar areas, but the larger cells in (a) are split into two or three cells in (b). For both cases, $U = 1.1$, $L = 0.9$, $r = 5.1$, and $p = 2.5$.

at the focus of Figure 2.5 is an example of this failure; in subfigure *(a)*, SPA using $U = 1.1$ considers this a single wide storm. As U increases, multiple cells are detected in the same region of reflectivity. Changing U does not change the set of point which are not assigned a flag, but it does move some points from the strong to weak set. If the points moved were the only strong points near other weak points, the latter points will no longer be included in the cluster. This has the effect of destroying weaker bridges between the strongest regions in the image. This is conducive to avoidance of the false split problem as the centroids of these cells are now more representative of the strongest reflectivity values rather than a weighted average which falls over a weaker, bridging region.

Reducing the number of strong pixels and thus the number of pixels included in any storm does reduce the sample size of points constituting each identified storm cell. In subfigure *(d)* of Figure 2.5, many pixels with reflectivity above 40 dBZ are excluded. This reduces the explanatory power of the tracked cells. After consideration of performance on several sample images, an intermediate value of $U = 1.7$ was selected in a trade-off between identifying cohesive, representative cells and including as many strong values as possible. Across all sample cases, this was the most inclusive value that did not result in cells with obvious bridges. In the case of a ridge of high reflectivity, however, it is possible that sufficient strong points will be flagged along the ridge for SPA to identify an elongated storm cell. Figure 2.6 shows a squall line forming a challenging case for SPA. When U is increased enough to split the line, an excessive number of pixels with high reflectivity are excluded from the resulting clusters.

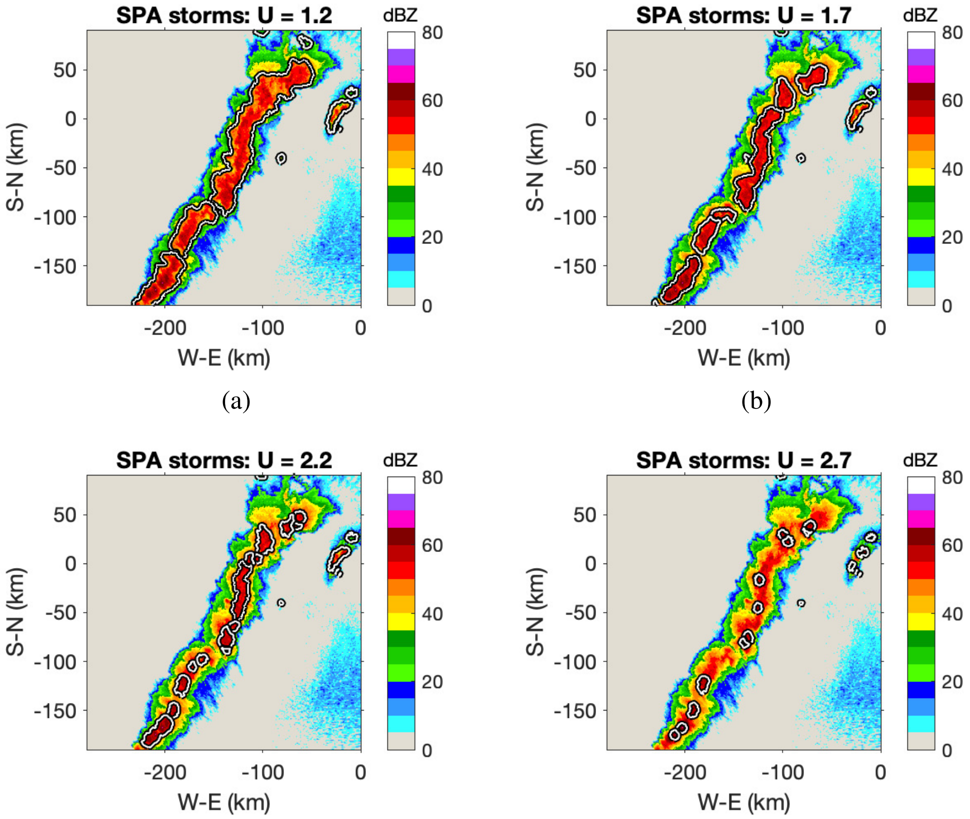
The reach parameter r is provided in units of pixels. Because the input images were generated at a resolution of 1.0 km per pixel, any pixels closer than 5.1 km are considered connected. Connectivity is recorded internally in the form of a graph,



(a) (b)

(c) (d)

Figure 2.5: Region of reflectivity observed by the KTLX radar on May 4, 1999, with detected SPA storm cells using different values of U outlined. (a): $U = 1.1$. (b): $U = 1.4$. (c): $U = 1.7$. (d): $U = 2.0$. As U increases, more distinct storm cells are identified, but each cell includes fewer pixels. For all subfigures, $L = 0.9$, $r = 5.1$, $p = 2.5$, and $n = 2$.



(a) (b) (c) (d)

Figure 2.6: Comparison of storm cells identified by SPA using different values of U. For all four tests the displayed reflectivity is a scan from the KICT radar in Wichita, KS at 04:42 UTC. (a): Most of the reflectivity above 40 dBZ is included in three elongated cells. (b): Each cell is replaced with two to four smaller cells. The total area included in cells has decreased. (c): The cells again shrink, but one particularly long cell is still visible from 0 to 50 km south of the radar. Points with reflectivity of 55 dBz are sometimes excluded. (d): Only small regions around the very strongest points in the image are included in cells. Cells approach the shape of circles with radius r around single strong point.

where each pixel is assigned a list of all other pixels to which it is connected.

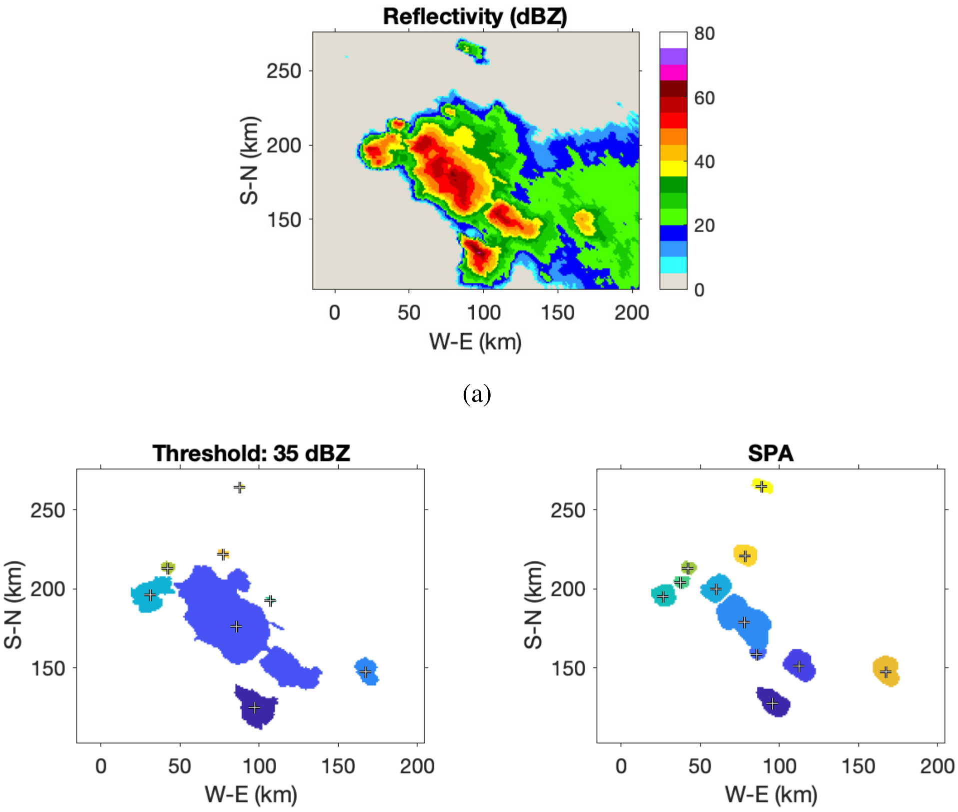
In the clustering step, connected-components labeling was implemented according to the one-pass method [19]. This method begins by assigning the first label to a single pixel. All pixels connected to that pixel are added to a stack. Pixels are then removed from the stack one at a time. As pixels are removed from the stack, they are assigned the same label. Any pixels which are connected to the newly removed pixel but have neither been added to the stack nor assigned a label are then added to the stack. In this manner, when the stack is empty, all pixels less than r km from a pixel with the initial label have been assigned that label. A single connected component has been identified. This process is repeated starting from the next unlabeled pixel, assigning successive labels.

To satisfy the requirements of SPA, an additional constraint was added to the connected-components method. When a pixel is removed from the stack, it is checked for strong or weak flags. Strong points are assigned the current label and their neighbors are added to the stack as in the original method. Weak points are only labeled. This distinction implements the dilation step of SPA in which weak points may fill out a cell if and only if they are connected to strong points.

After applying erosion and dilation $n + 1$ times, an output image is constructed of the same size as the Cartesian image. In this output, each pixel which was associated with a storm cell has the value of that cell's label. Pixels associated with no storm, whether excluded from the algorithm input or left unflagged within the algorithm, have a value of zero. The storm cells reported by this implementation of SPA are representations of the distinct areas of reflectivity in the initial radar image. For each cell, the centroid $(\bar{x}, \bar{y})$ is extracted from the k pairs of (x, y) coordinates of pixels corresponding to reflectivity Z:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i Z_i}{\sum_{i=1}^{n} Z_i}, \ \bar{y} = \frac{\sum_{i=1}^{n} y_i Z_i}{\sum_{i=1}^{n} Z_i}. \tag{2.4}$$

Although the (x, y) values are weighted by reflectivity, the application of SPA has already removed potential large regions of low reflectivity from consideration, and there is little deviation from the centroids obtained without this weighting. Sample storm cells detected by SPA are shown in Figure 2.7 with the corresponding centroids. Sample storm cells and centroids extracted by an implementation of the TITAN single-threshold detector are also shown. The resulting centroids represent features serving as inputs to the track association stage.



(a)

(b)

(c)

Figure 2.7: (a): Reflectivity observed northeast of the KDDC radar on July 8, 2004 at 00:10 UTC. (b): Output of simple reflectivity threshold storm detector considering all pixels above 35 dBZ to belong to storms. Each color signifies a separate storm. Shapes are irregular and do not reflect internal structure. (c): Output of SPA on the same data. The largest, most irregular storm is broken into several cells better indicating its internal structure.

Chapter 3

Multiple Hypothesis Tracking

In the centroid-based model of storm tracking, after storm cells have been observed by an identification process, a track association stage is necessary to determine which cells represent the same entity at two or more times. Multiple Hypothesis Tracking (MHT) is a technique that incorporates data from multiple scans in order to resolve uncertainties in track assignment [15]. This work explores the utilization of MHT in the track association stage of storm tracking. This chapter presents the advantages of MHT in the context of alternative track assignment techniques, details the algorithm as it is generally applied, and finally addresses the changes made in order to apply MHT to weather radar data for storm tracking.

3.1 Background of Track Association

Tracking algorithms designed specifically for weather targets include a variety of approaches to the problem of associating storm cells across scans.

The storm tracking algorithm in [3] generates compact cells limited to the highest reflectivity values in the data, where each cell containing relatively few separate radar measurements. As each cell contains a single local reflectivity maximum, many cells may be extracted from a single region of generally high reflectivity.

To improve robustness in tracking these small targets, vertical associations of cells across multiple elevations, referred to as volume cells, are tracked. New single-elevation cells are assigned to volume cells and tracks by comparison with position forecasts of previous volume cells. The first volume cell forecast sufficiently close to a new observation is associated with that observation.

Another approach relies on overlap between new cells and forecasts of previous cells [6]. For each cell identified in a new scan, an estimate of historical position is made based on an approximation of steering level winds. The amount of overlap between cells in the previous scan and these estimates is determined. Rain rate is calculated for the pixels within each cell. If the pixels containing most of the rain rate within a new cell are located within the previous cell on an existing track, the new cell is associated to that track. Multiple cells, each with limited overlap with one track, may still be considered a split, and multiple tracks overlapping a single cell may indicate a merge. After all associations are made, the difference between each historical position estimate and the last centroid in the track with which it was associated is determined. The average of these errors is used as feedback to update the approximation of wind for future scans.

In TITAN, cells are assigned to tracks such that the total cost of assignments is minimized [7]. A cost matrix C is constructed such that there is a row for each existing track and a column for each new cell. C_{ij}, the value in row i and column j, is the cost of assigning new cell $z_{\mathrm{j}}(t)$ to existing track T_{i}. This cost is a weighted sum of the differences in centroid position and storm volume between $z_{\mathrm{j}}(t)$ and the last observed cell on a track $z_{\mathrm{i}}(t-1)$. Assignments are selected from this matrix such that each row or column contains only one assignment and the sum of the costs selected is minimized. The resulting solution is optimal in terms of the cost function.

When a cell is further from a track than appears possible, the cost of that assignment is set to an extremely large value such that the minimized solution will prefer any other assignment. The condition signaling impossible motion is centroid displacement faster than 100 km hr^{-1}. If such an assignment is selected in the optimal solution, as when a storm is further away from all existing tracks than the motion constraint allows, a new track is initiated. The counterpart to this case, when a track is distant from all new observations, corresponds to track termination.

Centroid motion artifacts caused by the cell identification stage may interfere with this approach. When two storms merge, the centroid is displaced, potentially beyond the maximum allowed motion threshold. TITAN attempts to address this by searching terminated tracks for merge participants, or initiated tracks for split participants, when the motion constraint is violated. This category of error may also occur in the absence of a merge or split. Representing a cell primarily by its centroids limits the ability to distinguish between centroid motion from rapid shape change and centroid motion due to the motion of the entire cell.

The Enhanced TITAN (ETITAN) algorithm first attempts to select obvious assignments by comparing new observations to storm position forecasts obtained by the application of TREC [12]. New cells which have significant overlap with cell forecasts from tracks are assigned to those tracks. The remaining storms are assigned to tracks according to a modification of the original TITAN method intended to tolerate centroid motion artifacts. The maximum displacement constraint, above which storms may not be assigned to tracks, increases with the size of the new observation and the last observation on the track. Otherwise, ETITAN associates cells to tracks in the same manner as TITAN.

The SCIT algorithm also uses forecasts of storm positions to assign new observations to tracks [2]. If a new observation is within a threshold distance of the expected position of previous tracks, it is assigned to the track with the closest expected position. Forecasts are extracted from a least-squares fit on previous, equally-weighted positions. No specific procedure addresses merging or splitting of tracks.

Comparison of new cells to forecasts is an essential approach to the track association methods reviewed. This is a natural technique for ensuring tracks reflect the actual motion exhibited by storms. TITAN ignored this feature in favor of minimizing change in position between scans, but ETITAN added forecast comparison as an initial check before minimizing costs [12]. An improvement in this integration of forecasts, then, would improve the effectiveness of the track association stage. MHT provides a natural opportunity to make such an improvement.

Storm tracking can be viewed as a specific case of a more general object tracking problem referred to as multiple target tracking. In this context, MHT is a preferred method for track association [15]. The key advantage of MHT is its deferment of tracking decisions. MHT is structured to select the best track assignments possible at a given time while also allowing alternate assignments as new later data arrives through maintenance of multiple track assignment hypotheses.

The general structure of MHT was first proposed in order to track aircraft [20]. This approach generates a new hypothesis for each possible assignment of a newly observed cell to tracks present in the previous scan. Hypotheses are scored according to the agreement of the new observation with a Kalman filter estimating the state of the track to which it is assigned. Observed cells may also be interpreted as the initiation of new tracks or as false alarms assigned to no track. Either case forms a new hypothesis as if the cell had been assigned to a track. Despite the extensive memory requirements of the resulting expanding tree of hypotheses, only hypotheses with extremely low probability are removed. Hypotheses which are sufficiently similar may also be combined.

Later implementations of the technique sought to improve the computational requirements [15]. By generating only the k-best hypotheses as each new scan arrives, rather than enumerating every possibility, combinatorial explosion is reduced [21]. To generate a limited number of hypotheses, a cost function is evaluated based on the likelihood of each observation given the state estimate of each track. Minimizing this cost function gives an optimal solution to the assignment problem. The second through k-best solutions are then found by application of Murty's k-best algorithm [22]. More details of this approach to MHT are included in section 3.2.

When compared to other object tracking techniques, advantages of MHT include support for track initiation and termination as well as the ability to model occlusion or other causes of missed detections [11]. This feature is essential for storm tracking considering the irregular and dynamic nature of storm cells. While MHT does not support merges and splits, it may interpret a brief merge or split as a missed detection and resume the track if the storms separate or rejoin at a later scan. MHT also provides a mathematically optimal solution to the assignment problem, assuming the assumptions about measurement noise are correct, due to its integration of Kalman filters. Tracking methods based on Kalman filters have demonstrated performance advantages over least-squares regression for prediction of storm tracks [14]. These advantages support the suitability of MHT for the task of storm tracking.

3.2 Algorithm

The approach to MHT used in this research is organized in a similar manner to that in [20], generating new hypotheses whenever observations are assigned to

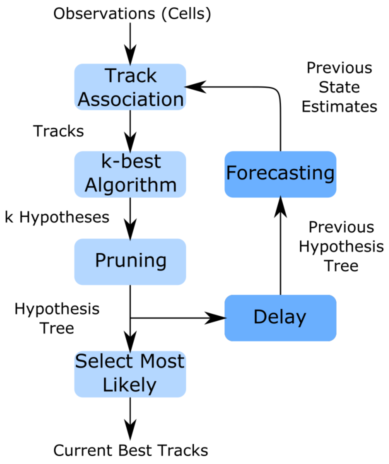


Figure 3.1: Flowchart illustrating the MHT implementation used. As new observations of storm cells arrive, they are compared to state estimates from previous hypotheses. This generates tracks and indicates where track association was ambiguous. The k-best hypotheses are then generated from these ambiguous associations. Hypotheses are added to the hypothesis tree; low quality hypotheses are pruned. The most likely hypothesis is selected as the tracking decision at this point. When more data arrives, the most likely hypothesis may change. All hypotheses in the tree, not just the most likely, are used to make track state estimates to which new data will be compared.

tracks [21]. Figure 3.1 illustrates the steps of this implementation of the MHT algorithm.

To create a hypothesis, MHT begins with a parent hypothesis, or set of tracks ending at the previous scan. For each track in the parent, the position and velocity of the tracked object are estimated using a Kalman filter. An ambiguity matrix indicates observations which fall within the validation region of each track defined by the uncertainty of the position estimate. The validation region is defined by the Mahalanobis distance

$$\gamma = \sqrt{(\vec{z} - \vec{\mu})^T \mathbf{S}^{-1} (\vec{z} - \vec{\mu})} \tag{3.1}$$

where $\vec{z}$ is the observation vector, $\vec{\mu}$ is the vector of mean values for the estimate, and $\mathbf{S}$ is the covariance matrix for the estimate [23]. This matrix is used to derive cost matrices reflecting the likelihood of each assignment in its validation region. k new hypotheses are then generated by Murty's algorithm.

The probability that each new hypothesis is correct is derived from the joint probability of two statements. First, the parent hypothesis must be correct. Unlikely parent hypotheses result in children that are less likely. The likelihood of the parent hypothesis was recorded when it was generated during the previous scan. Second, the new assignment of the cell to the track must be correct. This is based on the agreement of the new observation with a normal distribution estimating the state of the assigned track. Cells which are distant from the expected position of the tracked storm result in child hypotheses with low probabilities.

Hypotheses are stored in a track tree of configurable depth N, with different hypotheses originating from the same parent represented by branches from a single node. When a new scan arrives, leaving a node more than N scans in the past, the single branch from that node with the highest likelihood is selected and the other branches are pruned. The output of the algorithm is the hypothesis from the latest scan with the highest likelihood.

3.3 Implementation

The implementation of MHT used in this thesis is based on that presented in the previous section. As provided, it consists of a command line tool which operates on text files defining the configuration parameters and the observations at each scan. Descriptions of parameters and the values used for this work are enumerated in subsection 3.3.2.

For every point, an (x, y) coordinate pair must be provided. An observation ID defined here will appear in the output to identify which observations were ultimately assigned to which tracks, but will not affect the algorithm. A vector of 25 texture values may also be provided; this supports a cross-correlation filtering process. If the texture values are provided and this process is enabled, any track association for which the new observation and previous observation are too dissimilar in texture will be rejected.

Running the command line tool generates an output file enumerating all tracks and false alarms found. Tracks are described as a list of observations where each entry includes information about the state estimate of the track after the addition of that observation. For false alarms, only the position and ID of the observation are listed, as they were associated with no track or state estimate.

3.3.1 Assumptions for Storm Tracking

Tracking evolving, widely distributed storm cells requires modifications from the initial implementation, which was tested on camera image sequences [21]. MHT, as described, does not allow for the possibility of track merges and splits. This is problematic; merging and splitting behavior commonly occurs among storms. This conflicts with the one-to-one solutions which minimize assignment costs inherent to the k-best algorithm. A full expansion of MHT to allow one-to-many and many-to-one assignments is beyond the scope of this work. Instead, the use of SPA for the storm identification stage is intended to mitigate the influence of merges and splits on overall forecast quality. When complex storm shapes are represented as several smaller, more dense storm cells, MHT is able to assign observations to tracks in a one-to-one manner and, by aggregating tracks, track the apparent motion of the complex structure.

3.3.2 Configuration of MHT

Modification of the configuration parameters of MHT was required to optimize the performance of MHT on storm centroids. For this work, the initial parameters used were based on those from a previously unpublished survey, which evaluated performance on various cases of simulated storm tracks [24]. The parameters used and changes made are presented in this section.

When a new scan is completed, all observations within a certain Mahalanobis distance of a track's state estimate are considered for assignment to that track. The squared Mahalanobis threshold γ^2 limits the size of this validation region. In accordance with the survey, $\gamma^2 = 10.0$ is used.

Several MHT parameters control the performance of the Kalman filters estimating track states. The initial process, measurement, and state covariance estimates are all configurable. No values in this set were changed from those found in the survey. The initial state covariance matrix for a track, $\mathbf{P}(k)$, is given by

$$\mathbf{P}(k) = \begin{bmatrix} x & 0 & 0 & 0 \\ 0 & v_x & 0 & 0 \\ 0 & 0 & y & 0 \\ 0 & 0 & 0 & v_y \end{bmatrix}, \tag{3.2}$$

where the position variances x and y are 2.0 km and velocity variances v_x and v_y are 7.5 km scan^{-1}. The measurement noise $\mathbf{R}(k)$ is also described by a covariance matrix

$$\mathbf{R}(k) = \begin{bmatrix} R_x & 0 \\ 0 & R_y \end{bmatrix} \tag{3.3}$$

where $R_x = R_y = 2.0$ pixels^2. The process noise is described by a scaling factor q

rather than a full matrix; the value $q = 1.0$ was retained from the survey.

Hypothesis management is configured by a set of three parameters. When a new scan is processed, up to k child hypotheses are generated according to Murty's k-best algorithm. The tree of hypotheses may defer decisions between hypotheses until N additional scans have arrived. When another scan arrives and new child hypotheses would add a level to the track tree, a final decision is made at the root of the tree. The branch of child hypotheses with the highest total between its most recent children is selected as the new root node and other branches are pruned.

The third hypothesis management parameter, the likelihood ratio $G_{\min}$, restricts the generation of low-quality hypotheses. If the likelihood of a new hypothesis is evaluated to be less than $G_{\min}$ times that of the current most likely hypothesis, the new hypothesis is not added to the track tree. None of the hypothesis management parameters were addressed by the survey, so values of $k = 300$, $N = 3$, and $G_{\min} = 0.001$ are used as suggested by Hingorani and Cox [21].

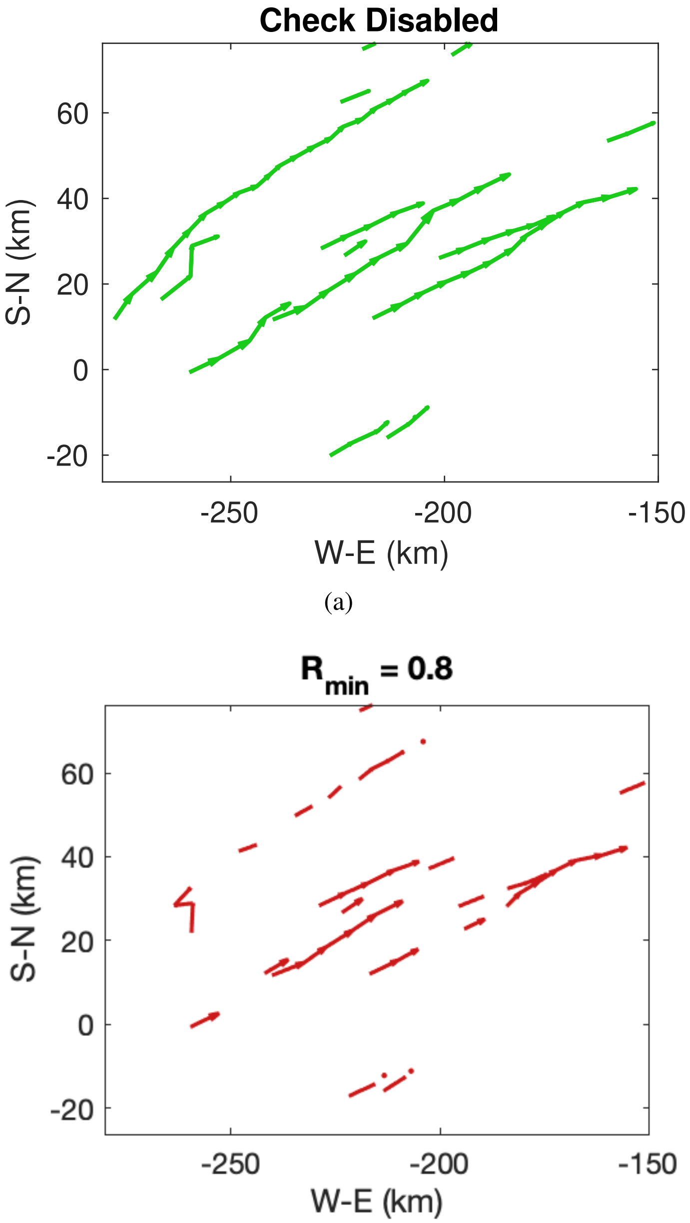
Another set of parameters affects the behavior of the tracker under special conditions. The probability of detection parameter P_D allows MHT to skip measurements in tracks. Higher values of P_D reduce the likelihood of hypotheses containing a missed observation. $P_D = 0.9$ is used to allow a moderate number of skips. Its counterpart, the mean false alarm ratio λ_{fa}, corresponds to observations which are assigned to no track. Hypotheses with such an assignment are less likely when P_{fa} is low. P_{fa} is set to 2.0×10^{-5}.

The mean new tracks parameter λ_N controls the other possible assignment for measurements which are unsuited to any track. Such observations may indicate the formation of a new track with initial state covariance $\mathbf{P}(k)$ at the location of the new observation. λ_N is set to 0.01. Tracks terminate and are removed from consideration for child hypotheses when no observation in an incoming scan is considered a likely assignment. This occurs when the likelihood of any assignment, or of a missed observation, is lower than the likelihood set by λ_x, the track termination ratio. The value selected for λ_x is 0.2. This parameter is provided to the implementation as a percentage whereas the other parameters are all provided as fractions.

The cross-correlation threshold $R_{\min}$ controls an additional check on the quality of track assignments. If enabled, assignments with dissimilar textures will be rejected. Similarity is too low when the best cross-correlation between sections of the new texture vector and the vector of the last observation in the track falls below this threshold. This check has been found unsuitable for use when observations are occluded [21]. The irregularity of storms suggested that this check should not be applied for storm tracking. Restricting association only to those storms which maintained similar texture vectors across scans would also worsen the effects of apparent centroid motion from growth, decay, or false mergers. To confirm the ineffectiveness of this check, MHT was applied to the same set of centroids with and without restricting associations by cross-correlation and using a range of values for $R_{\min}$. Figure 3.2 illustrates reduced track quality when associations are restricted. The threshold shortens and separates tracks by removing associations which otherwise would form consistent, isolated tracks.

In accordance with these results, the cross-correlation check is disabled for storm tracking by supplying a uniform field of zeros as the texture data for each corner. This ensures that the calculated cross-correlation is always perfect, preventing the removal of associations. To indicate that this step is disabled, the algorithm is initialized with $R_{\min} = -1$.

Table 3.1 collects the final values selected for the configuration parameters.



(a)

(b)

Figure 3.2: Tracks generated by MHT with (b) and without (a) the cross-correlation check. The radar data, not shown, is from the KINX radar in Tulsa, OK, taken at 01:02 UTC on 2004/05/30. The long, consistent tracks found without the restriction are split and shortened as parts of the tracks are removed by the threshold. Similar behavior is visible with lower thresholds as well. This led to the disabling of cross-correlation checks for this application.

Parameter	Description	Value
γ^2	Squared Mahalanobis threshold	10.0
x,y	Initial x and y position variances for $\mathbf{P}(k)$	2.0 px
v_x, v_y	Initial combined velocity variance for $\mathbf{P}(k)$	7.5 px scan^{-1}
R_x, R_y	Measurement noise covariance for $\mathbf{R}(k)$	2.0 px^2
q	Process noise covariance matrix scaling factor	1.0
k	Maximum hypotheses generated from one new scan	300
N	Maximum track tree depth	3 scans
$G_{\min}$	Minimum likelihood ratio for accepted hypotheses	0.001
P_D	Probability of detection	0.9
λ_{fa}	Mean false alarm ratio	2.0×10^{-5}
λ_N	Mean new tracks per scan	0.01
λ_x	Track termination chance	20%
$R_{\min}$	Unused minimum accepted cross-correlation	-1

Table 3.1: Configuration parameters selected for MHT.

Chapter 4

Evaluation

Chapter 2 presented the motivation for SPA as a storm cell identification algorithm and Chapter 3 presented MHT as a solution to the track association problem. Using the implementations described in both chapters, this chapter describes a combined SPA-MHT storm tracking procedure intended to leverage the strengths of both methods. The effectiveness of the combined method is evaluated on radar observations of a variety of storm conditions. Metrics of track quality are used to investigate the advantages and disadvantages of SPA-MHT.

The combined tracker operates on a set of scans. For each scan, radar data is transformed from native spherical coordinates to a Cartesian grid by inverse distance weighting interpolation. Each scan is represented by a single image of the reflectivity at 1 km above the elevation of the radar. SPA is applied to these images to extract a set of storm cells at each scan. The centroids of each cell are extracted and written to text files in a format appropriate for MHT. At this time, a parameter text file is also written to allow configuration of MHT from within MATLAB. For all tests in this chapter, the parameters in Table 3.1 were used.

After all input files are written, MATLAB invokes the MHT command-line tool with the appropriate arguments to use the inputs. The tool writes its outputs to another text file, which is then read by MATLAB. These outputs include a list of tracks and a list of centroids which were considered false alarms. Both lists are loaded and converted to formats suitable for plotting or for evaluation of algorithm performance.

4.1 Evaluation on Radar Data

Radar data collected by the Next-Generation Radar (NEXRAD) network was obtained from the National Centers for Environmental Information (NCEI) [25]. The enormous volume scans of these systems observed a variety of storm conditions to test SPA-MHT. For each case, the downloaded reflectivity data from within 300 km of the observing radar was interpolated to a Cartesian 1 km × 1 km grid at a constant altitude of 1 km above the radar. The resulting image is provided to SPA-MHT as input.

When observing severe weather, NEXRAD systems also track storm cells using the SCIT algorithm [8]. Each storm cell identified by SCIT is recorded with a label, position, and estimated motion vector. Collecting the labels present in sequential volume scans allowed reconstruction of SCIT track associations. These SCIT tracks provide natural points of comparison for the tracks generated by SPA-MHT. This data was also retrieved from the NCEI [25]. The cases on which SPA-MHT was evaluated are included in Table 4.1, arranged according to their origin.

Cases 1-3 were used in the original evaluation of SCIT [2]. These cases were used in the development of SPA-MHT, but the NEXRAD observations of these cases obtained from NCEI do not include the storm tracking information required for a full comparison. Cases 4 and 5 were selected from the development of SPA [13], and cases 6 and 7 were selected from data initially used to test the track evaluation metrics presented in Section 4.2 [26]. The following cases, 8-14, repre-

	Radar	Date	Start	End	SCIT	Overview Figure	Detail Figures
1	KIWA	08/06/1993	03:04	03:59	N		
2	KTLX	02/21/1993	17:06	18:33	N		
3	KLWX	10/06/1995	01:00	02:31	N		
4	KINX	05/30/2004	01:02	02:57	Y	A.1	4.1, 4.4, 4.6
5	KDDC	07/08/2004	00:02	01:56	Y	A.2	4.10, 4.11
6	KIWA	08/06/2003	18:00	19:59	N		
7	KLWX	10/06/2005	18:04	19:57	Y	A.3	
8	KTLX	05/04/1999	00:02	01:36	Y	A.4	4.2, 4.3, 4.7, 4.8, 4.9
9	KTLX	05/19/2013	20:02	23:57	Y	A.5	4.5
10	KTLX	05/03/2018	00:06	01:54	Y	A.6	
11	KTLX	06/07/2018	16:02	17:54	Y	A.7	
12	KTLX	08/03/2018	05:07	06:58	N		
13	KTLX	08/30/2018	13:00	14:56	N		
14	KTLX	09/21/2019	01:05	02:54	Y	A.8	4.12
15	KFDR	05/16/2003	03:00	04:59	N		
16	KICT	05/24/2008	04:01	05:56	Y	A.9	
17	KTWX	03/23/2009	22:03	23:59	Y	A.10	4.13
18	KVNX	04/18/2009	20:09	22:57	Y	A.11	4.14

Table 4.1: Sources of radar data used for evaluation, numbered for reference. The Radar column provides the four-letter identifier used for each WSR-88D system. The Date, Start and End columns specify the UTC timestamps of the first and last volume scan used. If SCIT tracks were available for the case, the SCIT column contains a Y; otherwise it contains an N. The Overview Figure for each case is located in Appendix A and illustrates the general layout of the test case. Any Detail Figures are located in Chapter 4 and focus on a smaller region as it evolves over multiple scans.

sent a variety of severe weather conditions observed by the KTLX radar. To guarantee a set of challenging cases, cases 15-18 were drawn from a survey of mergers between squall lines and isolated storms [27]. Cases 6, 12, 13, and 15, despite occurring after the adoption of SCIT for the NEXRAD network, did not have usable tracking data. These cases were only used during the development of SPA-MHT and were not considered in the evaluation.

The following specific cases illustrate the performance of SPA-MHT on data collected under various weather conditions. For each case, an image of the entire radar volume at the beginning of the sample data is provided in Appendix A without any tracks. Figures in this chapter, such as 4.1, show only detail areas from these cases. The nine subfigures of a detail area figure are divided into three columns and three rows. Each column of a detail area figure displays the state of the tracking algorithm at a different scan; the scans are arranged chronologically from left to right. The first row displays the tracks produced by SCIT overlaid on images of the reflectivity observed by the radar at three sequential scans. Each arrow within a track indicates a centroid identified at a different scan. Crosses on each track indicate the centroid associated with that track during that scan. Looking left to right along the row shows each SCIT centroid advancing along each SCIT track. The second row is organized in the same manner as the first, but generally displays the tracks produced by SPA-MHT instead of those resulting from SCIT.

Tracks which initiate after or are terminated before a scan are not included in the column showing that scan. If a track appears in a scan, a centroid was associated with that track either before and after or during that scan, even if the centroid is outside of the detail area. While scans are always arranged chronologically from left to right, they are not always immediately adjacent; some scans have been skipped to better illustrate storm motion over time.

In the first two rows of a detail figure, the reflectivity displayed has been faded to improve the visibility of tracks. This is not the case for the third row, which shows the same reflectivity images with colors accurate to the color scale at the bottom of each figure. The time at which each scan was initiated is also shown in the corner of each image in the third row, while the date is located at the top of the figure. The axes along the third row and first column indicate the position of the detail areas relative to the radar.

4.1.1 Isolated Storms

Strong, regularly shaped storms which remain consistent across multiple scans are easily tracked. SPA-MHT competently tracks storm cells in such cases, yielding tracks which agree with SCIT and with human interpretation. The tracks in Figures 4.1 and 4.2 agree with the visible motion of the underlying storms. The details of these tracks, however, demonstrate weaknesses of SCIT which are addressed by SPA-MHT.

Centroids identified by SPA were usually consistent between scans for isolated storms, remaining in a similar position relative to the underlying storm. The irregularity of SPA-MHT tracks was reduced compared to SCIT. Figure 4.1 contains an extreme example of an irregular SCIT track. The track found by SPA is smooth and consistent by comparison to the SCIT track, which deviates to the left and right. In the center of the detail area, the SCIT track reverses direction when it is assigned a centroid west of its previous position. These deviations do not represent the actual west-to-east motion of the storm. Instead, they are artifacts of the identification stage of SCIT, which moved the centroid in a manner disproportionate to the change in reflectivity. The SPA-MHT track, based on SPA centroids, better represents this

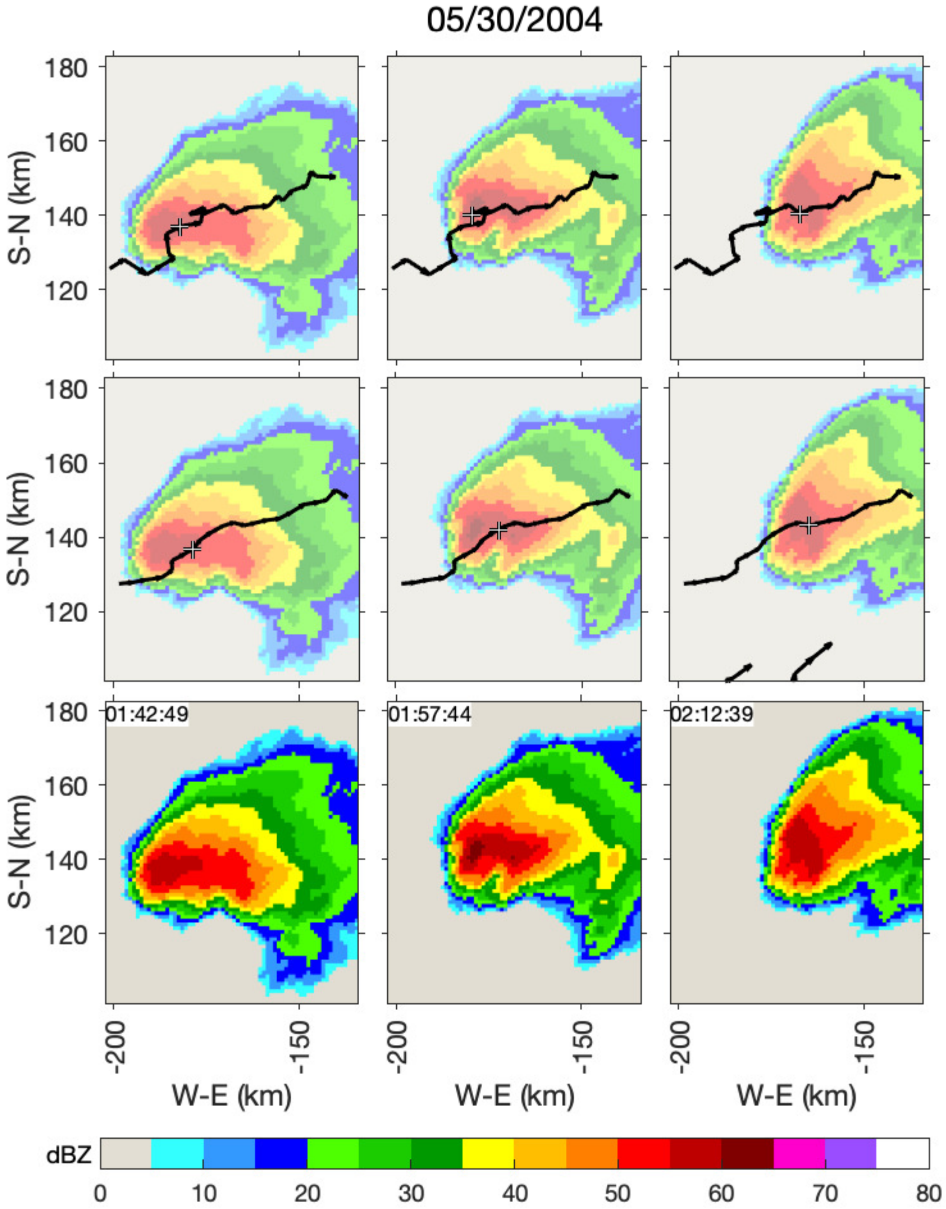


Figure 4.1: Detail area of three scans from the KINX radar in Tulsa, OK beginning from 2014/05/30 01:42 UTC. Tracks from SCIT and SPA-MHT are shown. The centroids identified by SCIT deviate widely from the linear motion visible to the eye, even appearing to reverse direction, due to shape changes in the corresponding cells. SPA-MHT yields more stable centroids for a more consistent track.

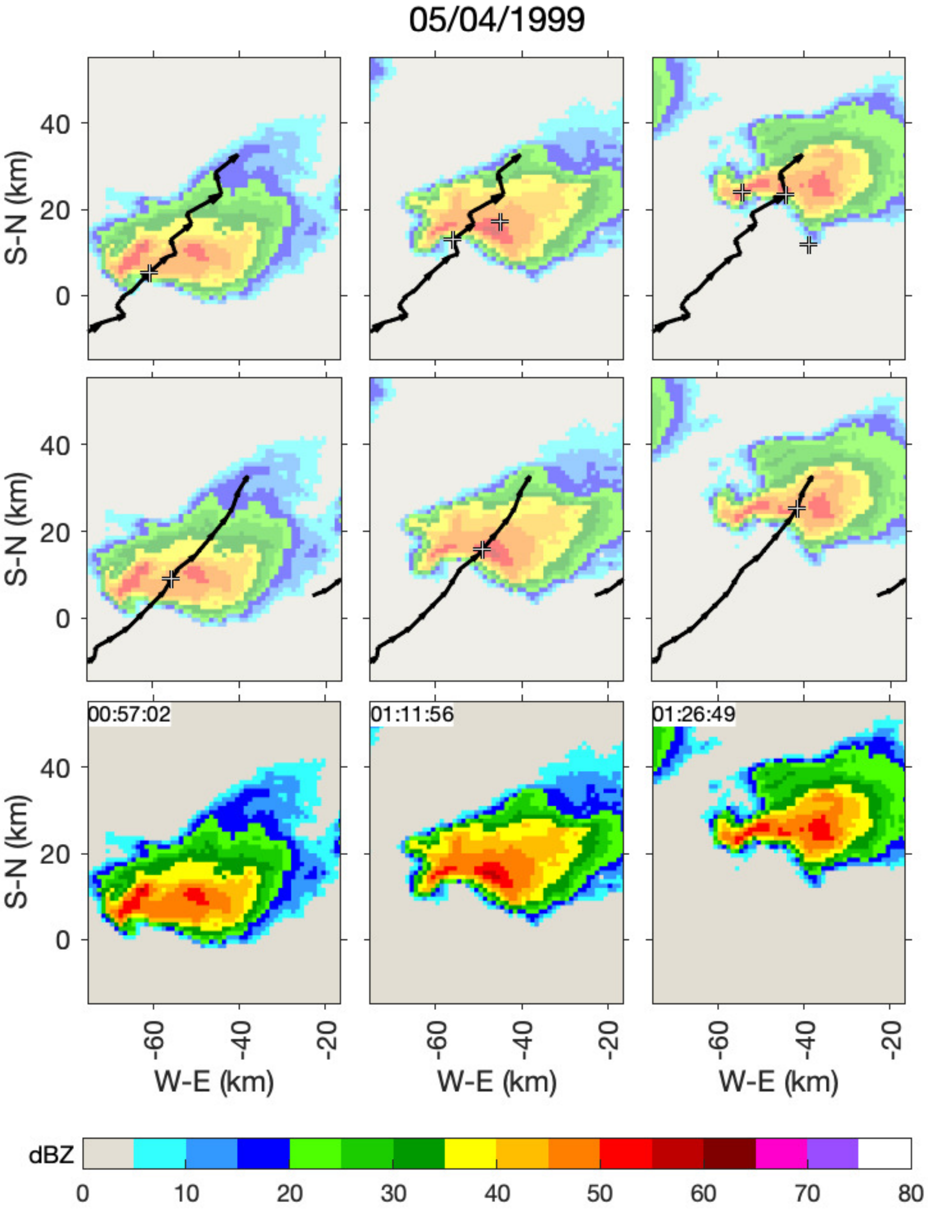


Figure 4.2: Detail area of three scans from the KTLX radar in Oklahoma City, OK beginning from 1999/05/04 00:57 UTC. Tracks from SCIT and SPA-MHT are shown. SCIT identifies several distinct cells in the second and third scans shown, resulting in an irregular track. SPA-MHT identifies a single cell for the storm, capturing the trend of motion to the northeast.

motion.

Similar behavior was identified in other test cases with isolated storms. In the case shown in Figure 4.2, a storm moves to the northeast. At first, SCIT and SPA-MHT each initialize a single track, identifying a single cell to represent this storm. As the storm moves, SCIT splits the storm, identifying one additional cell in the second scan and two additional cells in the third. This decision is defensible as the storm shape grows more irregular, but SCIT is unable to assign the additional cells to tracks. The additional cells are inconsistent, not identified in multiple consecutive scans, so no additional tracks are initiated. While the first track remains reasonably accurate, its predictive power is only applicable to the one of the three cells SCIT uses to represent the storm. The track also jumps each time a split occurs, leading to greater uncertainty in the track.

Likewise, in Figure 4.3, the long track produced by SPA-MHT is represented by multiple disjoint tracks in SCIT. The storm is initially tracked by both algorithms at the start of the test case 00:02 UTC, but the SCIT track jumps between a cell in the main body of the storm and another cell containing the small area at the tail of the storm. At 00:22 the tail cell is assigned to the original SCIT track and a new SCIT track is initiated which agrees with SPA. This confusion adds little to the overall model of the storm's motion. The new SCIT track at 00:22 attempts to describe the same feature which SPA-MHT has been tracking for almost twenty minutes. SPA-MHT held an advantage because the cell identified by SPA persisted between scans.

In general, SPA-MHT was much more likely to identify and track weak or short-lived storms. This sensitivity increased the number of unimportant or low-impact features tracked at any given time. Across several cases, many areas with low reflectivity were identified as storm cells by SPA. Tracking these cells potentially

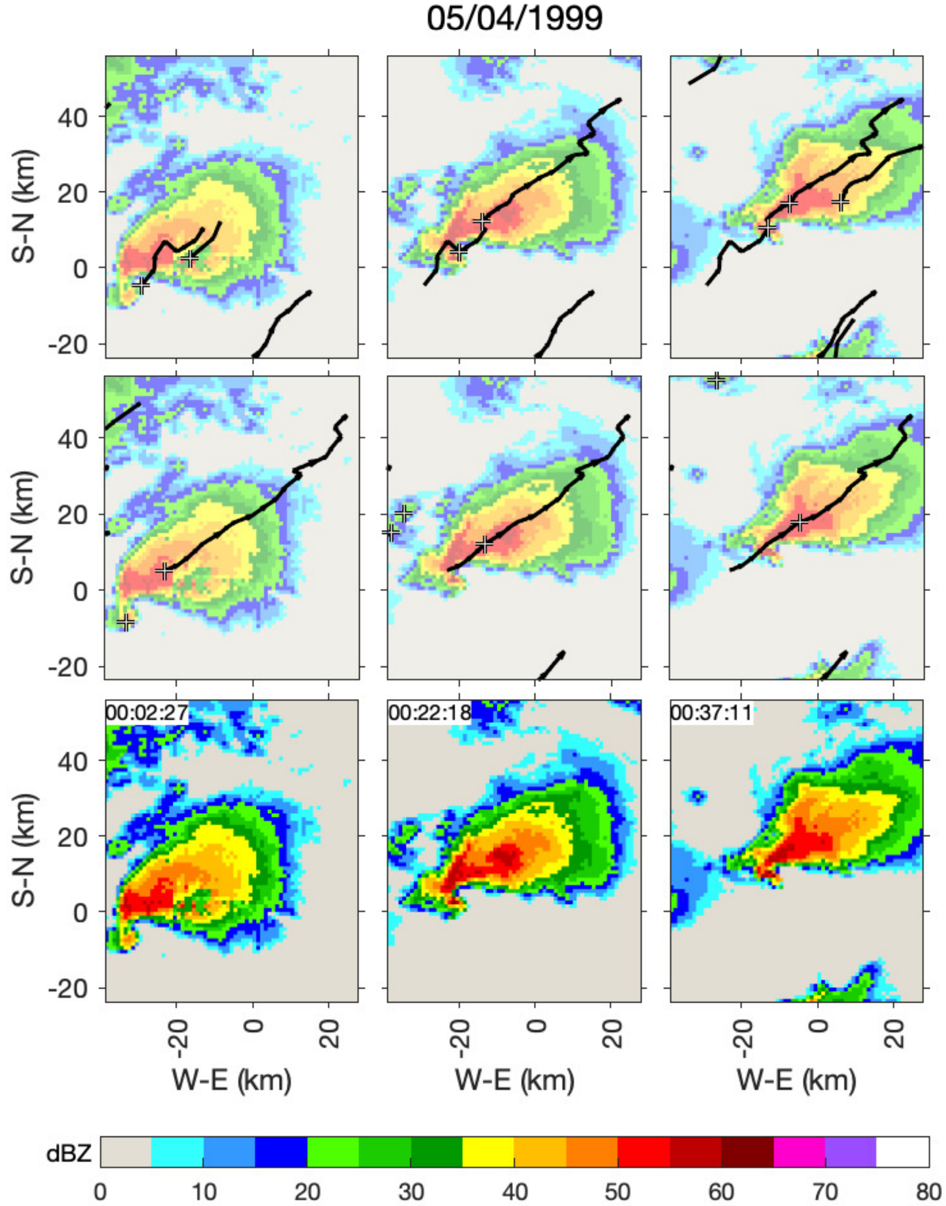


Figure 4.3: Detail area of three scans from the KTLX radar in Oklahoma City, OK beginning from 1999/05/04 00:02 UTC. Tracks from SCIT and SPA-MHT are shown. SCIT forms several disjoint tracks. The longest of these is initiated at 00:22 UTC, shown in the second column. By this time SPA-MHT has been tracking a similar feature since at least 00:02, the beginning of the test case. This SPA-MHT track is visible in all three scans.

diverts attention and processing power away from more significant features. When stronger precipitation is present, SPA does raise the threshold at which strong points are identified, automatically reducing the detection of extremely weak cells. Increasing the false alarm rate parameter for MHT, which was set very low for these cases, would result in higher likelihood scores for hypotheses which treated these cells as false alarms. This would correspond to a reduction to the number of these cells assigned to short tracks.

However, identification of small and weak storm cells is advantageous for early detection of storms. Figure 4.4 illustrates the early stages of a storm intermittently identified by SCIT. In the second scan at 02:02, SCIT begins to track this storm, which grows more intense by the third scan. SPA-MHT identifies and tracks the same cell consistently from 01:20, achieving 42 minutes of lead time. This allows earlier prediction and warning of developing severe weather. The trade-off between early detection of storms and the false alarm rate of tracking insignificant or non-weather features is not unique to SPA-MHT. Additional processing, such as the filtration of any tracks which contain no high-intensity storm cells, may mitigate the effect of false alarms.

4.1.2 Multiple Closely Spaced Storms

Tracking becomes more challenging for both SCIT and SPA-MHT when the storms observed are more tightly clustered. Both methods are capable of generating effective tracks under clustered conditions, as in Figure 4.5. In this case, the development of several storms in parallel results in some termination and initiation of tracks, but both methods generate several long, stable tracks identifying consistent features. The storms 0 km and 50 km north of the radar in the first scan are recognizable vi-

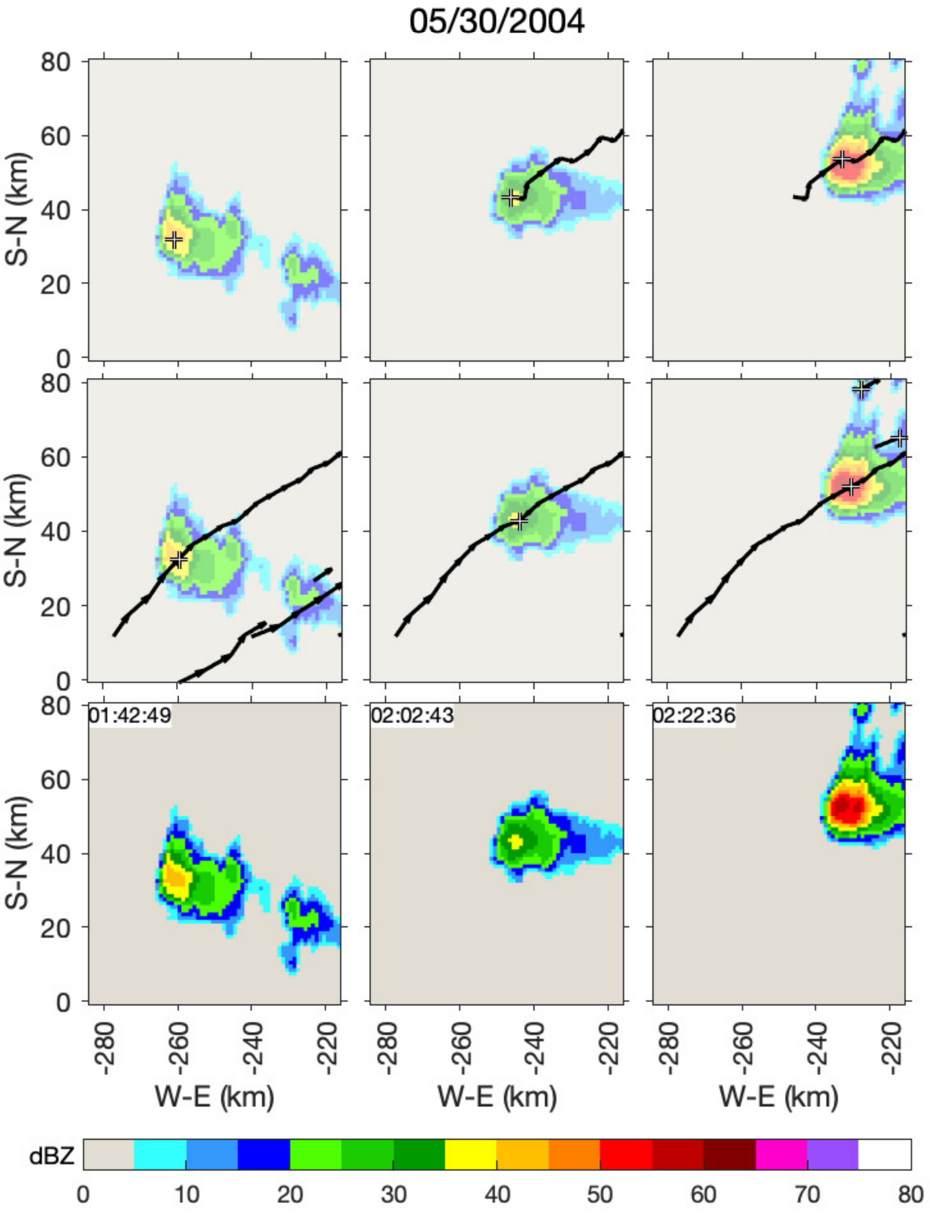


Figure 4.4: Detail area of three scans from the KINX radar in Tulsa, OK beginning from 2004/05/30 01:42 UTC. Tracks from SCIT and SPA-MHT are shown. SPA-MHT identifies a consistent track for the main storm as it initiates, while SCIT intermittently identifies the cell but does not associate it with cells in later scans until 02:02.

sually by the third included scan nearly 90 minutes later, despite the initiation and termination of short-lived nearby tracks in the interval. The southern storm in scan 2 results in another long-lived track from both methods before it splits at around 23:10.

Closer storms are more likely to engage in merges and splits. Both SPA-MHT and SCIT frequently interpreted merges as a track termination combined with a sharp jump in another track to the location of the merged centroid. This behavior is visible in SPA-MHT in Figure 4.6. The two storms in the center of scan 1 merge as they travel northeast, causing the southernmost SPA-MHT track to jump north to the new centroid in the second scan. SCIT, surprisingly, avoids this error, maintaining the two tracks in the center of the detail area throughout the observed interval. The three dimensional, spherical-coordinate identification stage of SCIT consistently identified two storm cells. SCIT effectively observed no merge. It is possible that the reduction of radar data from three dimensions to a two dimensional image concealed structure that better separated the two storms. Across several test cases, SPA appeared less likely than SCIT to break up storms into multiple cells. This tendency appeared advantageous in Figure 4.3, agreeing with visual inspection. Under close storm conditions where merges and splits occurred more frequently, however, SPA cells did not preempt these occurrences. This result was surprising given the adaptive design of SPA.

SCIT is not immune to merge or split errors. Figure 4.7 contains an example of two closely spaced storms moving in parallel. Both SCIT and SPA-MHT identify the two storms as separate cells at 00:27, and SPA-MHT successfully tracks both cells. In the next column, at 00:32, SCIT does not identify the northern storm and splits the southern storm into two cells. Both SCIT cells are incorrectly assigned to the two existing tracks. At 00:37, the identified cells are comparable to those

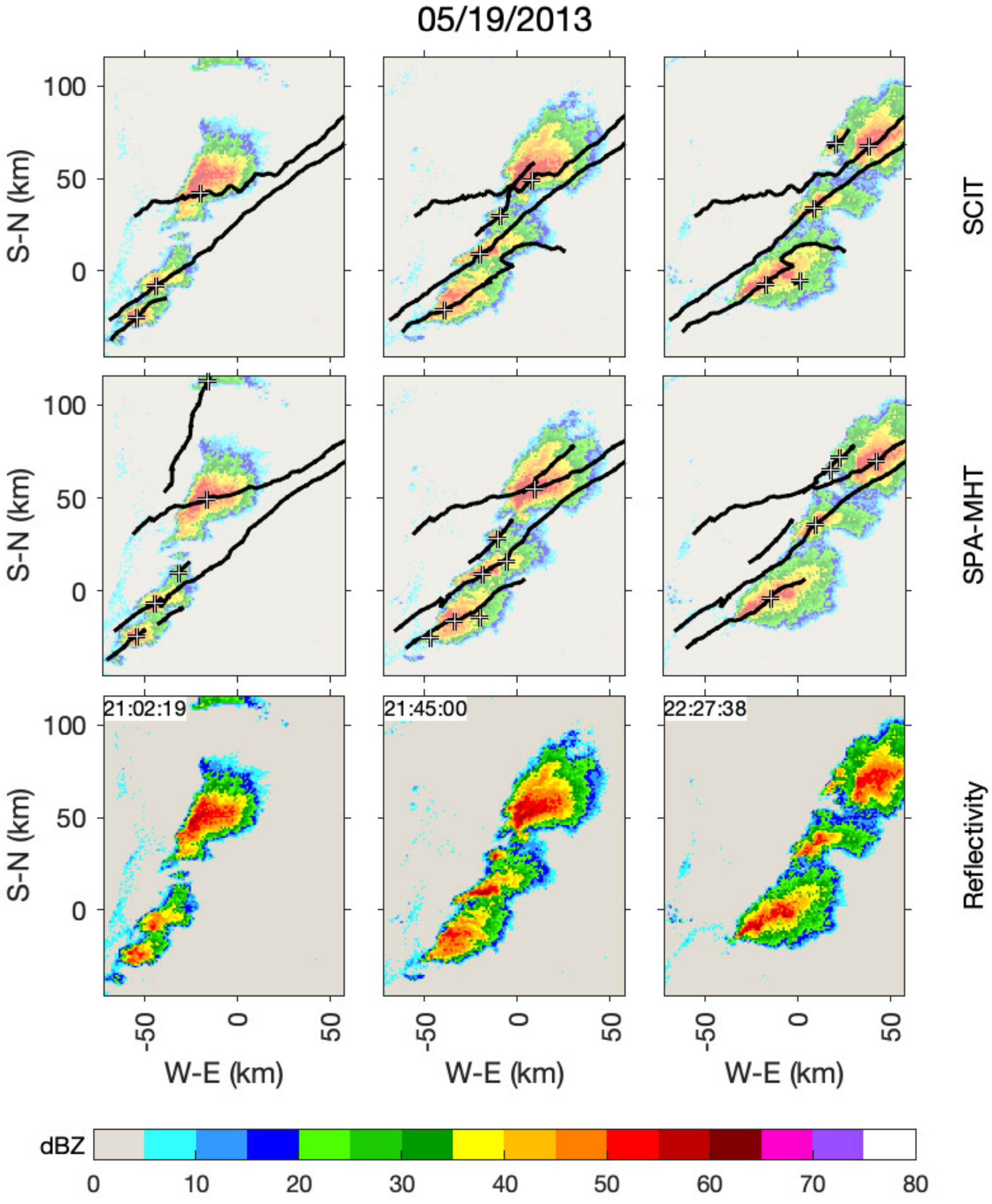


Figure 4.5: Detail area of three scans from the KTLX radar in Oklahoma City, OK beginning from 2013/05/19 21:02 UTC. Tracks from SCIT and SPA-MHT are shown. Multiple similar storms are consistently tracked by both methods over a period of hours.

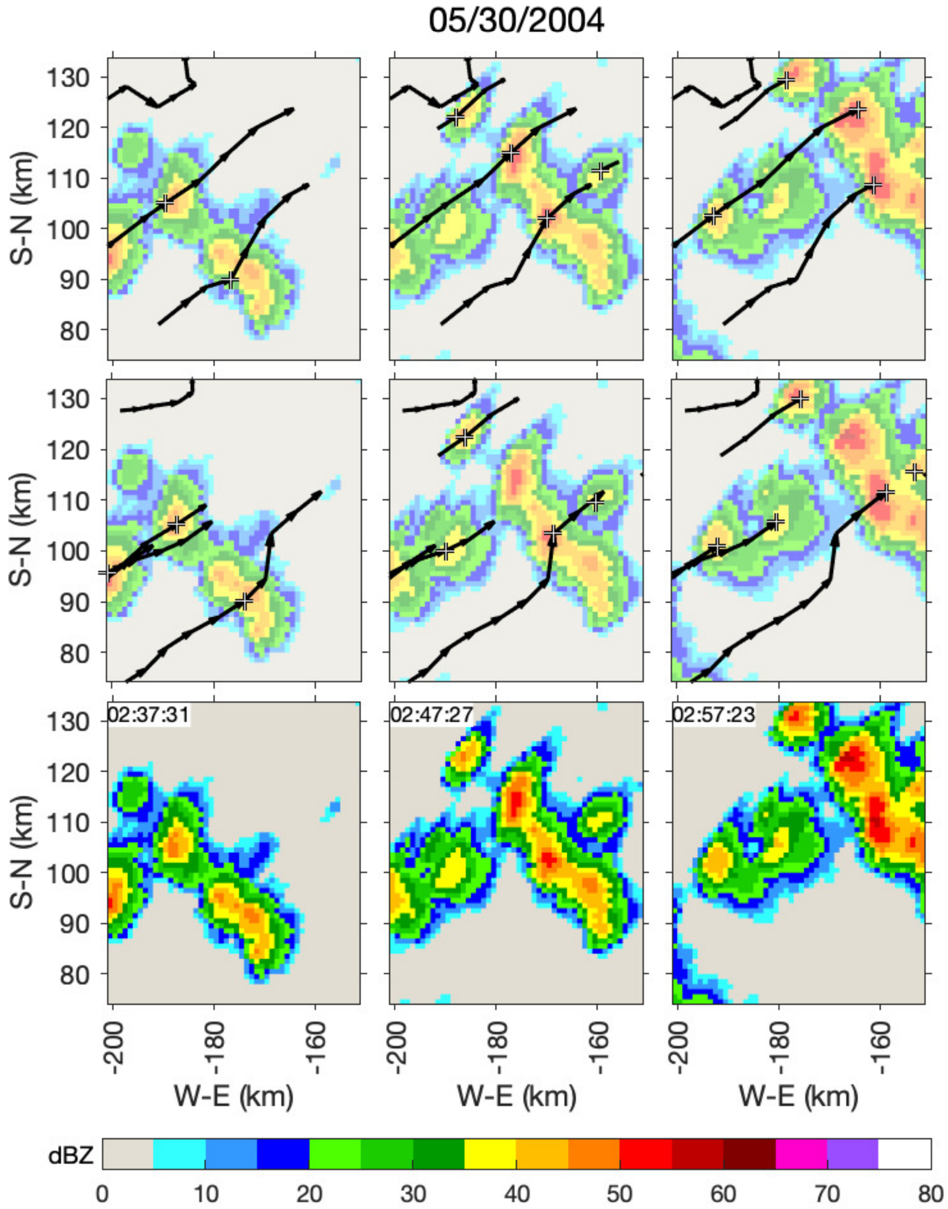


Figure 4.6: Detail area of three scans from the KINX radar in Tulsa, OK beginning from 2004/05/30 02:37 UTC. Tracks from SCIT and SPA-MHT are shown. When the two storms in the center of the area merge, SPA-MHT continues the track from the southern cell, causing a jump. The track of the smaller cell is terminated. SCIT interprets the two storms as separate cells throughout the scans, avoiding the merge problem.

from the first scan, but the track jumps from the previous error lead to incorrect assignments. This pattern, where the need to track a cell which was only briefly identified caused later tracking errors in SCIT, was repeated in several other scans and test cases.

The use of MHT for track association, however, was able to improve tracking performance in such situations. The track in Figure 4.8 includes a single scan where the storm tracked by SPA-MHT splits into two cells. Neither centroid is close enough to the single track state estimate to be considered likely. Rather than terminate the track, losing five scans of previous track positions, MHT interprets the lack of a centroid as a missed detection. When the next scan was processed, this hypothesis was supported by the identification of a single cell near the position predicted by the track. Track history preceding the brief split was preserved by MHT.

To directly test this capability, MHT was applied to the centroids detected by SCIT. The result, shown in the second row of Figure 4.9, demonstrates that MHT avoids the error made by SCIT association and recognizes the missed detection instead. In the second scan, the second cell identified in the southern storm is correctly assigned to no existing track, and when the original cells are again identified, both tracks are correctly continued.

Interaction between storms adds an additional challenge to tracking. The storms visible in Figure 4.10 are extremely close, merging shortly after scan 1. SCIT assigns the resulting cell to the long southbound track. This track persists into the second scan. A short southern track briefly coexists with the long track, but continues to the southeast when the long track terminates. The short track has effectively replaced the long-lived track, describing the same area of reflectivity. While it is possible that this interpretation accurately represents motion above the

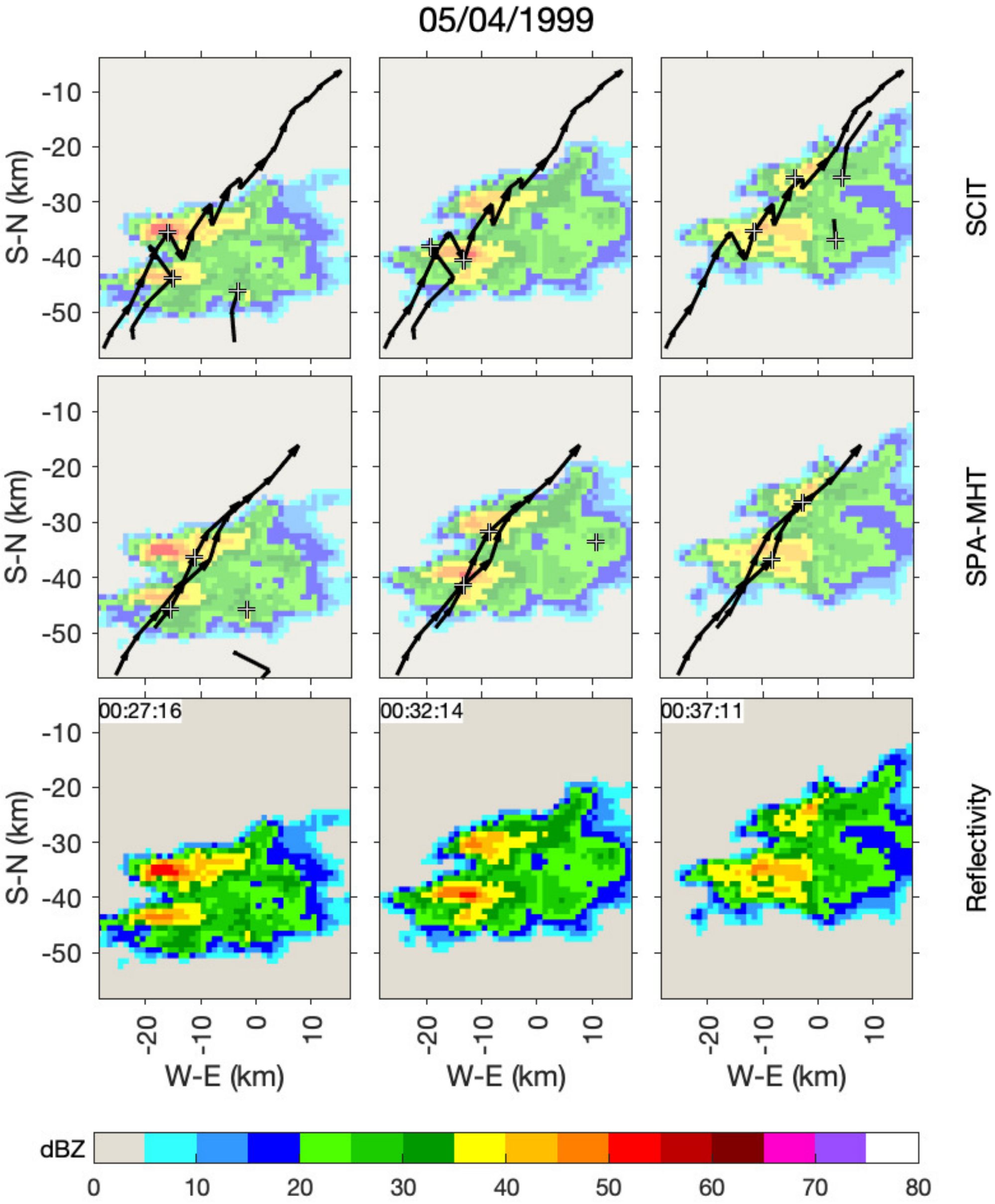


Figure 4.7: Detail area of three consecutive scans from the KTLX radar in Oklahoma City, OK beginning from 1999/05/04 00:27 UTC. Tracks from SCIT and SPA-MHT are shown. SCIT incorrectly assigns storms to tracks following a change in identified cells during the second scan. This results in a significant deviation of the tracks. SPA-MHT avoids this error by identifying the same pair of a northern and a southern cell across each scan.

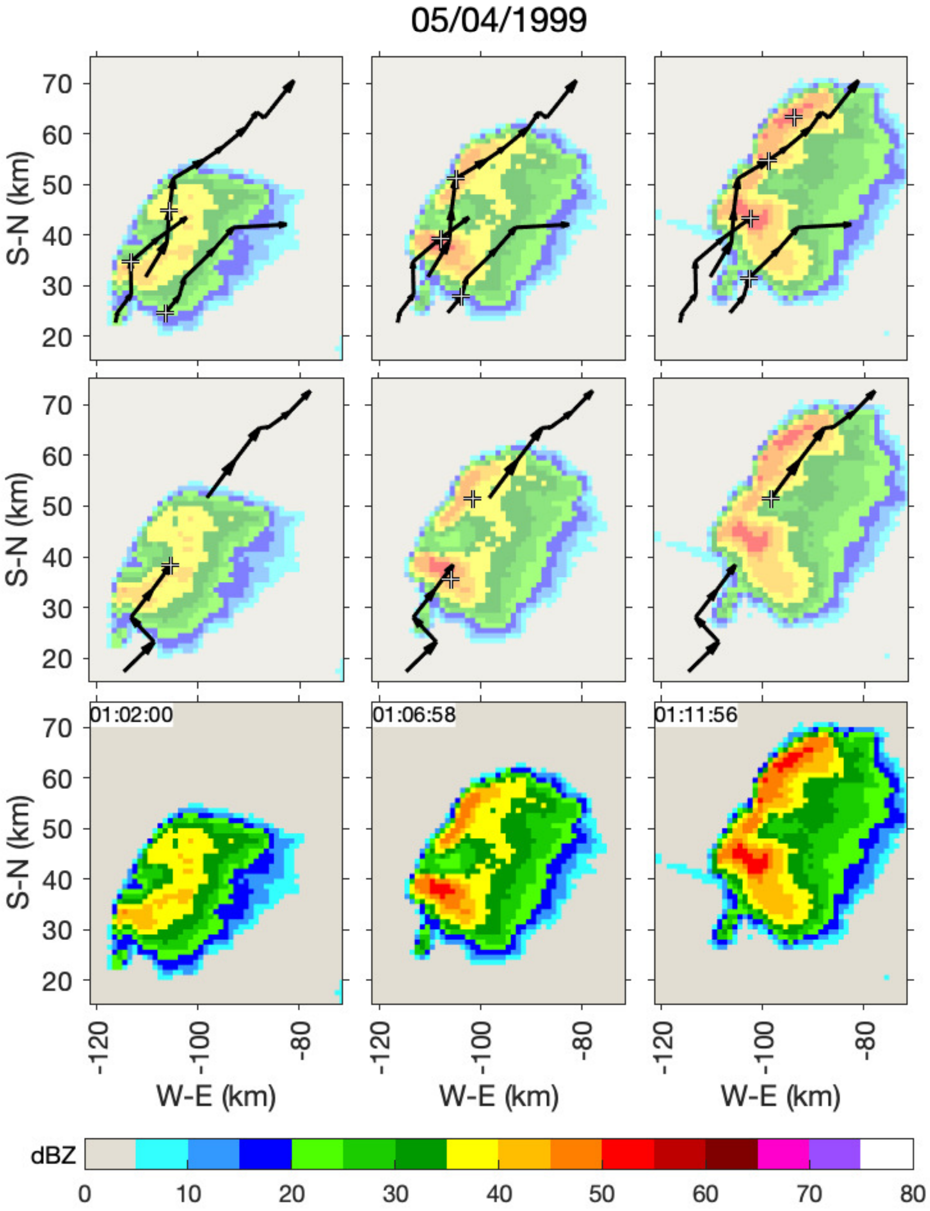


Figure 4.8: Detail area of three consecutive scans from the KTLX radar in Oklahoma City, OK beginning from 1999/05/04 01:02 UTC. Tracks from SCIT and SPA-MHT are shown. SCIT splits a storm into three to four cells, while SPA-MHT attempts to track one cell. Support for missed detections from MHT allows the track to survive a brief split visible in the second column. The southwestern and northeastern tracks are recognized as one track by SPA-MHT.

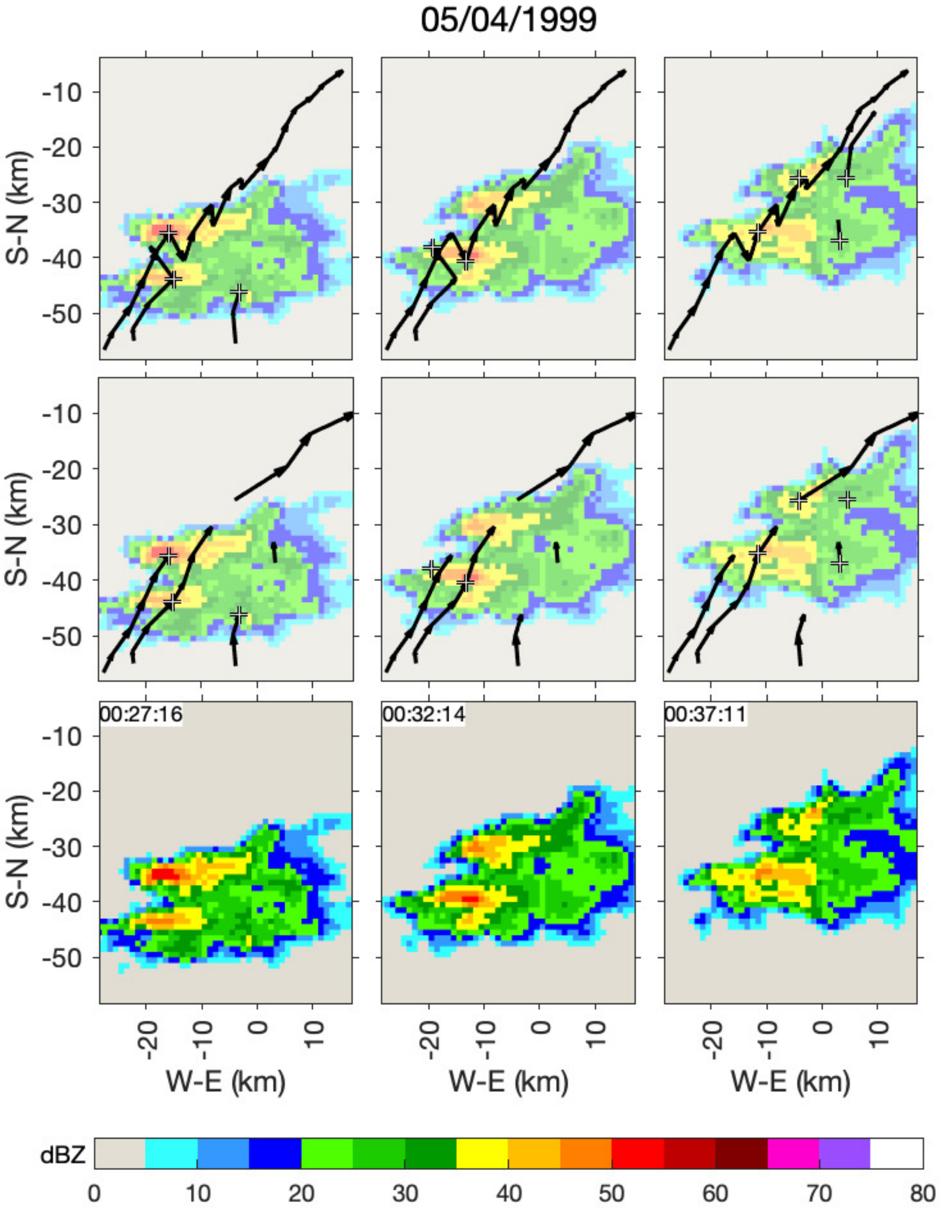


Figure 4.9: Detail area of three consecutive scans from the KTLX radar in Oklahoma City, OK beginning from 1999/05/04 00:27 UTC. Tracks from SCIT and from application of MHT to SCIT centroids are shown. MHT correctly identifies that the western storm in the second scan does not belong to either existing track. The northern track is correctly resumed in the third scan.

interpolated image, the single track produced by SPA-MHT is a better match to the visual. In the tracks identified by SPA-MHT, it is the western storm in scan 1 that dominates the collision and continues. This appears more accurate, as the final motion vector identified by both methods is closer to the initial motion of the western storm.

Figure 4.11 contains similar storms from the area slightly to the southeast of Figure 4.10. Again the consistent SPA-MHT track explains the same motion as two disjoint SCIT tracks. In the absence of storm structure not visible in these images, the continuous track is preferable to disjoint tracks, as it maintains a more confident estimate of storm state.

Extreme clustering results in the breakdown of both SCIT and SPA-MHT. The multitude of growing, decaying, and colliding storms present in Figure 4.12 resists tracking by both algorithms. SPA-MHT attempts many tracks which remain short due to rapidly changing features. SCIT successfully tracks several cells, potentially outperforming SPA-MHT, but fails to track the most intense region in the south of the system. Neither method represents the overall west-to-east motion of the system, which occurs in part due to growing storms on the leading edge. Instead, most tracks in both methods run from south to north. The number of cells identified changes radically from scan to scan due to merges, splits, and inconsistent identification. SPA, which identified consistent storms when there were few reflectivity maxima, is less effective when maxima are more evenly distributed throughout the storms.

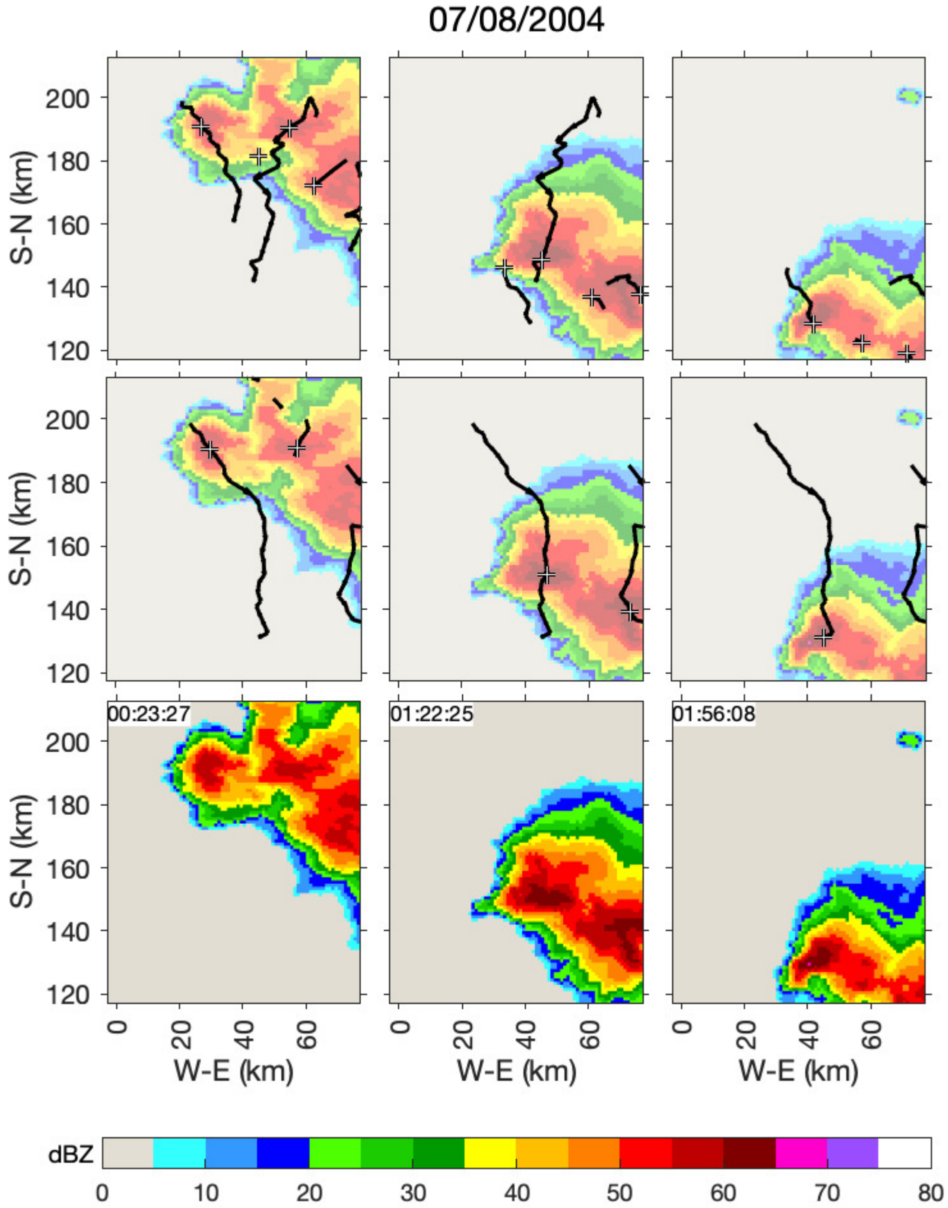


Figure 4.10: Detail area of three scans from the KDDC radar in Dodge City, KS beginning from 2004/07/08 00:23 UTC. Tracks from SCIT and SPA-MHT are shown. SCIT uses two irregular tracks to explain storm motion, while SPA-MHT uses one, more smooth track. The SPA-MHT track persists for longer.

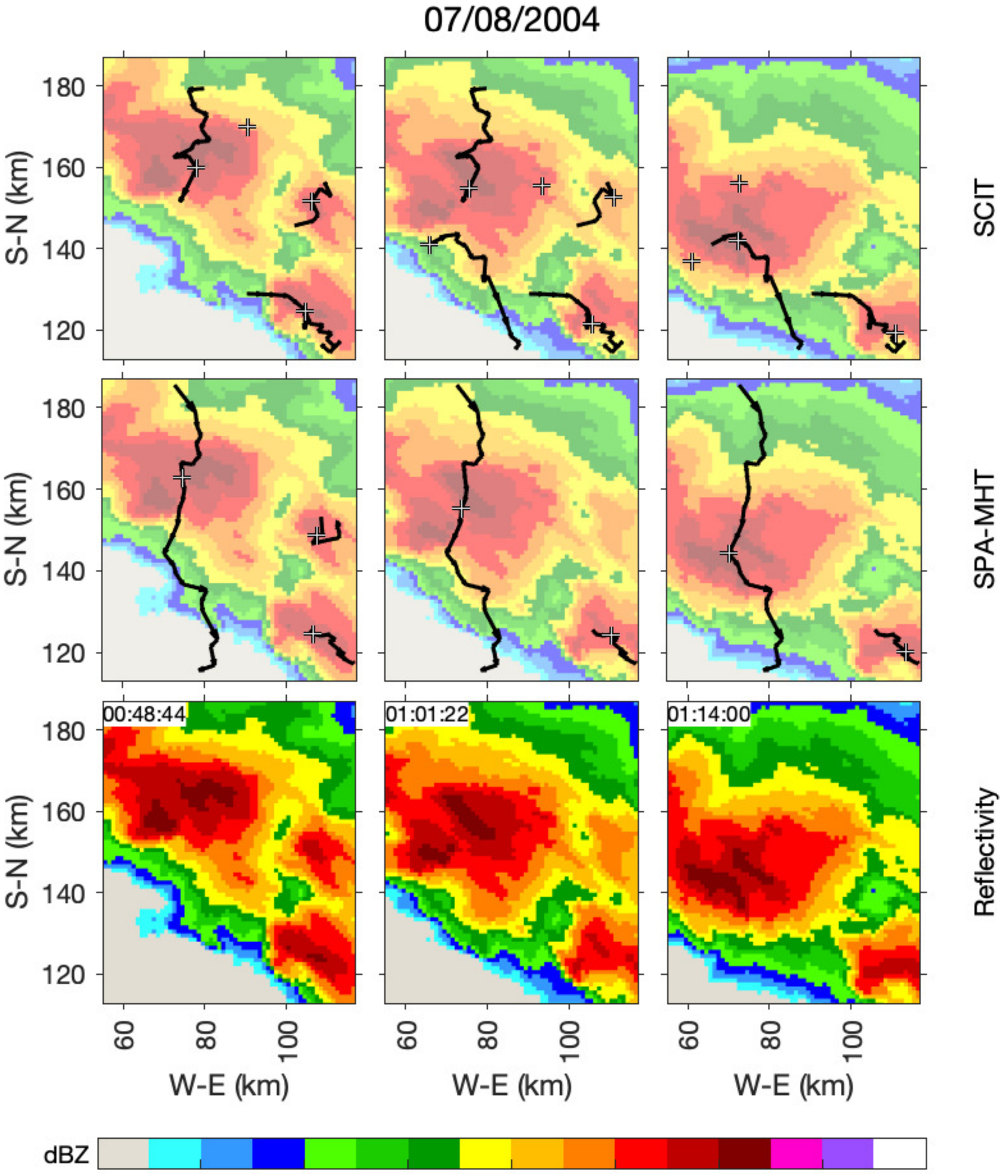


Figure 4.11: Detail area of three scans from the KDDC radar in Dodge City, KS beginning from 2004/07/08 00:48 UTC. Tracks from SCIT and SPA-MHT are shown. The western track from SCIT is terminated after a southern track is identified in the second scan, discarding information. SPA-MHT uses one long, continuous track to explain the same storm motion.

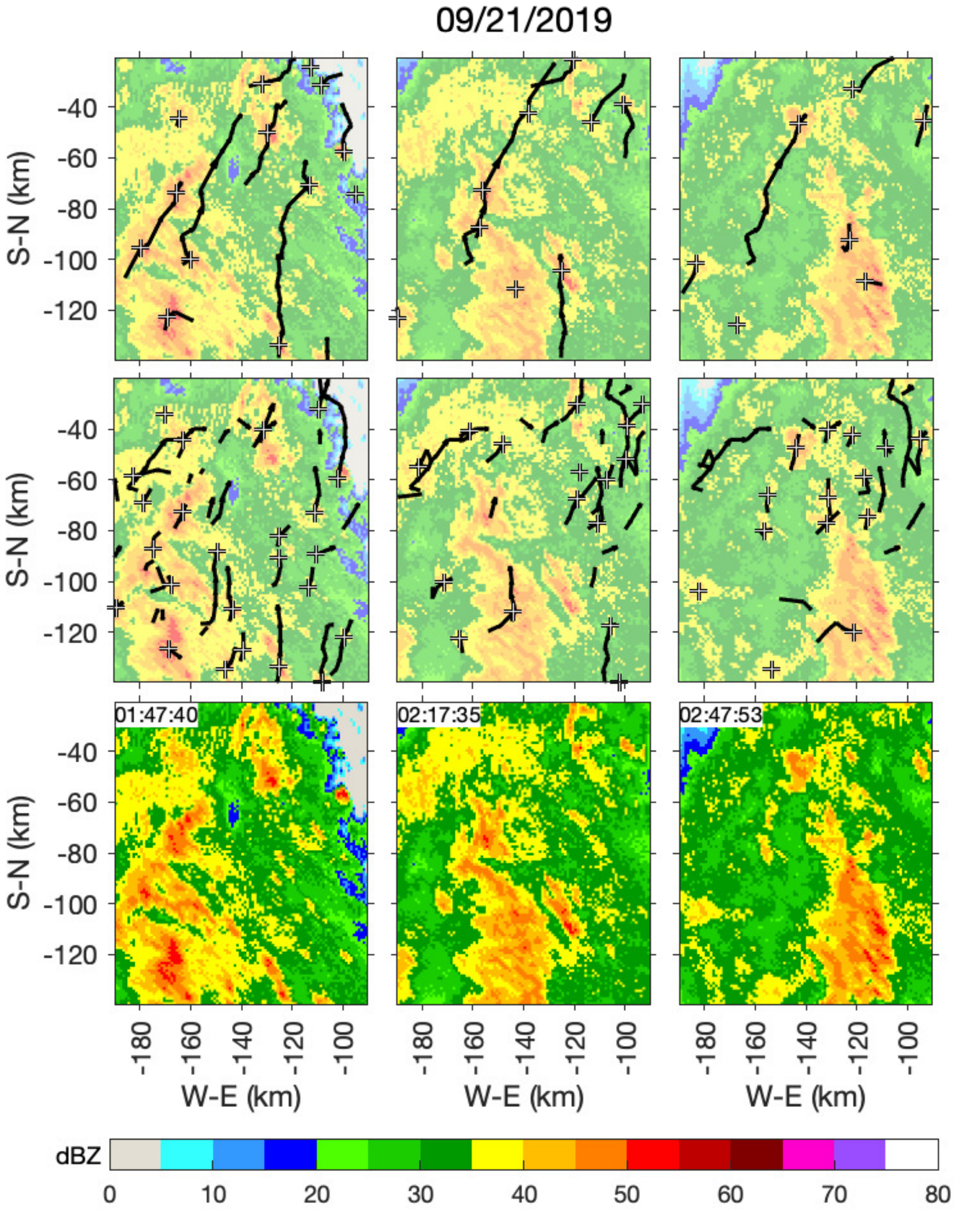


Figure 4.12: Detail area of three scans from the KTLX radar in Oklahoma City, OK beginning from 2019/09/21 01:47 UTC. Tracks from SCIT and SPA-MHT are shown, but poorly represent the west-to-east motion of the widespread storms. SCIT generates sparse tracks with some tracking effectiveness while SPA-MHT attempts many short tracks. Neither algorithm clearly identifies the motion of the intense southern region.

4.1.3 Squall Lines

A set of test cases observing the interaction of isolated storm cells and squall lines provided additional challenging tracking environments. In Figure 4.13, the initial narrow squall line is split into storm cells and tracked effectively by both algorithms.

As the squall line grows wider and more chaotic, features are more easily lost to merges or splits. The number of cells identified also visibly changes even when the area of storms does not change significantly. While some centroids are agreed upon by both algorithms and align with local reflectivity maxima, others are not repeated across methods. These spurious features are less consistent between scans than the features identified across methods and result in short and irregular tracks. The previously observed tendency of SPA-MHT to attempt tracking of weaker cells than SCIT made SPA-MHT to be somewhat more vulnerable in this situation. The average SCIT track in this region persisted for 7.08 scans on average compared to 6.29 scans for SPA-MHT tracks; the median duration for both sets of tracks was 5 scans. The apparent reduced density of SPA-MHT tracks visible in the figure is exacerbated by skipped measurements, which are not drawn in the figure.

SPA appears more sensitive than the SCIT identification stage to local changes of intensity. Only one strong point is required to identify a cell under SPA, and some regions within the squall line have high intensity variance, leading to the detection of additional spurious cells. The mechanism is similar to the increased detection of cells, compared to SCIT, in stratiform or other regions of weak reflectivity. The challenging squall line conditions do not affect successful tracking of more isolated cells ahead of and at the ends of the squall line.

Specific conditions within the squall line cases further degraded the performance of SPA-MHT. Figure 4.14 shows an isolated cell overtaken and absorbed

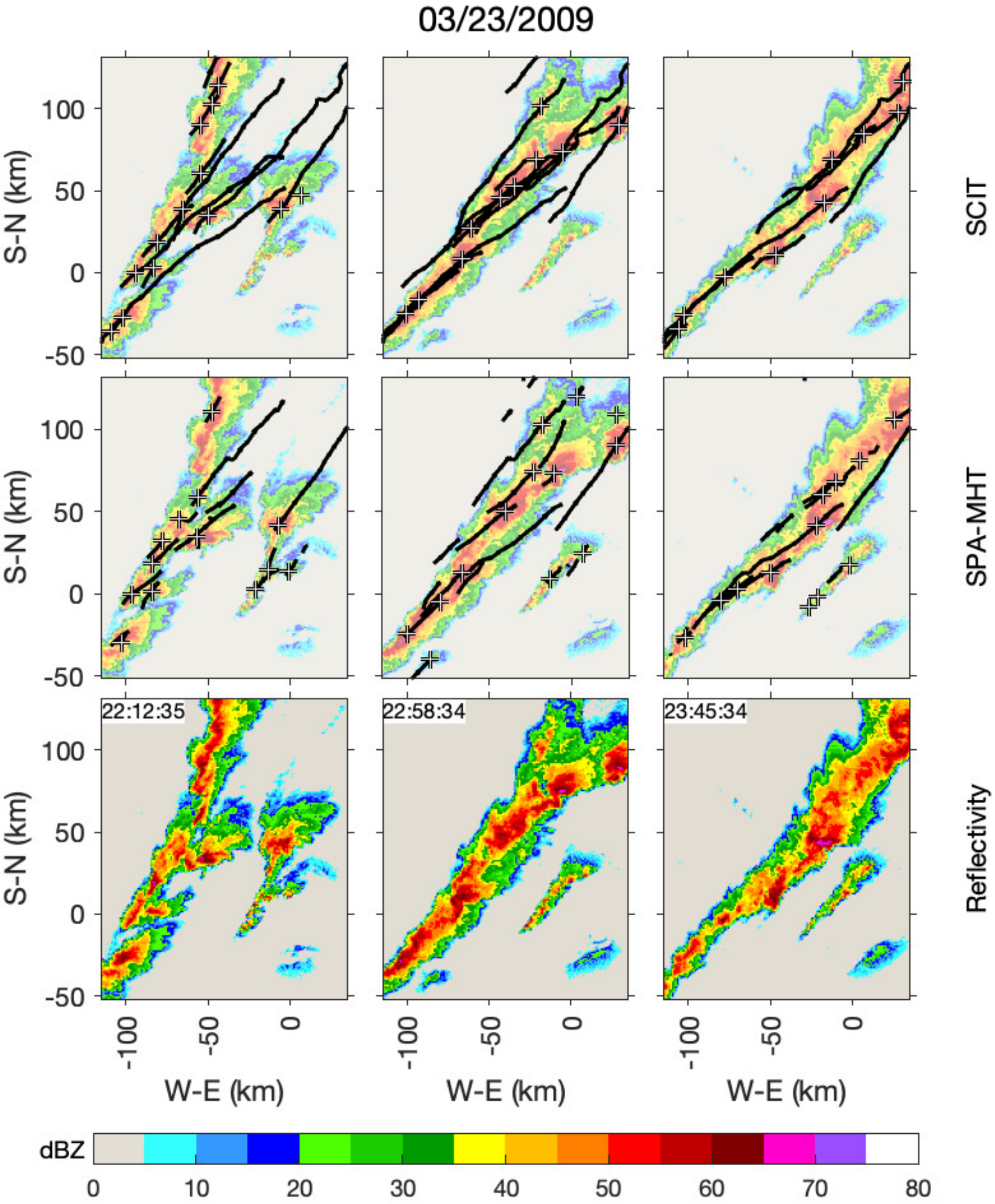


Figure 4.13: Detail area of three scans from the KTWX radar in Topeka, KS beginning from 2009/03/23 22:12 UTC. Tracks from SCIT and SPA-MHT are shown. Both algorithms demonstrate successful tracking across early scans, but the length of tracks decreases as the squall line grows wider and more intense. Many of the tracks visible in the third column are shorter and more fragmented for both methods.

into a squall line. In the second scan, the central merge between the cell and line has clearly visible structure. SCIT identifies three centroids from this structure, terminates tracks from the line, and continues the track from the isolated cell. SPA-MHT fails to distinguish the structure with multiple centroids, resulting in termination of nearly all tracks from before the merge. This feature extraction failure occurs when a wide spatial spread of pixels are similar in strength, forming a ”bridge” of strong points and causing the recursive process to erode only the edges of the storm. The SCIT feature extraction stage exceeded expectations on this test case, preserving longer tracks.

Challenging test cases, including the squall line mergers, stressed both algorithms. Improvements in SPA-MHT track consistency apparent in the simpler cases were less applicable to the challenge cases. The assumption that the recursive design of SPA would automatically break merged storms and squall lines into distinct, consistently identified cells was not well supported. The presence of strong point bridges in linear cases was able to prevent this breaking.

4.2 Quantifying Storm Tracking Performance

On visual inspection, SPA-MHT appears to generate intuitive tracks for isolated and closely spaced cases. Tracks smoothly follow isolated storms with less deviation than SCIT tracks. The missed detection support of MHT confers resistance to brief merges and splits. Permissive identification of cells results in more tracks of minor or irrelevant storms, but is advantageous for early detection of developing storms. The chaos of challenging squall line cases appears to negate these advantages. To validate these results, metrics to quantify storm tracking performance were evaluated.

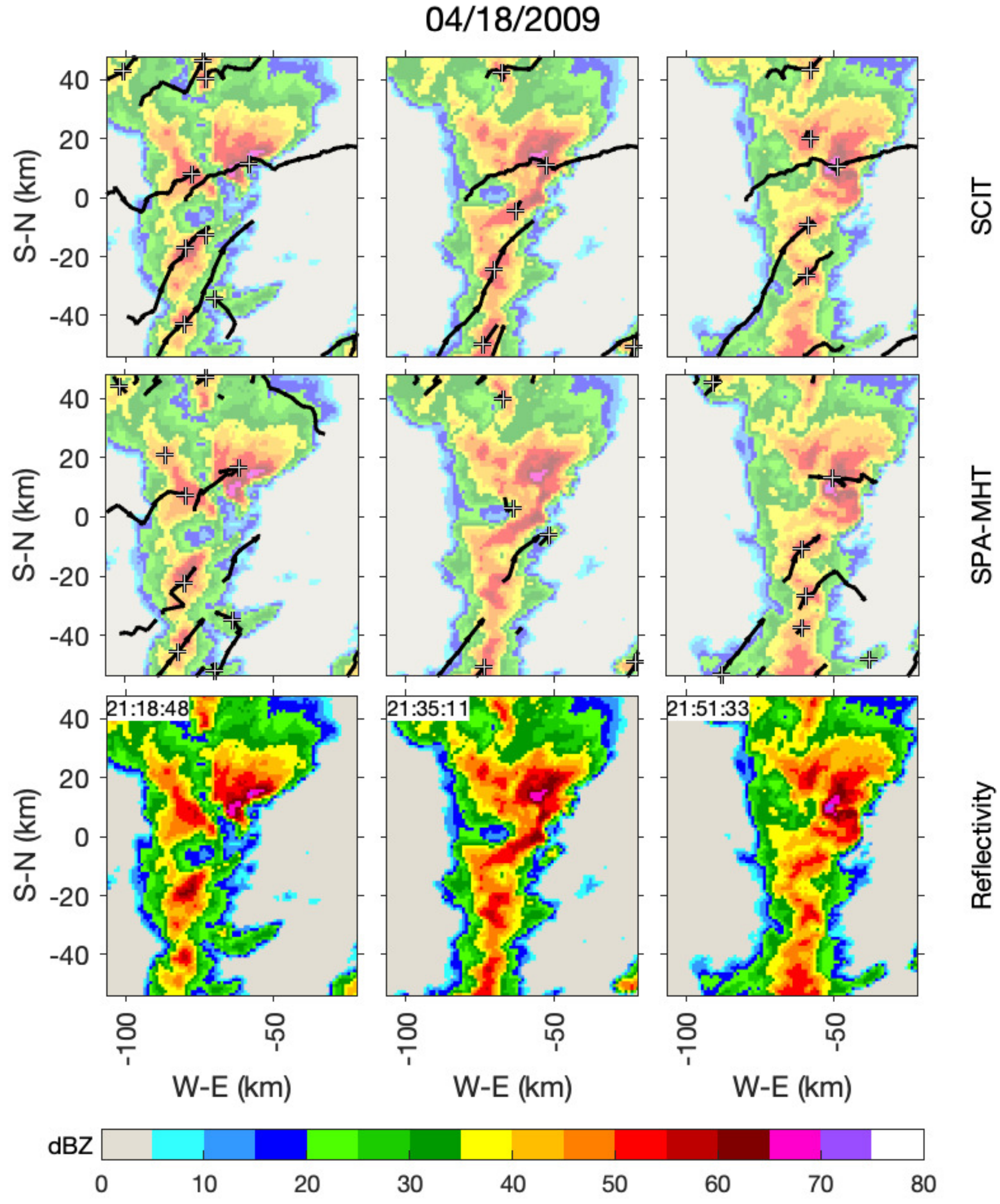


Figure 4.14: Detail area of three scans from the KVNX radar in Vance Air Force Base, OK beginning from 2009/04/18 21:18 UTC. Tracks from SCIT and SPA-MHT are shown. A squall line overtakes and merges with an isolated cell. SPA-MHT is challenged by the dense storm structure near the merger.

Direct numeric comparison of storm tracking algorithms is challenging. Forecasts may be evaluated on the basis of probability of detection (POD) and false alarm rate (FAR) if predictions are interpreted as binary decisions [7]. In such a design, a binary forecast image is generated where pixels are nonzero wherever the extent any storm cell is predicted to cover. This image is compared to the extents of storm cells detected in the actual data. A detection occurs when the forecast image and the data agree that a storm is present; a prediction in the forecast with no storm in the data is a false alarm. POD, FAR and similar scores have frequently been used to evaluate forecast accuracy despite drawbacks [10], [12], [17], [28]. The binary decision required for these metrics does not support forecasts of intensity such as growth and decay. Errors in predicted displacement are only represented by the amount of overlap between storm cells and forecast cells, biasing these metrics against methods which identify smaller cells. Efforts to measure intensity bias, growth and decay forecasts, or the displacement errors of forecasts rely on intensive human analysis [1].

The previous metrics emphasize the forecast stage of storm tracking. For this work, which evaluates the feature extraction and association stages, an alternative is preferred. A variant of POD may be calculated for the feature extraction stage alone if the binary image is generated with nonzero pixels wherever the algorithm identifies a storm [2]. To score track association, the ”percent correct” metric

$$PC = \frac{n_{\text{correct}}}{n_{\text{total}}} \times 100\% \tag{4.1}$$

relates the number of correct track associations to the total number of associations possible.

Both of these approaches require a source of ground truth. For feature extraction

POD, the intended storm cells must be provided, and for percent correct, every intended association must be provided. The use of another algorithm to obtain this ground truth precludes comparison; the maximum score possible for these metrics is obtained by perfect agreement with the ground truth. Obtaining unbiased ground truth for these metrics is achievable via human forecasters, but this approach is labor-intensive.

A proposed alternative to better support comparison between methods are the three metrics of median track duration, mismatch error in vertically integrated liquid (VIL), and linearity error [26]. These track metrics represent trade-offs in tracking algorithm design such that an algorithm optimized for one metric naturally will be penalized on one or both of the others. This allows for side-by-side comparison of algorithms applied to the same data without the need for ground truth data. The relative scores on each metric illustrate strengths and weaknesses of algorithms under consideration.

Track duration counts the number of scans in which each track is maintained. Severe, long lived storms ideally yield long, consistent tracks. Breaking such a track into several shorter tracks sacrifices the history of each track, reducing forecasting accuracy. An algorithm may optimize for track length by permitting associations even when the cells under consideration are distant or dissimilar, but such an algorithm will generate tracks with high linearity or mismatch error. The length of tracks over a test case is summarized by the median to reduce the impact of outliers.

The VIL mismatch of a track measures the standard deviation of VIL along that track. VIL is a statistic representing the amount of precipitation across the footprint of a storm cell [29]. For a storm which is not growing or decaying, VIL should remain consistent across scans. If such a storm is tracked correctly, the standard deviation of its VIL will be low. A high VIL mismatch indicates that the storm represented by the track may not be the same entity which was initially tracked.

Linearity error measures the deviation of the centroid positions assigned to a track from the line of best fit for that track. High linearity error may be caused by centroid motion artifacts, merge and split behavior, or incorrect track associations leading to jumps. Algorithms which optimize only to minimize linearity error, however, will improperly track some situations in which the actual storm motion is nonlinear, causing poor scoring in the duration or VIL mismatch metrics.

Of these three metrics, median track duration T and linearity error $\overline{e_{x,y}}$ were calculated for each test case. The duration of an individual track included skipped scans where the track was continued after the skip. Skips occurring at the end of a track, which resulted from from MHT evaluating a missed detection as more likely than a track termination, were not included in the track duration. The linearity error was calculated from the mean of the root-mean-square errors (RMSE), denoted by $e_{x,y}$, for tracks longer than the median duration. For a track with line of best fit

$$ax + by + c = 0 \tag{4.2}$$

consisting of n storm cells with centroids (x_n, y_n),

$$e_{x,y} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{(ax_n + by_n + c)^2}{a^2 + b^2}} \tag{4.3}$$

based on the minimum distance from each centroid to the line of best fit [30].

VIL mismatch was not evaluated due to the difficulty of obtaining VIL measures for storms identified by SCIT. While the footprint of storms reported by SPA-MHT is available and allows calculation of VIL, the source of SCIT data used reports only centroids. The VIL associated with SCIT cells is reported by a different NEXRAD

	Radar	Date	Scans	SCIT		SPA-MHT	
				T	$\overline{e_{x,y}}$	T	$\overline{e_{x,y}}$
4	KINX	05/30/2004	23	4	1.44	4	1.21
5	KDDC	07/08/2004	28	5	2.16	7	1.75
7	KLWX	10/06/2005	20	4	1.62	5	1.55
8	KTLX	05/04/1999	20	4	1.62	7	2.01
9	KTLX	05/19/2013	56	4	1.51	5	1.20
10	KTLX	05/03/2018	17	4	1.52	4	1.48
11	KTLX	06/07/2018	22	3	1.38	4	1.26
14	KTLX	09/21/2019	19	3	1.45	4	1.66
16	KICT	05/24/2008	26	3	1.50	3	1.41
17	KTWX	03/23/2009	26	4	1.33	5	1.45
18	KVNX	04/18/2009	42	5	1.58	5	1.43
Average			27	3.90	1.56	4.81	1.49

Table 4.2: Median track duration T and linearity error $\overline{e_{x,y}}$ for tracks found by SCIT and by SPA-MHT using test cases listed in Table 4.1. Cases with no or unusable SCIT data are excluded.

product [8]. Evaluation of the VIL mismatch metric remains a potential area of future work.

Table 4.2 contains the median track duration and linearity error calculated for the tracks generated by SCIT and SPA-MHT applied to each test case. Trivial tracks identified in only one scan were removed before calculating the metrics. Cases for which SCIT was unavailable are not included.

Across the test cases, SPA-MHT achieved comparable or superior track duration to SCIT. The longest median duration occurred on case 5, which featured a cluster of storms with slow, easily tracked movement, and case 8, which contained more long-lived but isolated storms in addition to some clustered storms. Case 4, which also featured long-lived storms, yielded tracks with a relatively low duration because of the initiation of many new tracks approximately halfway through the case. These initiations, such as those in Figure 4.4, were often identified later by SCIT. The penalty SPA-MHT accepts for earlier detection is a higher number of tracks, including short tracks of weak cells.

The median duration was somewhat nonspecific; the range of results was only between 3 and 7 despite the wide variance in weather conditions and number of scans per case.

For most cases, SPA-MHT achieved lower linearity error than SCIT. More consistent identification on isolated storms reduced linearity error compared to the deviating tracks of SCIT for cells which moved in straight lines. The resistance to brief splits and merges also contributed, as jumps in the track increased RMSE. As tracking complexity increased, the measured linearity error grew more similar between methods. SCIT achieved a lower linearity error on case 14, containing very chaotic and widespread storms as shown in Figure 4.12, and case 17, a challenging squall line merge pictured in Figure 4.13.

SPA-MHT also demonstrated unusually high linearity error on case 8 even as it effectively identified long tracks. Out of the 24 tracks on which RMSE was estimated in this case, two outliers accounted for the difference in linearity error between SCIT and SPA-MHT. One of these outliers was among the longest tracks identified in the case, accurately tracking the motion of a cell that split from the central storm before changing direction. This track is shown, alongside a SCIT track which eventually tracks the same underlying storm, in Figure 4.15. Its extreme spatial extent allowed for a high RMSE.

Averages for the median duration and linearity error of each method are also provided in Table 4.2. The cases surveyed are very diverse, trending from simpler cases to more complex squall lines, and the sample size is limited. The mean of the median durations suggests a commanding track length advantage for SPA-MHT. Given the four to six minute update time of NEXRAD, the average SPA-MHT length of 4.81 scans yields 3.6 to 5.5 minutes of tracking beyond that of SCIT.

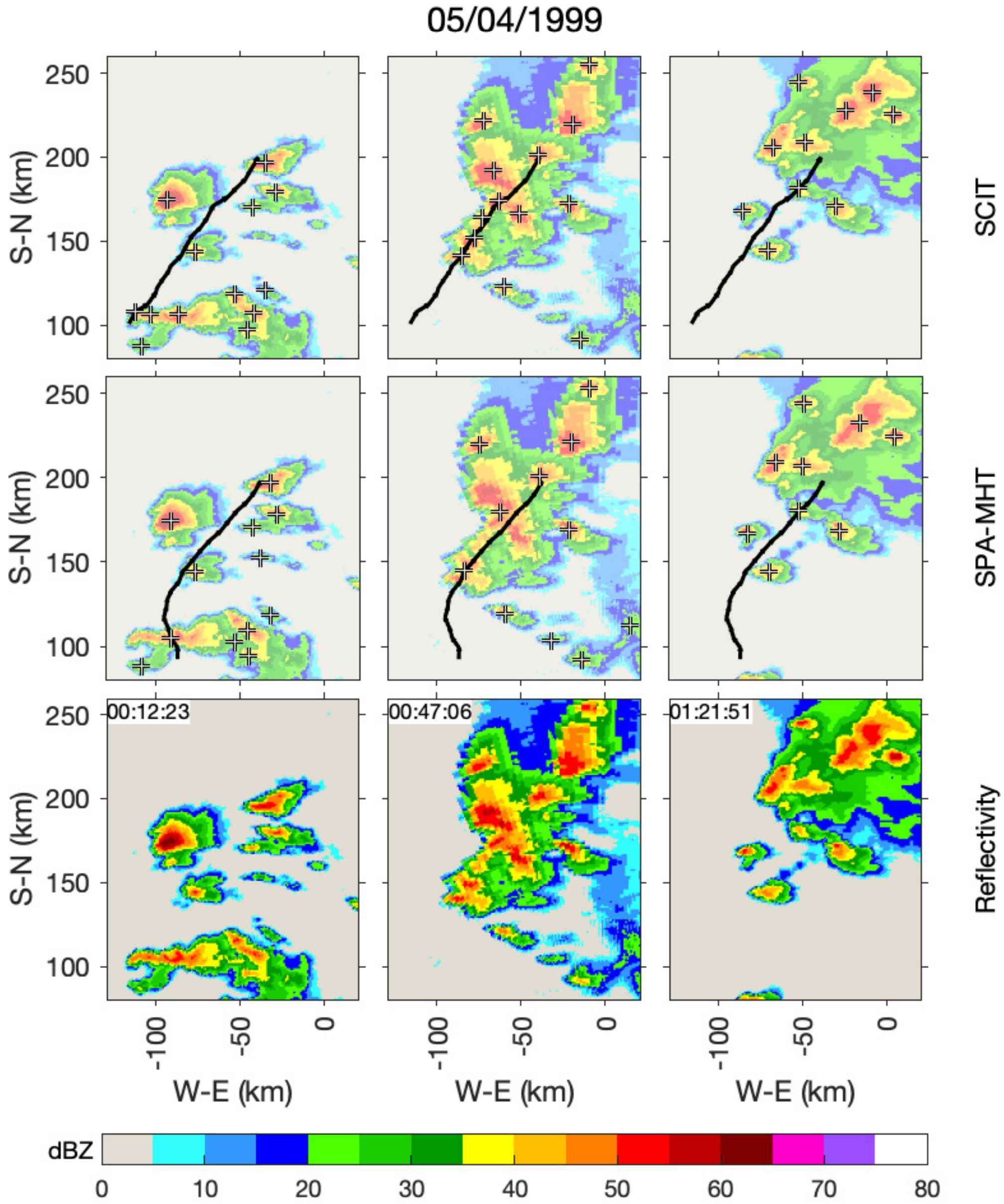


Figure 4.15: Detail area of three consecutive scans from the KTLX radar in Oklahoma City, OK beginning from 1999/05/04 00:12 UTC. One track from SCIT is shown alongside one from SPA-MHT. Other tracks are suppressed. The SPA-MHT track initially travels northwest, but the eventual trend of the track is to the northeast. The transition results in a highly nonlinear track. SCIT begins its track further west, avoiding the change in direction.

This average improvement is driven by the simpler cases; the more comparable performance on challenging squall line cases demonstrates that SPA-MHT has not achieved superiority on such cases.

While SPA-MHT achieves lower linearity error than SCIT on eight of the eleven evaluated cases, the average SPA-MHT error of 1.49 is similar to the SCIT average of 1.56. Considering the apparent smoothness of SPA-MHT tracks noted in this section, the closeness of this summary statistic was surprising. The same traits that allowed SPA-MHT to maintain longer tracks allowed the formation of curved or bent tracks in split or merge scenarios. In addition to increasing the sample size by identifying more test cases, future work may include evaluation of a similar metric using higher-order lines of best fit. This would provide a more realistic metric for the tracking of storms which demonstrate curving motion, such as those in case 8.

In the absence of VIL data, it remains possible that the performance advantages demonstrated by SPA-MHT on these metrics are due to decisions which would penalize VIL mismatch. However, this is not supported by visual assessment. SPA-MHT appears to identify cells which consistently track the same underlying storm in isolated and closely spaced cases. The ability to resume tracks after a brief merge or split should also benefit the VIL mismatch metric. It is on the more complex cases, especially squall lines, that SPA-MHT tracks become more visually confusing even as linearity error and median duration converge with SCIT. The predicted result of assessing the VIL mismatch metric is superiority on simple cases which converges to SCIT-comparable performance on squall lines, paralleling the trends noted for track duration and linearity error.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Effective storm tracking algorithms improve understanding of weather and enable advance warning of hazardous weather conditions. Centroid-based trackers implement an intuitive approach by tracking the most intense regions of storms, usually based on sequential volume scans of reflectivity from weather radar. The accuracy of this category of tracking algorithm for short-term forecasting depends on three stages: storm identification, track association, and forecasting. SPA-MHT addresses the first stage with SPA, a technique adapted from image processing which uses the erosion and dilation operations from mathematical morphology to identify storm cells consistently between scans. MHT is utilized for the track association stage, allowing track assignments to incorporate information from multiple scans.

This research demonstrated the advantages of SPA-MHT on observations of weather conditions of low to moderate complexity. When storms were isolated or occurred in closely spaced clusters, SPA-MHT generated similar or superior tracks to those generated by the SCIT algorithm. This improvement was suggested by the visual appearance of SPA-MHT tracks, which were often more smooth and longer than their counterparts, and was supported by numerical assessments of track duration and linearity. SPA identified cells which were more consistent from scan to scan, reducing the effect of storm shape changes on the resulting tracks. MHT made similar decisions to SCIT when there was little ambiguity, and its use of multiple scans of information allowed more accurate decisions when the correct assignment was briefly ambiguous. SPA-MHT also began tracking developing storms faster, providing increased lead time in forecasting. When applied to more challenging cases, such as squall lines, the performance of SPA-MHT degraded to be comparable to or marginally worse than SCIT. In the absence of separate reflectivity maxima, SPA was often unable to divide squall lines into consistent cells, resulting in different numbers of cells from scan to scan for short and irregular tracks.

5.2 Scientific Impact

The demonstrated tracking improvement on isolated storms and storm clusters suggests that the techniques used in SPA-MHT are promising routes to improve storm tracking. The evaluation of SPA provides further evidence for the utility of erosion- and dilation-based cell identification of consistent storm cells. This research also represents the first evaluation of MHT on actual weather data.

5.3 Future Work

Future research in this area falls into three main categories: further evaluation of track quality, improvements to cell identification, and improvements to track association. The first category includes the collection of additional data to confirm the trends identified in 4.2. More examples of each test case would provide more confidence in the metrics used. The collection of longer data may also reveal trends in the length of tracks, as for most test cases, only two hours of scans were used. An additional metric which was not evaluated in this research, VIL continuity, would assist in quantifying track assignment performance [26]. This metric indicates the tendency of an algorithm to assign similar cells to the same track.

To improve the cell identification stage, modifications to SPA may be explored. While consistency on isolated cells was already improved over SCIT, the behavior of SPA on large, intense ridges of reflectivity can be improved. Configuration of SPA parameters was insufficient to improve this performance. The use of a line observation, as opposed to the point observation of a centroid, is one potential avenue for this improvement; this would require significant alterations to the tracking algorithm as a whole. Alternately, an additional limit to the clustering step of SPA preventing the identification of extremely long cells might result in better representations of squall line observations as many adjacent cells.

Modifications to optimize the behavior of MHT on storm cell targets are also likely to improve performance. The current implementation of MHT, intended for the tracking of hard targets such as aircraft, relies on one-to-one assignment of cells to tracks [21]. A split guarantees the loss of all tracking history for all but one of the resulting storms; only one is allowed to continue the track. Likewise, a merge may only incorporate the history of one of its participant storms. An expansion of MHT which allows merge and split decisions, along with a principled evaluation of their likelihoods, would prevent the loss of information for these situations. This would, however, require substantial modifications to the hypothesis generation of MHT, possibly preventing an efficient implementation.

A potential simpler expansion of MHT is the incorporation of size and intensity information into the probabilities of track initiation and termination. The current model uses constant values for these probabilities; any hypothesis in which a new track begins incorporates the this constant into its total likelihood. In the case of true storm initiation, however, smaller storms are more likely candidates than large, developed storms. Likewise, track termination is more likely after several scans of storm decay than it is at the peak of development. Incorporation of this type of information into the hypotheses generated by MHT provides an appraoch to better adapt MHT for weather targets, improving the storm tracking algorithm as a whole.

Appendix A

Radar Data

The following figures illustrate the start and end states of test cases used to evaluate SPA-MHT in Chapter 4. For each case, the upper figure contains the constant-altitude Cartesian interpolation of the radar reflectivity observed during the first volume scan. The lower figure contains the corresponding reflectivity data for the final scan. Range rings are overlaid on each figure, spaced every 50 km from the radar out to the extent of the interpolated data at 300 km.

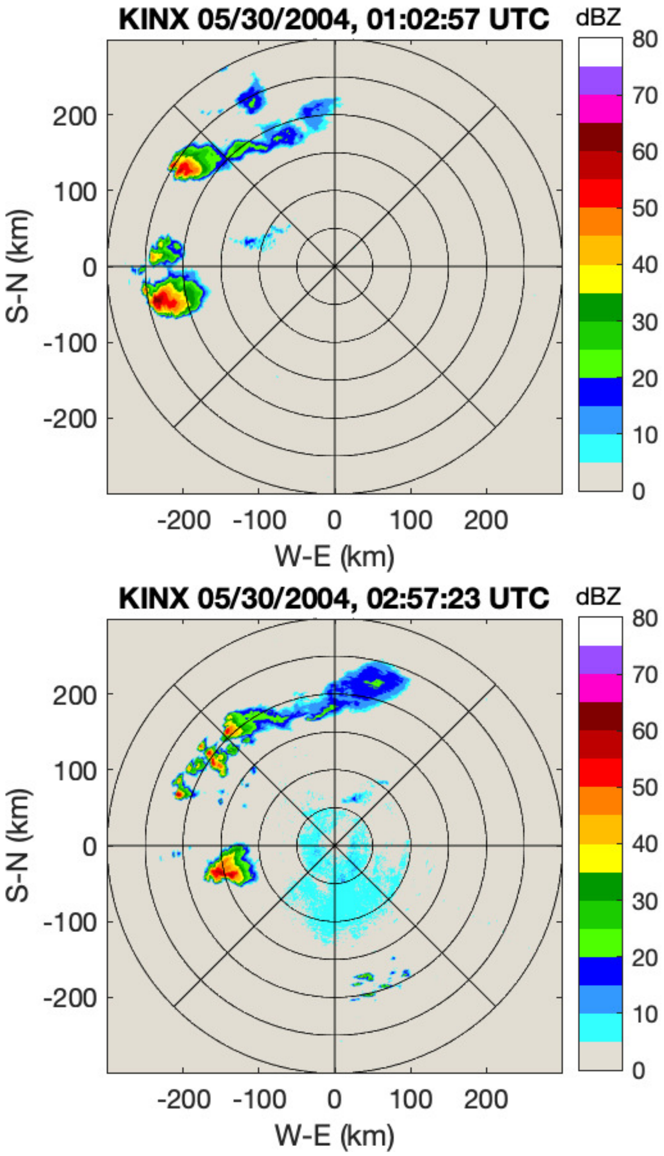


Figure A.1: Overview of case 4, observations by the KINX radar in Tulsa, OK. The first volume scan was initiated at 01:02 UTC on May 30th, 2004. The final volume scan was initiated at 02:57 UTC.

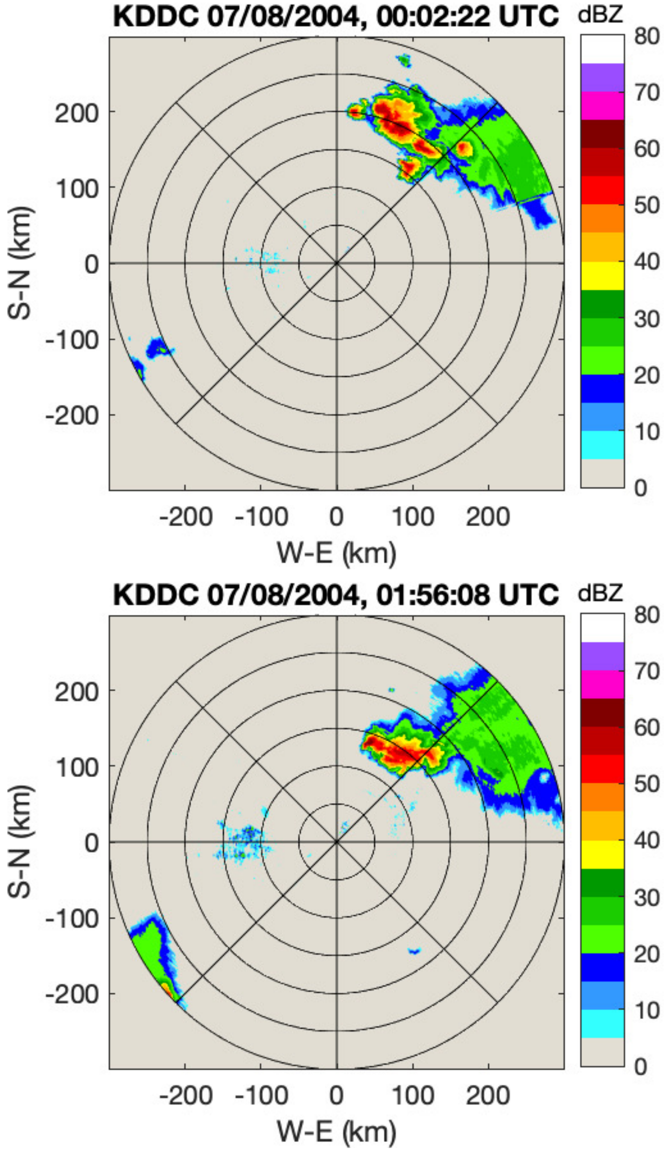


Figure A.2: Overview of case 5, observations by the KDDC radar in Dodge City, KS. The first volume scan was initiated at 00:02 UTC on July 8th, 2004. The final volume scan was initiated at 01:56 UTC.

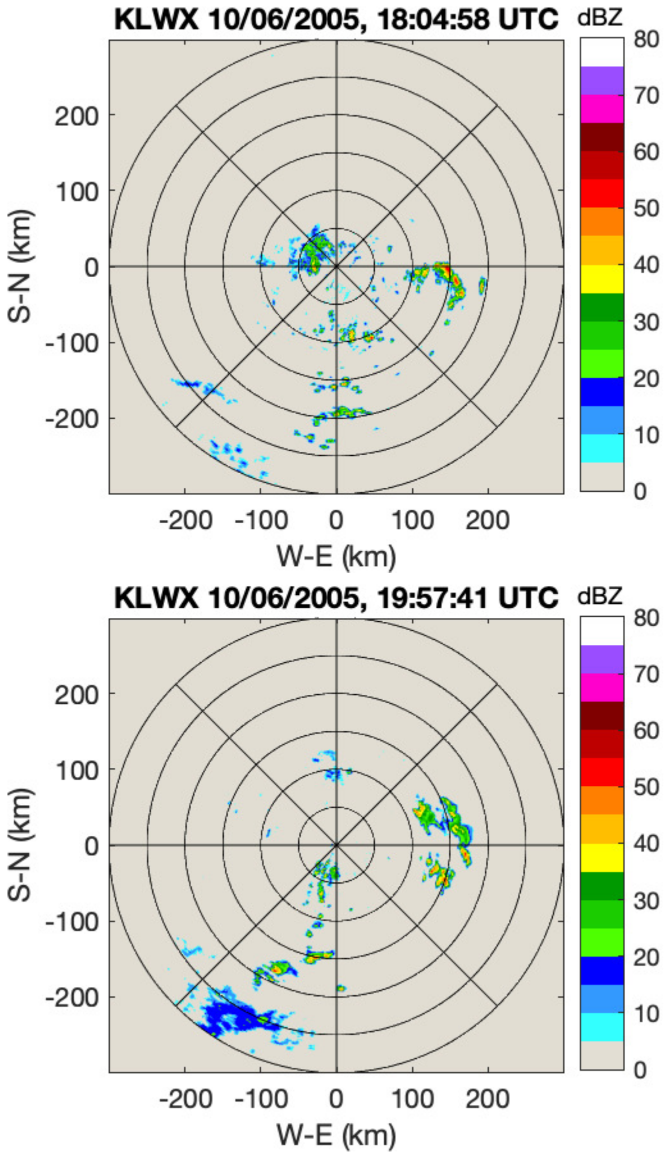


Figure A.3: Overview of case 7, observations by the KLWX radar in Sterling, VA. The first volume scan was initiated at 18:04 UTC on October 6th, 2005. The final volume scan was initiated at 19:57 UTC.

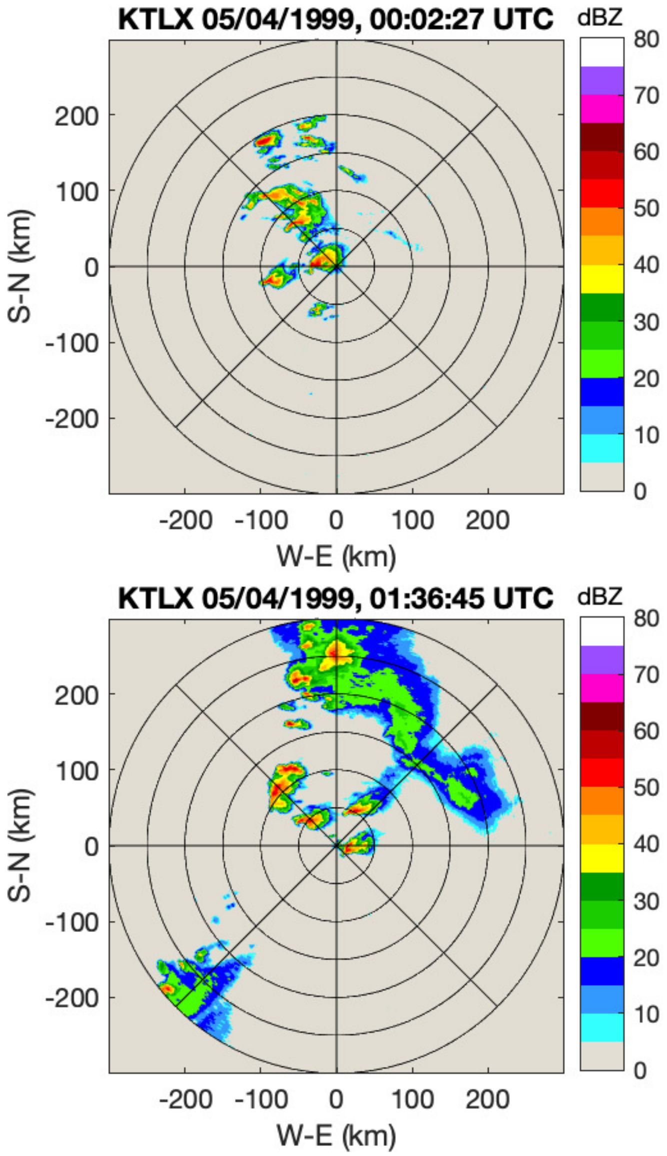


Figure A.4: Overview of case 8, observations by the KTLX radar in Oklahoma City, OK. The first volume scan was initiated at 00:02 UTC on May 4th, 1999. The final volume scan was initiated at 01:36 UTC.

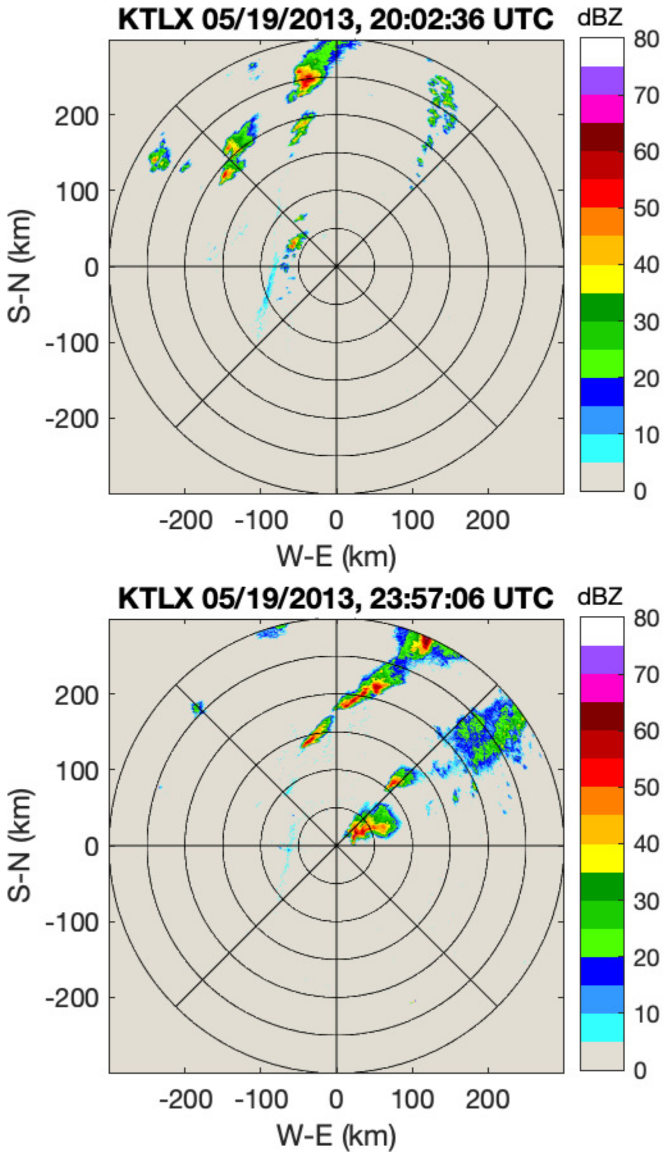


Figure A.5: Overview of case 9, observations by the KTLX radar in Oklahoma City, OK. The first volume scan was initiated at 20:02 UTC on May 19th, 2013. The final volume scan was initiated at 23:57 UTC.

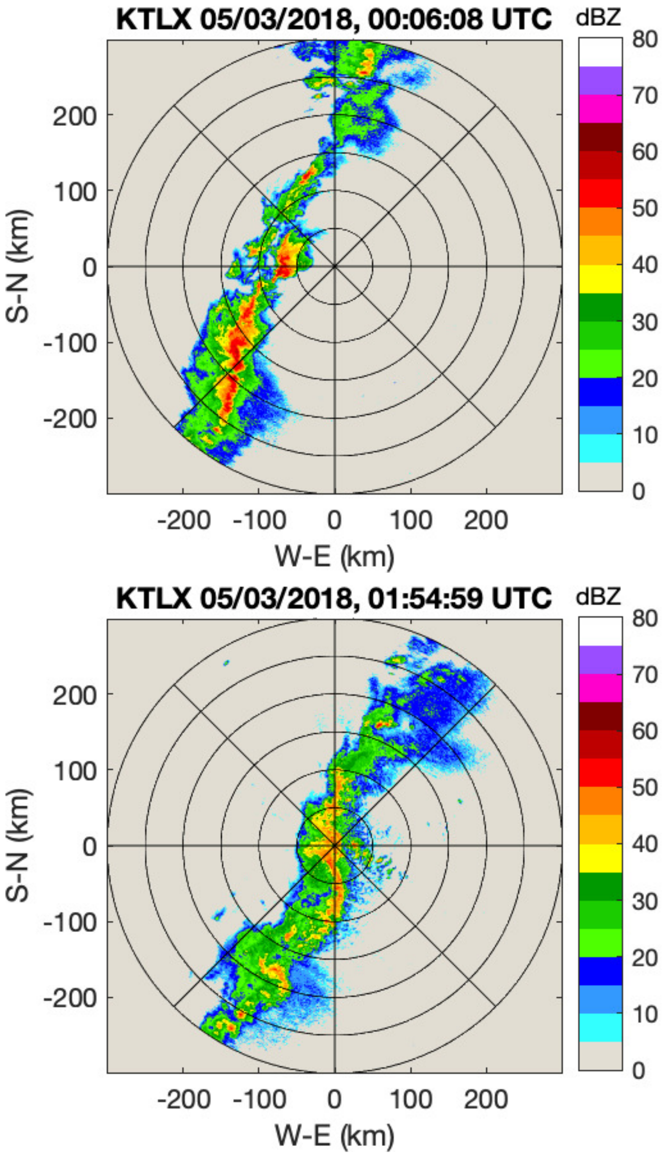


Figure A.6: Overview of case 10, observations by the KTLX radar in Oklahoma City, OK. The first volume scan was initiated at 00:06 UTC on May 3rd, 2018. The final volume scan was initiated at 01:54 UTC.

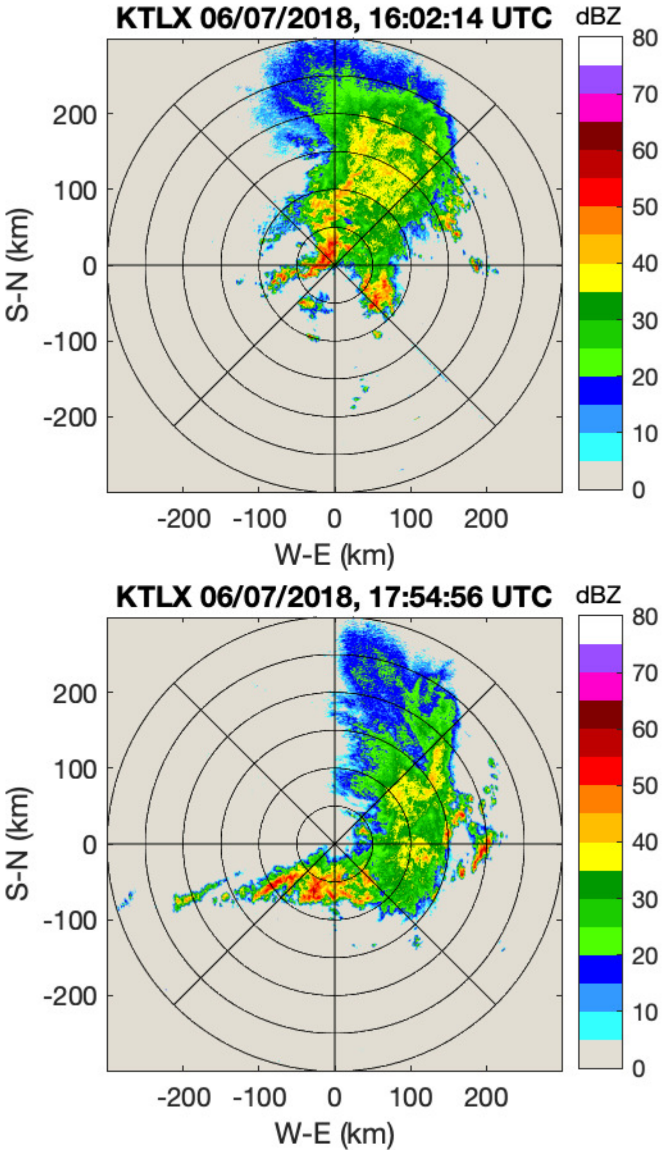


Figure A.7: Overview of case 11, observations by KTLX radar in Oklahoma City, OK. The first volume scan was initiated at 01:02 UTC on May 30th, 2004. The final volume scan was initiated at 02:57 UTC.

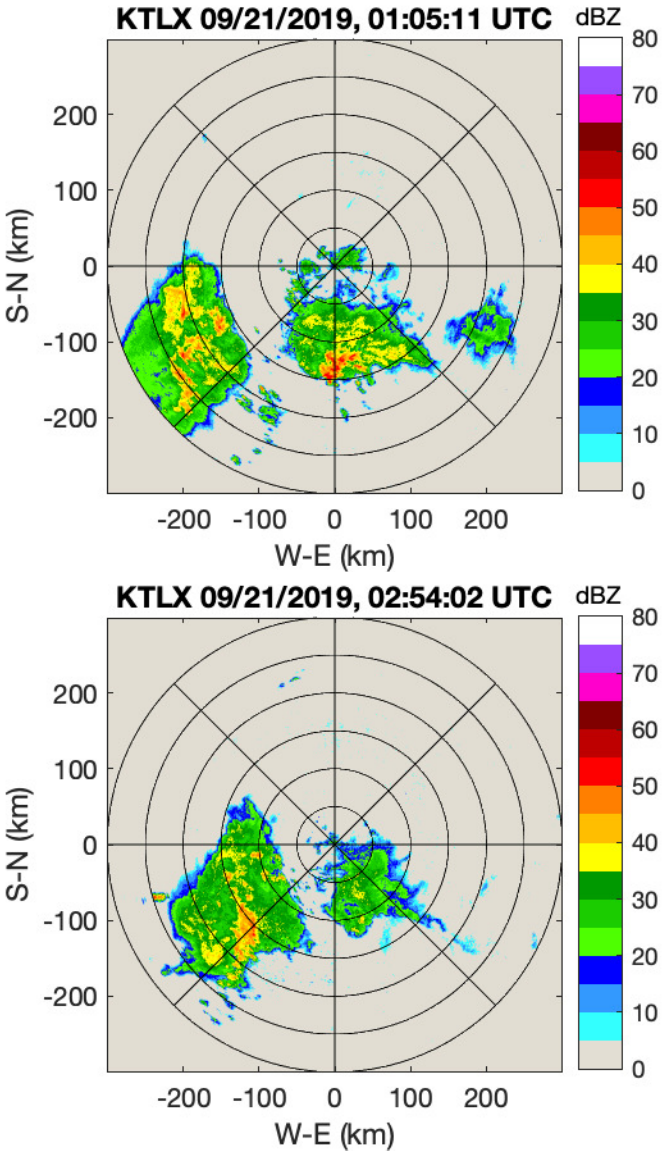


Figure A.8: Overview of case 14, observations by the KTLX radar in Oklahoma City, OK. The first volume scan was initiated at 01:05 UTC on September 21st, 2019. The final volume scan was initiated at 02:54 UTC.

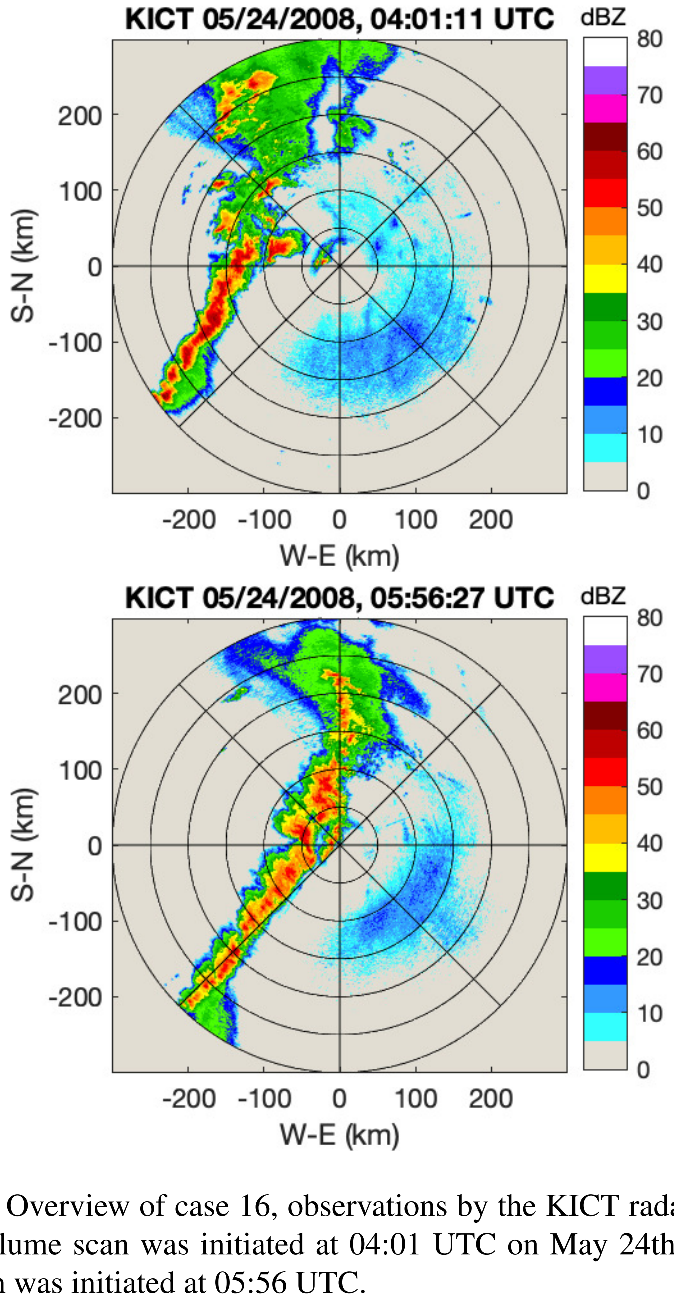


Figure A.9: Overview of case 16, observations by the KICT radar in Wichita, KS. The first volume scan was initiated at 04:01 UTC on May 24th, 2008. The final volume scan was initiated at 05:56 UTC.

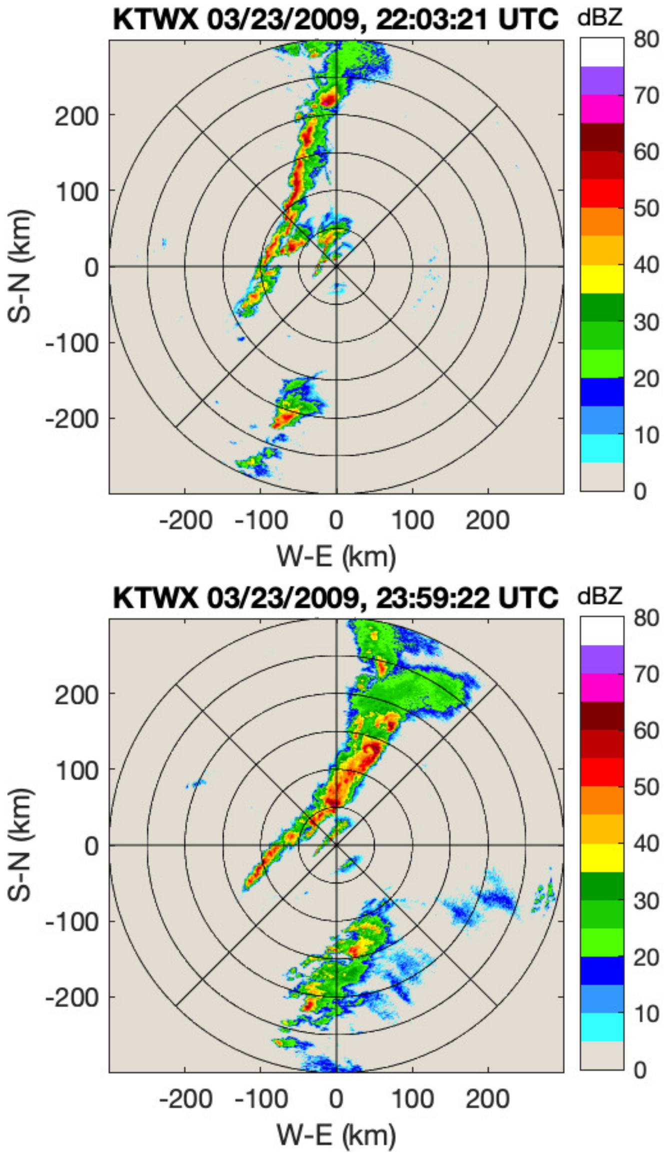


Figure A.10: Overview of case 17, observations by the KTWX radar in Sterling, VA. The first volume scan was initiated at 22:03 UTC on March 23rd, 2009. The final volume scan was initiated at 23:59 UTC.

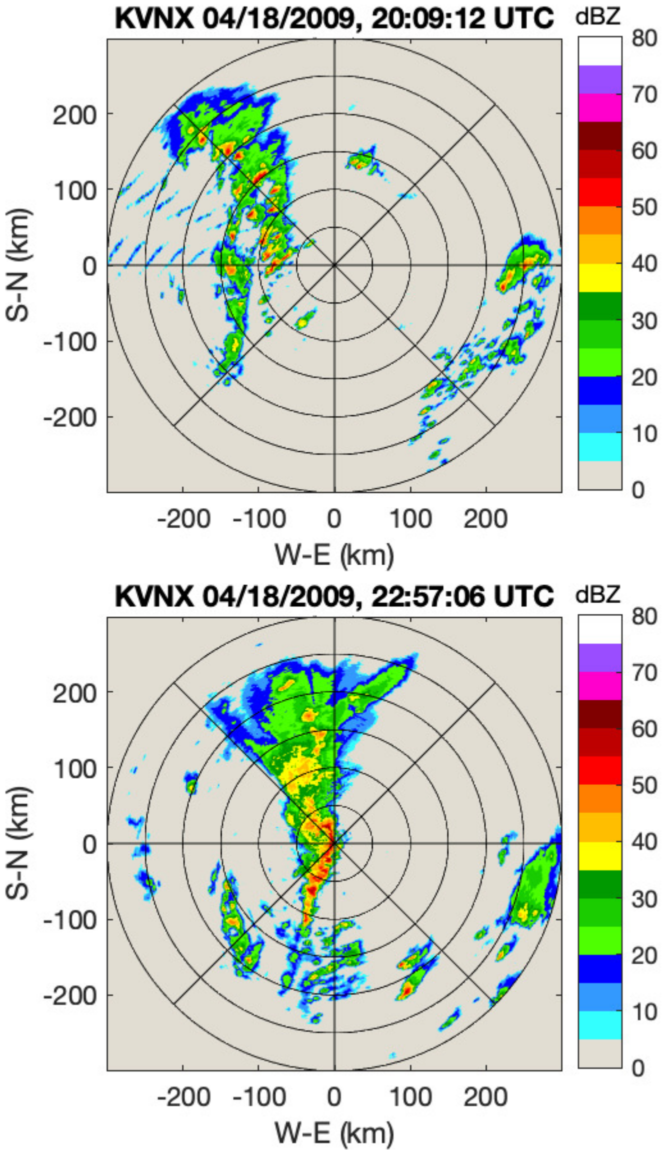


Figure A.11: Overview of case 18, observations by the KVNX radar at Vance Air Force Base, OK. The first volume scan was initiated at 20:09 UTC on April 18th, 2009. The final volume scan was initiated at 22:57 UTC.

References

[1] J. W. Wilson, E. E. Ebert, T. R. Saxen, R. D. Roberts, C. K. Mueller, M. Sleigh, C. E. Pierce, and A. Seed, “Sydney 2000 forecast demonstration project: Convective storm nowcasting”, *Weather and Forecasting*, vol. 19, no. 1, pp. 131–150, 2004. DOI: 10.1175/1520-0434(2004)019⟨0131:SFDPCS⟩2.0.CO;2.

[2] J. T. Johnson, P. L. MacKeen, A. Witt, E. D. W. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, “The storm cell identification and tracking algorithm: An enhanced wsr-88d algorithm”, *Weather and Forecasting*, vol. 13, no. 2, pp. 263–276, 1998.

[3] R. K. Crane, “Automatic cell detection and tracking”, *IEEE Transactions on Geoscience Electronics*, vol. 17, pp. 250–262, 1979.

[4] R. E. Rinehart and E. T. Garvey, “Three-dimensional storm motion detection by conventional weather radar”, *Nature*, vol. 273, no. 5660, pp. 287–289, May 1978. DOI: 10.1038/273287a0.

[5] J. Tuttle and R. Gall, “A single-radar technique for estimating the winds in tropical cyclones”, *Bulletin of the American Meteorological Society*, vol. 80, no. 4, pp. 653–668, Apr. 1999. DOI: 10.1175/1520-0477(1999)080⟨0653:ASRTFE⟩2.0.CO;2.

[6] D. Rosenfeld, “Objective method for analysis and tracking of convective cells as seen by radar”, *Journal of Atmospheric and Oceanic Technology*, vol. 4, no. 3, pp. 422–434, 1987. DOI: 10.1175/1520-0426(1987)004⟨0422:OMFAAT⟩2.0.CO;2.

[7] M. Dixon and G. Wiener, “Titan: Thunderstorm identification, tracking, analysis, and nowcasting—a radar-based methodology”, *Journal of Atmospheric and Oceanic Technology*, vol. 10, no. 6, pp. 785–797, Jan. 1993. DOI: 10.1175/1520-0426(1993)010⟨0785:TTITAA⟩2.0.CO;2.

[8] C. Stephenson, *Interface control document for the RPG to class 1 user*, Document Number 2620001X, WSR-88D Radar Operations Center, Norman, OK, Jan. 2018.

[9] C. E. Pierce, P. J. Hardaker, C. G. Collier, and C. M. Haggett, "GANDOLF: A system for generating automated nowcasts of convective precipitation", *Meteorological Applications*, vol. 7, no. 4, pp. 341–360, 2000. DOI: 10.1017/S135048270000164X.

[10] C. Mueller, T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and J. Yee, "Ncar auto-nowcast system", *Weather and Forecasting*, vol. 18, no. 4, pp. 545–561, 2003. DOI: 10.1175/1520-0434(2003)018⟨0545:NAS⟩2.0.CO;2.

[11] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey", *ACM Comput. Surv.*, vol. 38, no. 4, Dec. 2006. DOI: 10.1145/1177352.1177355.

[12] L. Han, S. Fu, L. Zhao, Y. Zheng, H. Wang, and Y. Lin, "3d convective storm identification, tracking, and forecasting—an enhanced TITAN algorithm", *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 4, pp. 719–732, 2009. DOI: 10.1175/2008JTECHA1084.1.

[13] B. Root, T. Yu, and M. Yeary, "Consistent clustering of radar reflectivities using strong point analysis: A prelude to storm tracking", *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 2, pp. 273–277, Mar. 2011. DOI: 10.1109/LGRS.2010.2070787.

[14] V. Manfredi, S. Mahadevan, and J. Kurose, "Switching kalman filters for prediction and tracking in an adaptive meteorological sensing network", in *2005 Second Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2005. IEEE SECON 2005.*, Sep. 2005, pp. 197–206. DOI: 10.1109/SAHCN.2005.1557075.

[15] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking", *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, Jan. 2004. DOI: 10.1109/MAES.2004.1263228.

[16] V. Lakshmanan, K. Hondl, and R. Rabin, "An efficient, general-purpose technique for identifying storm cells in geospatial images", *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 3, pp. 523–537, 2009. DOI: 10.1175/2008JTECHA1153.1.

[17] W. Hand, “An object-oriented technique for nowcasting heavy showers and thunderstorms”, *Meteorological Applications*, vol. 3, no. 1, pp. 31–41, 1996.

[18] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2008.

[19] A. AbuBaker, R. Qahwaji, S. Ipson, and M. Saleh, “One scan connected component labeling technique”, in *2007 IEEE International Conference on Signal Processing and Communications*, Nov. 2007, pp. 1283–1286. DOI: 10.1109/ICSPC.2007.4728561.

[20] D. Reid, “An algorithm for tracking multiple targets”, *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979. DOI: 10.1109/TAC.1979.1102177.

[21] I. J. Cox and S. L. Hingorani, “An efficient implementation of reid’s multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 138–150, Feb. 1996. DOI: 10.1109/34.481539.

[22] K. G. Murty, “An algorithm for ranking all the assignments in order of increasing cost”, *Operations Research*, vol. 16, no. 3, pp. 682–687, 1968.

[23] C. W. Therrien, *Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics*. New York, NY, USA: John Wiley & Sons, Inc., 1989.

[24] B. Root, “Performance analysis of storm cell tracking with multiple hypothesis tracking”, unpublished.

[25] National Centers for Environmental Information. (2020). Radar data, National Oceanic and Atmospheric Administration, [Online]. Available: https://www.ncdc.noaa.gov/data-access/radar-data (visited on 01/17/2020).

[26] V. Lakshmanan and T. Smith, “An objective method of evaluating and devising storm-tracking algorithms”, *Weather and Forecasting*, vol. 25, no. 2, pp. 701–709, 2010. DOI: 10.1175/2009WAF2222330.1.

[27] A. J. French and M. D. Parker, “Observations of mergers between squall lines and isolated supercell thunderstorms”, *Weather and Forecasting*, vol. 27, no. 2, pp. 255–278, 2012. DOI: 10.1175/WAF-D-11-00058.1.

[28] J. W. Wilson, N. A. Crook, C. K. Mueller, J. Sun, and M. Dixon, "Nowcasting thunderstorms: A status report", *Bulletin of the American Meteorological Society*, vol. 79, no. 10, pp. 2079–2100, Oct. 1998.

[29] D. R. Greene and R. A. Clark, "Vertically integrated liquid water—a new analysis tool", *Monthly Weather Review*, vol. 100, no. 7, pp. 548–552, 1972. DOI: 10.1175/1520-0493(1972)100⟨0548:VILWNA⟩2.3.CO;2.

[30] R. Larson and R. Hostetler, "Precalculus: A concise course", in, 1st ed. Houghton Mifflin, 2007, p. 522.