11-2012

An Analysis of Deep Convection Initiation Environments

Noah Lock
University of Nebraska-Lincoln, nlock@huskers.unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/geoscidiss
Part of the Atmospheric Sciences Commons, and the Meteorology Commons

http://digitalcommons.unl.edu/geoscidiss/33

This Article is brought to you for free and open access by the Earth and Atmospheric Sciences, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations & Theses in Earth and Atmospheric Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
AN ANALYSIS OF DEEP CONVECTION INITIATION ENVIRONMENTS

by

Noah A. Lock

A THESIS

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Master of Science

Major: Earth and Atmospheric Sciences

Under the Supervision of Professor Adam Houston

Lincoln, Nebraska

November, 2012
Initiation is the part of the convective life cycle which is currently least understood and least well forecast. The inability to properly forecast the timing and/or location of deep convection initiation degrades forecast skill, especially during the warm season. The goals of this research are to examine the spatiotemporal distribution of thunderstorm initiation points and to determine which atmospheric parameters (and ultimately processes) are most important for the initiation of thunderstorms. The spatiotemporal distribution of thunderstorm initiation points shows the expected peaks during summer and during the afternoon. The warm season also produces significant concentrations of initiation points near mountains, mainly in the western part of the analysis domain. The selected atmospheric parameters computed at initiation points are compared with those obtained from nearby areas where storms did not form. Analysis of these parameters shows that there is no threshold of any single parameter that effectively discriminates between initiation and non-initiation in all cases. However, case-by-case comparison of the values showed that lift is most often the factor that distinguishes the thunderstorm initiation environment from other areas.
ACKNOWLEDGEMENTS

I wish to acknowledge the assistance of my advisor and committee members.

I also want to thank the other members of the Severe Storms Research Group, in particular George Limpert and Cody Oppermann, for numerous productive discussions.

I would like to acknowledge Dr. Valiappa Lakshmanan, who provided the WDSS-II software and assistance in using it properly.

This work would not have been feasible without the use of the computing resources at the Holland Computing Center at the University of Nebraska-Lincoln and the assistance of the staff, in particular Dr. Adam Caprez, who helped us run some of our processing on the Open Science Grid.

I also wish to thank my sources of funding for my time here, including the Earth and Atmospheric Sciences department and National Science Foundation Award No. AGS-0757189.

Lastly, I wish to acknowledge the sources of data for this project, the National Climatic Data Center (radar and model data), Iowa State University (lightning data), and the PRISM Climate Group (precipitation data).
## TABLE OF CONTENTS

CHAPTER 1: Introduction .................................................................................1

CHAPTER 2: Spatiotemporal Distribution of Initiation Points .........................3
  2.1: Introduction ..........................................................................................3
  2.2: Methodology .......................................................................................4
    2.2.1: Radar-based thunderstorm identification ......................................4
  2.2.2: Selection of Initiation Points .............................................................7
  2.3: Results ..................................................................................................8
  2.4: Discussion ............................................................................................12
  2.5: Conclusions ........................................................................................22

CHAPTER 3: Environmental Conditions of Thunderstorm Initiation ...............23
  3.1: Introduction ..........................................................................................23
  3.2: Description of Parameters ...................................................................26
  3.3: Methodology ........................................................................................33
    3.3.1: Selection of Points for Parameter Selection ....................................33
      3.3.1.1: Initiation Points .......................................................................33
      3.3.1.2: Null Points .............................................................................34
    3.3.2: Attribution of Parameter Values to Cataloged Points .....................37
  3.4: Results ..................................................................................................38
    3.4.1: Analysis of Raw Values ..................................................................38
    3.4.2: Surface-based vs Elevated Convection .........................................42
    3.4.3: Pairwise Differences .....................................................................47
3.5: Discussion ..................................................................................59
3.6: Conclusions ..................................................................................65

CHAPTER 4: Summary and Conclusions ..............................................67

APPENDIX A: StratFilter Algorithm ...................................................71

APPENDIX B: ThOR Training and Verification .....................................76
  B.1: Training .....................................................................................76
  B.2: Verification ................................................................................84
LIST OF MULTIMEDIA OBJECTS

Fig. 2.1: Map of study area .................................................................5
Fig. 2.2: Kernel density of initiation points ...........................................8
Fig. 2.3: Elevation map for study area ..................................................9
Fig. 2.4: Annual distribution of initiation points .....................................10
Fig. 2.5: Kernel density of monthly initiation points .............................11
Fig. 2.6: Diurnal distribution of initiation points .................................12
Fig. 2.7: Kernel density over expanded domain .................................14
Fig. 2.8: Annual precipitation anomalies .............................................15
Table 2.1: Summary of initiation points and tracks by year .................16
Fig. 2.9: Overlay of precipitation and initiation points for selected months ....18
Fig. 2.10: Kernel density of initiation points with 50 km threshold ...........20
Fig. 2.11: Annual distribution of initiation points with 50 km threshold ....21
Fig. 2.12: Diurnal distribution of initiation points with 50 km threshold ....22
Table 3.1: List of parameters to be used ..............................................26
Table 3.2: Distance thresholds used ...................................................33
Fig. 3.1: Illustration of initiation and null point identification methods .......35
Fig. 3.2: Raw CAPE and CIN values ..................................................40
Fig. 3.3: Raw values of other parameters ............................................41
Table 3.3: Separation values for raw data at 120 km ............................42
Fig. 3.4: CAPE and CIN for surface-based cases .................................44
Table 3.4: Separation values for surface-based data ...............................45
Fig. 3.5: Other useful parameters for surface-based cases ......................45
Table 3.5: Separation values for elevated data ......................................46
Fig. 3.6: Convergence, omega, and Δz* for elevated cases .........................46
Fig. 3.7: Comparison of mixing ratio deficit in surface-based and elevated cases ......47
Fig. 3.8: Normalized differences of CAPE and CIN ........................................49
Fig. 3.9: Normalized differences for other parameters .....................................50
Fig. 3.10: Histogram of normalized differences in 0-LFC convergence ..............52
Table 3.6: Ranking of best parameters by significant differences at 60 km ..........54
Table 3.7: Ranking of best parameters by significant differences at 120 km ........55
Table 3.8: Ranking of best parameters by significant differences at 180 km .........56
Table 3.9: Frequency that each parameter was the only one different ...............59
Fig. A.1: Score as a function of reflectivity gradient .....................................73
Fig. A.2: Fuzzy logic membership functions ..................................................74
Fig. A.3: Example application of stratFilter algorithm .....................................75
Table B.1: List of events used for ThOR training ...........................................77
Table B.2: List of parameters derived from ThOR training ..............................78
Fig. B.1: Radii for correct and incorrect continuations at 5 minutes .................80
Fig. B.2: Radii for correct and incorrect continuations at 10 minutes ..............82
Table B.3: List of events used for human tracking in verification .....................85
Table B.4: List of possible outcomes ............................................................87
Table B.5: Verification statistics based on comparison to human tracks ............87
Table B.6: Aggregate statistics for short events ............................................88
Table B.7: Aggregate statistics for long events .............................................89
CHAPTER 1: INTRODUCTION

Deep convection is key to the redistribution of heat and momentum within the atmosphere. The distribution of precipitation is also closely related to deep convection, with up to 70% of annual rainfall in portions of the Great Plains estimated to come from thunderstorms (Changnon 2001). In addition, hazards associated with thunderstorms (tornadoes, hail, wind, flooding) have been responsible for an average of 57% of annual insured catastrophe losses for years since 1953 (Folger 2011). The reason to focus on thunderstorm initiation in particular is that initiation is the part of the convective life cycle which is least understood and least skillfully forecast (Markowski et al. 2006). In this work, identification of thunderstorm initiation principally relies on radar data. Analysis focuses on the Great Plains from 2005-2007.

The first component of this study is an analysis of the spatiotemporal distribution of the “first initiation”. These results show that the spatiotemporal distribution is consistent with expectations, with maximum activity in summer afternoons. However, there are significantly fewer initiation points in 2007 than either 2005 or 2006. Possible explanations for this are explored.

The second component of the study is a comparison of atmospheric parameters derived from 20-km RUC-II model analysis fields at the identified initiation points with the same parameters calculated at null points, defined as locations where thunderstorm initiation did not occur. The approach is a modification of the typical “ingredients-based” approach (e.g., Johns and Doswell 1992) used to examine thunderstorm initiation. Rather than the traditional ingredients of instability, moisture, and lift, this study will examine deep convection through the “relevant factors” of buoyancy, dilution, lift, and inhibition.
These approaches are very similar, although the factor-based approach explicitly considers inhibition and dilution, processes that negatively affect the development of convection. The primary result is that lift is the most important of the four factors in determining the location of thunderstorm initiation.
CHAPTER 2: SPATIOTEMPORAL DISTRIBUTION OF INITIATION POINTS

2.1. Introduction

One of the first national thunderstorm climatologies to go beyond counting the number of days with thunder was done by Changnon (1988). That study found the maximum in thunderstorm events over the lower Mississippi valley during the cool season, with increasing thunderstorm count and a northwestward movement of the maximum until it resides in the central High Plains in summer. Carbone et al. (2002) found that during the warm season convection frequently initiates over the high, sloping terrain of the Rockies in the early afternoon and shifts eastward during the evening and overnight hours. This study mainly focused on continental-scale precipitation features, and only during the warm season. Wilson and Roberts (2006) studied convection initiation using a wealth of storm-scale data collected on 44 days during IHOP_2002. They found two peaks in convection initiation, a surface-based peak during the afternoon hours and an elevated peak during the overnight hours. Tucker and Li (2009) analyzed an 11-year sample of warm-season precipitation within the Arkansas-Red River basin using hourly radar-based precipitation data, and showed that storms were most frequent in August and during the afternoon and evening, consistent with activity that is mostly driven by diurnal heating. Advantages of the dataset used in this study over the others are:

- features are identified and tracked at the storm scale and 5-minute granularity;
- data covers all seasons (rather than just the warm season) over multiple years;
- the spatial domain is large.
The high spatiotemporal resolution of the data is necessary to accurately identify and locate the initiation points. In order for the analysis to be complete, it needs to cover a sizable domain and incorporate data from all seasons over multiple years. While the studies mentioned above have had some of these characteristics, none have simultaneously possessed all of them.

This work serves as the first component of a broader examination of observed deep convection. The objective of this component is to identify the spatiotemporal distribution of thunderstorm initiation points over the Great Plains from 2005-2007. Focus will be placed on identifying seasonal, diurnal and spatial patterns and the interannual variability of these patterns. A longer record of data is required to establish robust interannual trends. Nevertheless, the results presented here can establish a benchmark for future work.

2.2. Methodology

2.2.1. Radar-based thunderstorm identification

Accurately identifying thunderstorm initiation points requires first identifying and tracking individual thunderstorms, since it is the initiation of individual thunderstorms that is of interest. Level-II radar data were downloaded from the NCDC archive (NCDC 2011b) for 2005-2007 for 44 radars covering the Great Plains (Fig. 2.1). Identification of thunderstorms from these data was done using the Thunderstorm Observation by Radar (ThOR; Lahowetz et al. 2010) algorithm, which involves the following key steps: 1) remove non-meteorological echoes using a neural network quality control algorithm (the w2qcnn algorithm of Lakshmanan et al. 2007a); 2) merge the data from individual radars into a common three-dimensional grid (the w2merger algorithm of Lakshmanan et al.
2006); 3) attenuate stratiform precipitation using fuzzy logic (the stratFilter algorithm described in Appendix A); 4) identify candidate thunderstorms through image segmentation of radar reflectivity to form reflectivity clusters (the w2segmotionll algorithm described by Lakshmanan et al. 2009); 5) track these clusters over time; 6) associate lightning to clusters along the tracks to classify tracks as thunderstorms. The w2qcnn, w2merger, and w2segmotionll algorithms are included in the Warning Decision Support Services – Integrated Information (WDSS-II; Lakshamanan et al. 2007b) package.

The horizontal extent of the grid used by w2merger is shown by the black box in Fig. 2.1. The grid spacing was 0.014° latitude x 0.011° longitude, approximately 1 km x 1 km.

Fig. 2.1: Map showing radars used in the study (labeled black dots), the area within 300 km of those radars (stippled), and the analysis domain (black box).
The w2segmotionll algorithm operated on the composite reflectivity fields (the maximum reflectivity within each vertical column) that were generated by w2merger at 5-minute granularity and subsequently modified by stratFilter. The reflectivity clusters developed by the w2segmotionll algorithm are constrained to have a composite reflectivity value between 30 and 70 dBZ and a minimum area of 50 km².

The algorithm to create tracks from the reflectivity clusters starts by identifying a cluster that has not been placed on a track. The 0-6 km mean wind from the North American Regional Reanalysis (NARR; Mesinger et al. 2006; NCDC 2010) is used as the initial motion estimate for the first 10 minutes of each candidate track. As the track gets longer, a reliable motion estimate can be obtained from the position history of the storm. Between 10 and 30 minutes the motion estimate is the weighted average of the NARR and position history estimates, and after 30 minutes, the motion estimate is entirely based on the position history. Moving downstream through the observed clusters at subsequent times, the tracking algorithm creates all unique candidate tracks that begin at the given cluster. The candidate track with the lowest mean error (difference between actual position and projected position) over the duration of that candidate track is chosen as the correct track, provided that track contains at least two clusters. For details regarding the procedures used to train and verify ThOR’s tracking algorithm, refer to Appendix B.

Cloud-to-ground lightning data are used to determine if particular tracks generated by ThOR are thunderstorms; only those tracks that have at least one cluster at the same time and location as a strike are counted as thunderstorms (cluster positions and
shapes are interpolated to account for the lightning data being at one-minute granularity and the clusters being at five-minute granularity). Some thunderstorm tracks (and possible initiation points) are omitted since only cloud-to-ground lightning data are used to identify thunderstorms and not all thunderstorms produce cloud-to-ground lightning. It is unknown how many tracks are omitted for this reason, though it is reasonable to think that the majority of thunderstorms in this domain produce at least one cloud-to-ground strike.

2.2.2. Selection of Initiation Points

The times and locations of thunderstorm initiation are determined from the final thunderstorm tracks output by ThOR. Since the first cluster on a track represents the time at which sufficient reflectivity was observed by radar, it is unlikely that this represents the actual time and location of initiation. The initiation point to be used is found by extrapolating the storm’s track backward 15 minutes from the appearance of the first cluster.

The primary interest of this study is the “first initiation” within an area, rather than initiation of new cells within an area where thunderstorms were already present, such as a pre-existing multicell system. The concept of excluding initiations located near pre-existing storms was used by Wilson and Roberts (2006) when assessing the initiation mechanisms for convection observed during IHOP. To ensure the initiation points represent “first initiation”, the candidate initiation points are checked to see if they are within 100 km of “established” storms at the time of initiation. “Established” storms are defined as thunderstorms that are at least 15 minutes old (30 minutes old when considering the backward extrapolation described above). If the initiation point for a
track is found to be within 100 km of an established storm at initiation time, it is considered to be connected to the ongoing convection, so all points along that track are considered established. Candidate initiation points beyond the threshold distance from established storms are retained. The backward extrapolation makes it possible for initiation points to be identified that are outside the analysis domain. This is most likely to occur if the storm initiates upstream and moves into the domain. Since these are not true initiations but artifacts of the boundary, all initiation points west of 105°W (the western boundary of the domain) were removed.

2.3. Results

Kernel density analysis was applied to the retained initiation points using ArcGIS, using a Gaussian kernel with a bandwidth of 75 km, and a cell size in the output raster of 10 km (Fig. 2.2). The bandwidth was chosen subjectively after testing several values since it represented the best balance between smoothing out noise and preserving important patterns. The cell size was chosen in a similar way.

Fig. 2.2: Kernel density of initiation points for (a) 2005, (b) 2006, and (c) 2007. Values are points per km$^2$. 
One of the obvious results is that initiation points cluster along the high terrain features found in the western part of the domain, from the northern Sierra Madre Oriental near the Rio Grande through New Mexico to the Colorado Rockies, with another isolated cluster near the Black Hills in South Dakota (Fig. 2.3). This is expected, as the focus of this study is on first initiation of deep convection, and many warm season convective episodes originate over the high terrain around 105°W (Carbone et al. 2002). The primary reason for convection to initiate here is heating of elevated terrain (Tucker and Crook 2005), which implies these regions will be most active when heating is maximized, i.e., summer afternoons. Beyond the topography, the other consistent trend is that there are more initiation points in the southeastern part of the domain, which is also not surprising as suitable thermodynamic profiles are found there during a greater portion of the year.

Fig. 2.3: Elevation map for study area.
Binning the initiation points by month shows that the maximum occurs in the summer, as expected (Fig. 2.4). Both 2005 and 2006 show a sharp peak in August. The maximum in 2007 is in June, though counts are similar in June, July, and August. The August peak is consistent with the results from Tucker and Li (2009). The geographic distributions (Fig. 2.5) indicate that April through September are active in virtually all areas of the domain and contain the vast majority of the mountain-related initiations.

The most striking feature of the seasonal distribution data is the much lower number of initiation points seen in the 2007 warm season compared to the other years. These months are largely responsible for the reduced density of initiation points for 2007 as a whole (Fig. 2.2).

Fig. 2.4: Distribution of initiation points by month for each year.
Analysis of the diurnal trends in initiation points (Fig. 2.6) shows a similar and expected shape for each year, with a primary peak during the afternoon. There are subtle indications of the 08 UTC peak described by Wilson and Roberts (2006). The reduction in initiation points in 2007 is spread over the entire diurnal cycle. The peak time for 2006 is 20 UTC, though in the other years the maximum is at 18 UTC.
2.4. Discussion

The high concentration of initiation points in mountainous areas near the edge of the analysis domain inevitably brings up two major questions: 1) Are these actual thunderstorms or merely ground clutter returns from the mountains themselves? 2) Are the tracks represented by those initiation points originating there or initiating upstream and then moving into the domain?

It is unlikely that a meaningful fraction of the initiation points result from contamination by returns from the mountains themselves. Radar returns from the mountains themselves are automatically removed prior to the data being archived at NCDC. Any remaining clutter should be removed by the neural network quality control algorithm. Any clutter that is left behind would need to have reflectivity over 30 dBZ covering an area of at least 50 km² to be counted as a cluster by w2segmotionl1. Furthermore, only the tracks associated with cloud-to-ground lightning are used, so lightning would have to occur near the ground clutter for it to contaminate the final
dataset. With these criteria in place, it is safe to say that these are actual thunderstorms, not simply radar returns from the mountains themselves.

The issue of artificial enhancement of initiation point density along the western boundary of the analysis domain was dealt with by removing points that likely represent storms moving into the domain rather than initiating in it. The high concentration of initiation points near the edge of the domain suggests a need to expand the domain westward in areas where radar coverage permits. The steps described in section 2.2 were re-run for a supplemental domain extending from 33°N-43°N and 102°W-108°W using the four radars with significant coverage beyond 105°W (CYS, FTG, PUX, FDX). The resulting initiation points west of 104°W were used to replace the original initiation points west of 104°W. Since the tracks away from the boundary are unlikely to change, all existing initiation points east of 104°W were left unaltered. Application of the same kernel density analysis to the combined domain (Fig. 2.7) shows that some of the initiation point maxima shift west to better align with the location of mountains. This supports the original conclusion that mountainous areas are a focus for initiation. Beam blockage by the mountains may be contributing to the apparent lack of initiation points west of 106°W in Colorado.
Fig. 2.7: Kernel density of initiation points for the combined domain. The expansion from the original domain is shown by the additional black box.

The overall distribution of thunderstorm initiation is consistent with other studies (e.g., Changnon 1988; Tucker and Li 2009), initiation is most frequent during summer afternoons, and is often focused by terrain. The presence of interannual variation is expected as well, but the 30-40% drop in initiation point counts in 2007 compared to the other years is unexpected.

Based on the lack of initiation points, one might expect that 2007 was a dry year in the Great Plains. However, annual precipitation data obtained from PRISM (Daly et al. 1994; PRISM 2012) and plotted in Fig. 2.8 indicates the opposite. Dong et al. (2011) identified hydrological year 2006 (October 2005 – September 2006) as one of the driest on record in Oklahoma, and hydrological year 2007 (October 2006 – September 2007) as one of the wettest.
Fig. 2.8: Annual precipitation anomalies (% of normal) for 2005-2007 (a-c, respectively) based on PRISM data.

Analysis of track and initiation point counts for each year (Table 2.1) shows that even though there were slightly fewer tracks in 2007 those tracks contained almost twice as many clusters on average. Since the clusters are regularly spaced in time, more clusters per track means that the tracks in 2007 also lasted significantly longer. The increased number of clusters in 2007 (an increase of 81% from 2005 and 57% from 2006) is consistent with above normal precipitation.

The number of initiation points depends both on the number of tracks and how many of the candidate initiation points (one per track) are retained. As described in section 2.2, only those points away from established storms are retained. If convection is dominated by large complexes, it is less likely that candidate initiation points will be retained, as new tracks are likely to be located near one or more established tracks. The
ratio of retained initiation points to tracks is slightly lower in 2007, which, along with the longer tracks, suggests storms organized into larger convective systems rather than isolated cells.

Table 2.1: Summary of initiation points and tracks for each year.

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td># Initiation Points</td>
<td>37,596</td>
<td>33,001</td>
<td>19,282</td>
</tr>
<tr>
<td># Tracks</td>
<td>194,141</td>
<td>218,310</td>
<td>185,699</td>
</tr>
<tr>
<td># Clusters</td>
<td>971,654</td>
<td>1,122,997</td>
<td>1,761,985</td>
</tr>
<tr>
<td>Mean Clusters per Track</td>
<td>5.00</td>
<td>5.14</td>
<td>9.49</td>
</tr>
<tr>
<td>% Initiation Pts Retained</td>
<td>19.36</td>
<td>15.12</td>
<td>10.38</td>
</tr>
</tbody>
</table>

Tucker and Li (2009) found that 70-80% of the storms in the Arkansas-Red River basin (a subset of the domain used in this work) were small, short-lived, ordinary cells, though these storms produced only about 1% of the total precipitation. The vast majority (86%) of the precipitation came from the 1% of storms they identified as MCSs (duration of at least 6 hr, and an axis of at least 100 km). The number of initiation points scales with the number of tracks (modified by the degree of convective organization, as noted above), which is dominated by isolated storms. However, the amount of precipitation is dominated by the frequency and location of large convective systems that produce precipitation more efficiently over a larger area. The most likely explanation for the unusual characteristics of the initiation point numbers in 2007 (more precipitation, fewer initiation points, longer tracks) is an unusually large frequency of MCSs relative to isolated storms. Another possibility for such a pattern could be tropical cyclones and remnants thereof, which could also produce very heavy precipitation, long tracks and
relatively few “first initiations”. An example of this is the precipitation track of Tropical Storm Erin during August 2007 (Fig. 2.9d), which was characterized by precipitation significantly higher than surrounding areas, and a similar density of initiation points.

To analyze the spatial relationship between initiation points and precipitation, four months from the data set have been selected, and are shown in Fig. 2.9. June 2005 and August 2006 are chosen to represent “typical” warm-season months. In these cases, there is reasonably good spatial agreement between higher (lower) precipitation and greater (lower) density of initiation points, except in the higher terrain where deep, dry subcloud layers result in many initiation points and little precipitation. In particular, August 2006 features a band of high precipitation from New Mexico through the Texas Panhandle and into Kansas that coincides with a very high concentration of initiation points.
Fig. 2.9: PRISM monthly precipitation (shading) and initiation points (dots) for (a) June 2005, (b) August 2006, (c) June 2007, and (d) August 2007. Blue curve in (d) approximates the track of TS Erin and its remnants.

June and August are also chosen to represent the 2007 warm season. Both months featured localized very heavy precipitation. Goebbert et al. (2008) observed that a large
MCS developed on 19 June and generated a mesoscale convective vortex (MCV) that aided in the development of convection over the region for a couple of days. A second MCV that developed on 24 June developed into a deep warm-core circulation, and this feature remained nearly stationary until the end of the month and was largely responsible for the bullseye of precipitation in southeast Kansas and adjacent areas (Goebbert et al. 2008; Schumacher and Davis 2009). Though there are fewer initiation points overall, their concentration is greatest in the typical mountain locations and near the area of heaviest precipitation.

Other than along the path of Erin, most of the rain in August fell in the upper Midwest. There is a region of enhanced initiation point density over eastern South Dakota and southwestern Minnesota, on the northwest edge of the heavy precipitation area (Fig. 2.9d). This pattern suggests upscale growth of these “first initiations into one or more MCSs that moved east or southeast, which would be consistent with the observed tendency for long tracks.

While 100 km sounds like a large threshold to use for determining if an initiation point is “near” an established storm, it was not determined arbitrarily. The results of threshold distances of 50 km, 75 km, and 100 km were manually examined and the locations of retained initiation points compared to corresponding radar images to determine if they truly represented “first initiation”. It was found that the smaller thresholds were unable to adequately filter out initiation points near and within large convective systems. One possible reason for this is that it is the locations used to represent established storms are the centroids of the tracked clusters, and the clusters themselves can be quite large if there is a large area of contiguous high reflectivity, as
might occur in a squall line. While a new storm developing on the periphery of the system may be more than 50 or 75 km from the nearest centroid, it can still be directly connected to the pre-existing convective system. This occurred frequently enough that the larger threshold distance was required. To test the sensitivity of the results to this threshold distance, the same analysis was conducted using a distance of 50 km. The most significant change was a 150-200% in the number of initiation points. The annual spatial distributions (Fig. 2.10) show increased density everywhere (note the change in the color scale). The largest increases occur between 90°W and 100°W south of 45°N, from east Texas and Louisiana up to Nebraska (compare Fig. 2.10 to Fig. 2.2). Local maxima in initiation point density are still seen in the same mountainous regions noted before, but they are somewhat less pronounced, except near the Big Bend region of Texas.

Fig. 2.10: Kernel density of initiation points using a 50 km threshold distance to established storms. Values are in points per km$^2$ (note change in color scale from Fig. 2.2).
Analysis of the annual (Fig. 2.11) and diurnal (Fig. 2.12) distributions shows the same increase in initiation points for all months and hours. The relative minimum in initiation points in July 2005 is likely due to lightning data being unavailable from July 1-7 of that year. The basic pattern seen previously still exists, with a sharp peak in August for both 2005 and 2006 and a broad May-August peak in 2007 with considerably lower counts. The peak in the diurnal distribution has shifted a bit later into the afternoon, occurring at 20 or 21Z, though the curves look similar to those in Fig. 2.4 otherwise.

Fig. 2.11: Annual distribution of initiation points using a 50-km threshold.
2.5 Conclusions

One purpose of this work is to examine the spatiotemporal distribution of thunderstorm initiation points over the Great Plains and the interannual variability of those distributions as part of a broader examination of thunderstorm initiation in the region. A second purpose was to establish a benchmark for “typical” distributions of thunderstorm initiation over the Great Plains. The primary conclusions of this portion of the study are that: 1) initiation is favored over mountainous areas in the domain during the warm season; 2) there were significantly fewer initiation points in the 2007 warm season than in other years; 3) the amount of precipitation in a region is not necessarily well-correlated to the number of initiation points as precipitation also depends on the longevity and organization of the convection.
CHAPTER 3: ENVIRONMENTAL CHARACTERISTICS OF THUNDERSTORM INITIATION

3.1. Introduction

Fundamentally, deep convection initiation (DCI) requires that a volume of air be lifted to a level where it is able to realize considerable positive buoyancy over a significant depth. Positive area on a thermodynamic diagram for some lifted parcel is a necessary condition; however, the effects of dilution on the buoyancy that an actual updraft is able to realize cannot be neglected (Houston and Niyogi 2007; hereafter HN07). Also, the amount of lift that is needed depends on the amount of inhibition present below the level of free convection (LFC). These ideas are captured by the three “ingredients” of Johns and Doswell (1992) – instability, moisture, and lift. Consideration of the processes that govern convection suggests a slight modification to that approach in which DCI is examined in the context of two pairs of factors – buoyancy and dilution, and lift and inhibition. The “moist layer of sufficient depth” (Johns and Doswell 1992) has two roles. The first is to produce a parcel with sufficient $\theta_e$ to achieve positive buoyancy given the temperature profile and the assumptions of parcel theory. The second role (and primary reason the depth of the moisture is important) is to limit the dilution of the parcel as it ascends. Therefore, we contend that it is better to consider buoyancy and dilution as the governing factors. Furthermore, HN07 showed that a positive feedback exists between dilution and buoyancy. Lift and inhibition are paired since the amount of inhibition is what determines if a given amount of lift is sufficient to initiate a thunderstorm.
Buoyancy and inhibition are frequently assessed using parameters based on parcel theory, in particular convective available potential energy (CAPE; Moncrieff and Miller 1976) and convective inhibition (CIN). However, the collocation of significant CAPE and minimal CIN does not guarantee that deep convection will develop, even when a lifting mechanism is present (Ziegler and Rasmussen 1998; hereafter ZR98).

Lift of sufficient strength and depth to get the parcel to its LFC is assumed in the calculation of parameters based on parcel theory, including CAPE and CIN. Vertical motion is a quantity that is difficult to accurately diagnose in the atmosphere, due largely to sparse and flawed observational data. It has been known for quite some time that convergence lines are favored locations for convective initiation (Purdom 1982; Wilson and Schreiber 1986). Horizontal mass convergence at low levels creates pressure excesses near the ground and an associated upward pressure gradient force. This pressure gradient force will result in upward acceleration of air parcels that could be sufficient to get them to their LFC. Accordingly, low-level convergence is often used as a measure of lift in forecasting DCI, and airmass boundaries are favored locations for this to occur.

The explicit exclusion of parcel dilution is one of the major limitations of traditional parcel theory. Dilution occurs when a rising parcel entrains environmental air with lower $\theta_e$, which acts to reduce the amount of buoyancy the parcel can realize. Entrainment of environmental air can affect the parcel in two ways. One is the reduction of parcel $\theta_e$ by mixing it with environmental air with lower $\theta_e$. This process is relevant to both saturated and unsaturated parcels. Another process that affects saturated parcels is evaporation brought about by mixing with dry environmental air, which acts to cool the parcel. The theory of criticality proposed by HN07 is an effort to include the feedback
between buoyancy and parcel dilution in the process of convection initiation. In their numerical experiments, deep convection only occurred if the rate at which parcels could gain buoyancy through ascent exceeded the rate at which buoyancy was lost through dilution. The presence or absence of deep convection was found to be related to the lapse rate of the active cloud-bearing layer (ACBL), which is the layer above the LFC where “active” convection is occurring (Stull 1985). Though dilution is a cumulus-scale process that cannot be directly measured or computed from the data available, environmental parameters relevant to dilution (humidity, ACBL lapse rate and wind shear) can be measured, and that is the intent here.

The purpose of this work is to determine how often each of the basic factors (buoyancy, dilution, lift, and inhibition) is the difference between thunderstorms initiating and thunderstorms not initiating. Even though the reasoning outlined above applies to deep convection in general, the data in this study are generated from a subset of deep convection that produced cloud-to-ground lightning. The most accurate description of this dataset is thunderstorms that produce cloud-to-ground lightning; however, in the interest of brevity, these will be described as “thunderstorms”. This determination requires quantifying the factors at locations where initiation occurred as well as other locations where initiation did not occur as a point of comparison. These locations need to be related enough to make meaningful pairwise comparisons. To quantify the factors, a number of parameters will be computed from RUC-II model analysis data (Benjamin et al. 2004; NCDC 2011a), with the intent that they be independent of geography and/or season as much as possible. To be clear, the use of parameters is not intended as a search for an as-yet-undiscovered “magic bullet” to forecast thunderstorm initiation. Rather, the
relative importance of a parameter is used to indicate the importance of the factor it is measuring. Since multiple parameters may be used to measure the same basic factor, some insight can be gained on the most effective ways of quantifying each factor.

3.2. Description of Parameters

Though an essentially infinite range of parameters could be computed from RUC-II data, the parameters chosen for this study are intended to represent physical processes occurring in the environment that would affect the development of convection. The parameters to be computed from the RUC-II analysis data are shown in Table 3.1. The descriptions and justifications for the parameters used in this work follow.

Table 3.1: Parameters to be computed from RUC-II analysis data, listed by basic factors the parameters are designed to quantify.

<table>
<thead>
<tr>
<th>Lift and Inhibition</th>
<th>Buoyancy and Dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIN</td>
<td>CAPE</td>
</tr>
<tr>
<td>Maximum omega (magnitude)</td>
<td>ACBL lapse rate</td>
</tr>
<tr>
<td>Ht of maximum omega</td>
<td>LCL-LCL+2km CAPE</td>
</tr>
<tr>
<td>H_{LFC}</td>
<td>0 to top of ACBL mixing ratio difference</td>
</tr>
<tr>
<td>Convergence</td>
<td>ACBL wind shear</td>
</tr>
<tr>
<td>Subcloud wind shear</td>
<td></td>
</tr>
<tr>
<td>Δz*</td>
<td></td>
</tr>
</tbody>
</table>

**CIN.** One of the most commonly used metrics to forecast initiation is CIN, as it quantifies how much lift must be provided for a parcel to reach its LFC. There are three main “parcels” that are commonly used to compute CIN (and CAPE) – the surface parcel, the mixed-layer parcel, and the most unstable parcel. Here all three will be used with the purpose of comparing the outputs to evaluate the assumptions made about the properties
of the parcels responsible for initiating thunderstorms. The surface-based method assumes the parcels that initiate convection mainly originate near the surface. This assumption is clearly inadequate for cases of elevated convection, where convection occurs above a low-level inversion. The most-unstable parcel method assumes that the parcels with the highest $\theta_e$ are most relevant. In many cases, the surface-based and most-unstable methods are equivalent since the surface parcel is the one with the highest $\theta_e$. The mixed-layer method uses a “parcel” with the mean mixing ratio and potential temperature of the lowest 100 mb or lowest 1 km of the atmosphere (here the lowest 100 mb layer is used). This method is an attempt to account for mixing within the boundary layer, and it generally yields more conservative values of CAPE and CIN than the other methods. The virtual temperature correction (Doswell and Rasmussen 1994) is used in all parcel-based calculations. It is hypothesized in this work that there will be significantly less CIN in cases with storms, though there will also be null cases with minimal CIN, and that a combination of CIN with information about the lift present will be more useful.

**Maximum omega and height of max omega.** Lift is one of the important factors for the formation of thunderstorms. In the past, estimates of vertical motion were not available at the spatial and temporal resolution needed for forecasting thunderstorm initiation, so vertical motion was assumed or inferred from other fields. With the advent of the RUC and other similar models, estimates of vertical motion are available at 20-km grid spacing and hourly resolution. While this grid spacing is unable to resolve meso-$\gamma$ scale updrafts that directly initiate thunderstorms, it is worth testing the ability of this parameter to discriminate environments that do or do not initiate thunderstorms. Specifically, both the magnitude of maximum upward motion no higher than 100 mb
above the LFC and the height of that maximum value will be obtained from the RUC-II data, as both the strength and depth of lift may be relevant to initiation. It is hypothesized that both will be higher where initiation occurred. Additionally, the height of maximum upward motion is needed to compute $H_{\text{LFC}}$, which is described next.

$H_{\text{LFC}}$. As described by ZR98, $H_{\text{LFC}}$ is the ratio of the height of maximum upward motion (assumed to represent the top of the mesoscale updraft) and the height of the LFC. Values significantly greater than 1 suggest that initiation will occur since parcels are being lifted above their LFC. This ratio combines information about the depth of lift present and information about the depth of lift needed, so it is hypothesized that it will be quite effective in assessing the potential for thunderstorm initiation. However, the effectiveness of the metric is dependent on the quality of the vertical motion information available from the RUC.

Convergence. Many studies (e.g. Wilson et al. 1992; Xue and Martin 2006) have related DCI to areas of enhanced convergence. The depth of convergence has also been shown to be important (Wilson et al. 1992; ZR98), thus, the convergence over a deeper layer (such as parcel level to LFC) may be more useful than surface convergence alone. The computation of both surface and 0-LFC mean convergence will allow this hypothesis to be tested. The 0-LFC mean convergence is the integral of convergence from the parcel level to the LFC for the most unstable parcel, divided by the distance between the two levels. The continuity equation relates vertical motion to the divergence, and vertical motion at a given level is the integral of the divergence in the level below it. Though moisture flux convergence is frequently used in the forecasting of severe storms, Banacos and Schultz (2005) suggest that simple mass convergence provides essentially the same
information and is more physically sound. However, as noted by Doswell and Schultz (2006), divergence can be a rather noisy and volatile field due to sparse and potentially erroneous observations.

**Subcloud wind shear.** RKW theory (Rotunno et al. 1988) states that when the horizontal vorticity associated with the cold pool is equal and opposite the environmental vorticity strong vertical updrafts are created along the gust front. Though originally conceived to explain squall line maintenance, the basic physical principles can still be applied for thunderstorm initiation (Lee et al. 1991). Since we are interested in “first initiation”, there should be no “cold pools”, though there could be airmass boundaries, which have horizontal vorticity associated with them. Accordingly, the idea of environmental vorticity due to vertical wind shear balancing the vorticity associated with the airmass boundary to create enhanced updrafts along the boundary seems plausible. This idea was lent some credence by Lee et al. (1991), who evaluated a case of thunderstorm initiation along colliding boundaries and found that the removal of low-level shear in a model simulation diminished the convection. Here low-level shear is defined as shear between the parcel level and the LCL in order to represent subcloud shear and to be able to account for elevated parcels. It is hypothesized that increased subcloud shear is favorable for initiation.

**Δz*. First introduced by HN07, this measures how large a vertical displacement is needed for a parcel to reach its LFC. This is useful since it is related to the depth of lift needed to initiate a thunderstorm instead of just the strength of the lift. It is hypothesized that smaller values of Δz* will be found in cases with storms.
CAPE. The existence of positive CAPE is a necessary, but not sufficient, condition for thunderstorms to occur. In theory, more CAPE would produce a stronger updraft given that the parcel is able to reach the LFC. Similar reasoning as for CIN applies to the use of surface-based, mixed-layer, and most-unstable parcels to compute CAPE. It is hypothesized that there will be a significant overlap in the distributions of CAPE in the storm and no storm categories, making CAPE of little use in discriminating initiation and non-initiation environments.

ACBL lapse rate. This parameter was shown to be important in the success or failure of DCI by HN07 (here the ACBL is defined as a 1.5 km deep layer starting at the LFC). As concluded by HN07, larger lapse rates increase the vertical displacement of parcels caused by an airmass boundary due to reduced static stability. Also, steeper lapse rates above the LFC allow parcels ascending through the layer to gain buoyancy more rapidly. Parcels for which the gain of buoyancy through ascent exceeds the loss of buoyancy through entrainment are termed “supercritical” by HN07. The hypothesis of this work is that lapse rates will be larger in environments that initiate thunderstorms.

LCL-LCL+2km CAPE. This has been developed specifically for this work and is defined analogously to CAPE and CIN in height coordinates, except the limits of integration are the LCL and 2 km above the LCL. The depth of the layer was chosen to represent the region immediately above the cloud base where the feedback between buoyancy and dilution is most important. Since the sign of the virtual temperature difference may be either positive or negative in this layer, both positive and negative values are meaningful. Positive values indicate there is more CAPE than CIN in the layer, while negative values indicate the opposite. The purpose of this metric is to quantify how
quickly a parcel can gain buoyancy. It is hypothesized that the distribution of CAPE within the sounding is important, with more CAPE in the low levels of the sounding being more supportive of initiation since parcels are more able to overcome the negative effects of dilution. Though this effect is also measured by the ACBL lapse rate, a CAPE framework takes parcel moisture into account.

**0 to top of ACBL mixing ratio difference.** This is another parameter developed specifically for this work. It is designed to represent the cumulative potential entrainment a rising parcel might experience as it ascends to a level where it is significantly buoyant. The relevant quantity to measure the dryness of the environment relative to the parcel is the difference between parcel mixing ratio and environmental mixing ratio. By integrating this quantity from the parcel level to the top of the ACBL, the overall dryness of the environment during the critical early stages of convective development can be characterized. Though entrainment is known to be an important process in the evolution of convection, it is neglected by parcel theory and not very well understood. Ziegler et al. (1997) showed that in mesoscale updrafts along a dryline where thunderstorms develop the change in mixing ratio from the surface to the LFC is minimal, in contrast to nearby areas where they do not develop. This deepening of the moist layer is a result of persistent convergence and upward motion. It is hypothesized that if rising parcels must pass through deep dry layers before significant buoyancy is achieved then the likelihood of thunderstorm initiation will be reduced.

**ACBL wind shear.** While vertical wind shear is known to help in storm organization and severity, a number of studies have suggested that vertical shear above the boundary layer has a negative effect on storm initiation. Weisman and Klemp (1982) and Lee et al.
(1991) showed that increased vertical shear tended to decrease the maximum updraft speed of the convection and delay its onset. Possible mechanisms by which increased shear above the LFC can inhibit convection are increased entrainment and the advection of developing clouds away from the boundary layer updraft. Entrainment is a turbulent mixing process, and greater wind shear leads to greater turbulence, so greater wind shear may lead to increased entrainment. Wind shear also tilts the updraft, increasing the surface area subject to entrainment. For a mesoscale updraft to initiate a thunderstorm, rising parcels need to reach their LFC before being advected out of the updraft region (ZR98). Peckham and Wicker (2000) showed that stronger cross-dryline flow in the 0-5 km layer inhibited the growth of deep convection along the dryline since clouds were more rapidly advected away from the surface-based convergence band associated with the dryline. If sufficient dynamic perturbation pressure gradients were not yet established, then the incipient storms fell apart. It is hypothesized that there will be less wind shear in the ACBL in the cases of deep convection. The notion of wind shear having opposite effects depending on the layer of the shear is consistent with the results of Lee et al. (1991). For both this parameter and the subcloud wind shear parameter, the orientation of the shear with respect to a possible mesoscale boundary is relevant in addition to the magnitude of the shear. A limitation of the present analysis is that it is unable to consider the orientation of boundaries and only considers the magnitude of the shear.
3.3. Methodology

3.3.1. Selection of Points for Parameter Collection

3.3.1.1. Initiation Points

The initiation points found using the steps described in section 2.2 are grouped into hourly bins centered at the nominal RUC-II analysis times. The center of the bin is defined as $t_0$. Initiation points within $\Delta ii$ (Table 3.2) of each other are clustered into a single representative point. The method for determining which points should be grouped together is an adaptation of connected component analysis from graph theory (two points are considered “adjacent” if they are within $\Delta ii$ of each other). Within each group, the mean center of the group is defined as the mean latitude and mean longitude of all candidate initiation points in the group. The candidate initiation point nearest to this mean center is cataloged as the representative for this group (Fig. 3.1c). Candidate initiation points beyond $\Delta ii$ from all other initiation points (“isolated” initiations) are cataloged.

Table 3.2: Description of the distance thresholds and the values used in this chapter.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Description</th>
<th>Value (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta ii$</td>
<td>Distance used to determine clustering of candidate initiation points</td>
<td>50</td>
</tr>
<tr>
<td>$\Delta nt$</td>
<td>Threshold distance between candidate null point and any thunderstorm location</td>
<td>40</td>
</tr>
<tr>
<td>$\Delta ni$</td>
<td>Distance between initiation point and candidate null point</td>
<td>60, 120, 180</td>
</tr>
</tbody>
</table>

The motivation for this spatial grouping is to avoid biasing the final results by having many samples from the same location and time. This study is interested in
whether a given environment produces deep convection, so whether one storm or five
occur in that environment should not matter, and the sampling approach should reflect
that. Similar reasoning was used by Thompson et al. (2003), who used time and space
separation thresholds for their supercell climatology to avoid biasing their results to
single events with a large number of supercells.

3.3.1.2. Null Points

In order to be most useful, the null cases chosen in this study need to represent an
environment which is close to initiating convection, and maybe is missing just one
ingredient. The strategy adopted in this study for selecting points to represent the null
case environments takes points that are a distance Δni from the cataloged initiation points
(Fig 3.1d). Only the candidate null points beyond a threshold distance Δnt from all
thunderstorm locations within the hourly bin and Δni from all candidate initiation points
within a 3-hour bin centered at t0 are cataloged (Fig. 3.1d). This is to ensure that the
selected null point is actually away from areas of convection. The reason a 3-hour buffer
is used for initiation points is to avoid a situation where a sounding could be used to
represent a “null point” in one bin and then be used as a t-1 initiation sounding for an
initiation in the next hourly bin.

There is some uncertainty concerning an appropriate value for Δni, as the most
informative null points are those that are “close” to initiating convection and do not. As a
result of the 20-km grid spacing and the method of selecting the model gridpoint used to
compute the parameters, the minimum Δni to assure that under no circumstance could the
initiation point and the null point use the same grid point is 60 km, so this was chosen as
the minimum value of $\Delta n_i$ (and the value that is used in Fig. 3.1). The other values of $\Delta n_i$ are simply twice and triple this value.
Fig. 3.1: (a) Identification of initiation points by backward extrapolation of tracks. (b) Removal of initiation points within 100 km of established storms. The procedures illustrated in (a) and (b) are described in section 2.2. (c) Spatial clustering of nearby initiation points. (d) Selection of locations to be used for null points. (e) Removal of points with no adjacent RUC grid points with positive MUCAPE.

For “isolated” initiation points, candidate null points are identified at a distance $\Delta n_i$ from the initiation point in the 8 cardinal directions (Fig. 3.1d). For “grouped” initiation points, the shape of the group is approximated by a rectangle. The length of the rectangle is equal to the maximum distance between candidate initiation points in the group. The width of the rectangle is specified as twice the maximum distance from a candidate initiation point to the line connecting the maximally separated points. The actual latitude and longitude differences between those maximally separated points gives the “rotation” of the rectangle. Candidate null points are then identified $\Delta n_i$ from the corners of the rectangle and $\Delta n_i$ from the midpoints of its sides (Fig. 3.1d). As the aspect ratio (length/width) of the rectangle becomes larger, the diagonal search directions compress toward the long axis of the rectangle. This is beneficial since a linear pattern of initiation points is likely along an atmospheric boundary of some kind, so the most useful null points will be those along the boundary that initiated the convection. Having more candidate null points near the boundary makes it more likely that the null points sampled will have environments near the initiation/non-initiation threshold, as these null points will be the most informative.
This approach assumes that the locations selected as null points have environments with many characteristics that support thunderstorm development; however, something is missing since thunderstorms did not develop there. Though this approach will likely reduce separation in distributions of parameter values between the two categories as the null environments are similar to the initiation environments, it should allow for the isolation of the “missing ingredients” in each case. This approach also allows pairwise differences to be used to compare storm and no storm cases, which is a way to eliminate event-to-event variability in the convective environments. This approach will be discussed further in Section 3.4.3.

3.3.2. Attribution of Parameter Values to Cataloged Points

The atmospheric parameters described in section 3.2 will be calculated from hourly RUC-II analysis using an NCAR Command Language (NCL) script. The RUC-II has a grid spacing of 20 km, and produces new analyses and forecasts every hour. As described by Benjamin et al. 2004, analysis fields are developed using the 1-hr forecast from the previous run as the “first guess”. This first guess is then adjusted based on observations ingested from a variety of sources (for more details on the RUC-II data assimilation methods, see Benjamin et al. 2004). This “hot start” approach to developing analysis fields means that vertical motion has been “spun-up” at the analysis time.

The RUC grid point nearest the cataloged initiation or null point with positive most unstable CAPE (MUCAPE) will be used as the data source for that initiation or null point (Fig. 3.1e). The reason for this criterion is that positive MUCAPE is a necessary condition for deep convection to occur, and if the RUC grid point does not satisfy this condition then it is not a representative profile for an initiation point. This criterion is
used for null points since the values of the other parameters are trivial if the necessary condition is not met. If the nearest grid point for a cataloged initiation or null point does not possess positive MUCAPE, the NCL script then checks the next nearest grid point for positive MUCAPE. The process continues until it finds a grid point with positive MUCAPE or it has searched the four nearest grid points, whichever is first. If none of the four nearest grid points meet the criterion, then that initiation or null point will not be used further. If a grid point has positive MUCAPE, then the script will compute the remaining parameters. The virtual temperature correction is used in all parcel-based computations. The parcel ascent is treated as a pseudoadiabatic process.

For the initiation and null points within each hourly bin, two data sets will be collected. The first will be from the analysis valid at the central time of the bin (t₀). These data are intended to represent the environment within 30 minutes either side of the observed initiation. The second data set will be from the previous hour’s analysis (t₋₁) and it is intended to represent the pre-initiation environment, since t₋₁ is at least 30 minutes before the observed initiation time.

3.4. Results

3.4.1 Analysis of Raw Values

The chosen method for analyzing the raw parameter values for initiation and null points is the box-and-whisker plot. For all box-and-whisker plots shown hereafter, the box represents the middle 50% of the data, the black line is the median, and the whiskers extend to the maximum data value within 1.5 times the interquartile range. Outliers beyond this range are not plotted. Null points at each range (60, 120, and 180 km) are shown along with the initiation points. Also, correlation coefficients presented are based
on Spearman rank correlation throughout. The dataset contains 55,103 initiation points that are retained after the procedures described in section 3.2, and 324,000-352,000 null points (depending on range). This translates to an average of 6-7 null points at each range that can be paired with each initiation point, out of the 8 that were originally considered.

The plots of raw parameter values are shown in Fig. 3.2-3.3. One of the first things to notice is that all parameters show significant overlap between initiation and null distributions at all 3 ranges. This indicates that the spread in background environments overwhelms the differences between initiation and null points, and that thresholds that distinguish initiation from non-initiation for all cases do not exist. This is not surprising as the dataset includes both surface-based and elevated convection as well as a wide range of climate zones. This variation in environments should be most relevant to the thermodynamic variables, such as CAPE and Δz*. It is somewhat surprising that a dimensionless variable such as $H_{LFC}$, which is designed to account for variation in environments, would not perform better at this stage. This means that more sophisticated pairwise analysis is required to extract useful information from this data (section 3.4.3).

Another characteristic of most variables is that the separation in medians increases as range increases, suggesting that the favorable convection initiation environment can be more skillfully identified at a precision of 180 or 120 km than 60 km. This is expected, as the scale of features resolvable by a model with 20-km grid spacing is ~80 km.
Fig. 3.2: Box-and-whisker plots of raw CAPE and CIN values.
Fig. 3.3: Box-and-whisker plots of raw values for remaining variables.
The approach used in this work to assess discriminatory ability from box-and-whisker plots is to evaluate the absolute value of the difference in medians divided by the interquartile range of the initiation point values. This quantity will hereafter be referred to as the “separation”. Applying this technique to the raw data shows that the parameters with the most discriminatory ability are maximum omega, convergence (both surface and 0-LFC), Δz*, and mixed-layer CAPE and CIN. The separation values for these parameters at 120 km are shown in Table 3.3. The 120 km range is being used to represent all ranges since the order is similar at each range, although the values themselves are greater at 180 km and less at 60 km. Shear values seem to make very little difference, as all four boxes are essentially identical for both 0-LCL and ACBL shear.

Table 3.3: Separation values for raw data at each range.

<table>
<thead>
<tr>
<th>Range</th>
<th>Max omega</th>
<th>0-LFC conv</th>
<th>Sfc conv</th>
<th>Δz*</th>
<th>MLCIN</th>
<th>MLCape</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 km</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>120 km</td>
<td>0.16</td>
<td>0.14</td>
<td>0.14</td>
<td>0.12</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>180 km</td>
<td>0.21</td>
<td>0.18</td>
<td>0.17</td>
<td>0.17</td>
<td>0.12</td>
<td>0.11</td>
</tr>
</tbody>
</table>

3.4.2. Surface-based vs. Elevated Convection

Some of the parameters collected are unlikely to be relevant to elevated convection, specifically surface-based and mixed-layer CAPE/CIN and surface convergence, so it makes sense to separately examine surface-based and elevated convection. Also, knowing how much of the dataset they make up will help explain why
some parameters are more or less relevant in the whole dataset. The official definition of elevated convection is convection that does not ingest near-surface air (Glickman 2000); however, this cannot be directly determined from the data available. For this work, we define elevated storm environments as those with zero SBCAPE, and surface-based environments as those in which SBCAPE and MUCAPE are equal. This classification leaves a gray area in which MUCAPE is greater than SBCAPE, and both are non-zero; here it cannot be unambiguously determined which parcels are contributing to the convection, so they are not included in this analysis. Null points satisfying the criteria for either category are selected for that category whether or not the matching initiation point also falls into the category. In this dataset, approximately 60% were classified as surface-based, approximately 15% were classified as elevated, and the remaining 25% were indeterminate.

For surface-based cases, mixed-layer CAPE/CIN offers more discrimination than most unstable CAPE/CIN (Fig. 3.4). Compared to the plots for the entire dataset (shown in gray in the figures), CAPE is larger and CIN is closer to zero in the surface-based cases. It is worth noting that 53-55% of the null points had zero MUCIN, yet failed to initiate convection. Craven et al. (2002) noted that observed cloud base heights are better predicted by the mixed-layer LCL than the surface LCL, so it is reasonable that the mixed-layer parcel would be more useful in predicting initiation as well. It may be that the reason that the mixed-layer parcel is more useful is that it makes an attempt to account for parcel dilution.
Fig. 3.4: Most unstable (left) and mixed-layer (right) CAPE and CIN for surface-based cases. Offset gray box-and-whisker plots are identical to those in Fig. 3.2.

Other parameters that look to offer some discrimination for surface-based cases are shown in Fig. 3.5, and those are convergence (both surface and 0-LFC), maximum omega, and Δz*. Separation values for the surface-based cases at the 120 km range (Table 3.4) show that mixed-layer CAPE and CIN and both convergence parameters offer improved discrimination for surface-based cases than for the entire dataset (Table 3.3). The separations for omega and Δz* are about the same as for the whole dataset. The correlation between the two convergence variables is not particularly strong, 0.31, and both are negatively correlated with maximum omega, about -0.5 for both.
Table 3.4: Separation values for raw surface-based data at 120 km range, listed in order of separation.

<table>
<thead>
<tr>
<th>0-LFC conv</th>
<th>Sfc conv</th>
<th>Max omega</th>
<th>MLCIN</th>
<th>Δz*</th>
<th>MLCAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.21</td>
<td>0.19</td>
<td>0.17</td>
<td>0.15</td>
<td>0.12</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Fig. 3.5: Convergence (top), maximum omega, and Δz* for surface-based cases. Offset gray box-and-whisker plots are identical to those in Fig. 3.3.

By definition, surface-based and mixed-layer CAPE and CIN are of no use for elevated cases, leaving the most unstable parcel as the only useful choice for any parcel-based properties. Examination of the parameters found useful for surface-based parcels and plotted in Fig. 3.5 shows that convergence is not especially important for elevated
cases; however, maximum omega and Δz* are still relevant (Fig. 3.6). Separation values for elevated cases at the 120 km range (Table 3.5) show that the set of parameters that discriminate for elevated storms is considerably different than the set of parameters that discriminates well for surface-based storms.

Table 3.5: Separation values for raw elevated data at 120 km range, listed in order of separation.

<table>
<thead>
<tr>
<th>Δz*</th>
<th>MUCIN</th>
<th>LCLCAPE</th>
<th>Max omega</th>
<th>MRD</th>
<th>MUCAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.18</td>
<td>0.17</td>
<td>0.15</td>
<td>0.14</td>
<td>0.13</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Fig. 3.6: Convergence (top), maximum omega, and Δz* for elevated cases. Offset gray box-and-whisker plots are identical to those in Fig. 3.3.
However, maximum omega and Δz* are relevant for both, which suggests that they have a robust relationship with the initiation of thunderstorms. The fact that omega values are similar for elevated cases despite reduced mean convergence and reduced correlation between convergence and maximum omega (-0.17) suggests that the circulation containing the upward motion may be sloped, as suggested by Banacos and Schultz (2005). This would result in much of the convergence occurring upstream of the initiation point, both horizontally and vertically, and not being captured at the gridpoint used to represent the initiation at or above the height of the most unstable parcel. However, the upward motion is clearly being captured, and that is really the variable that matters.

The mixing ratio deficit (MRD) parameter shows some interesting behavior when the cases are divided into surface-based and elevated categories. The absolute values in the elevated cases are much lower, and confined to a narrow range (Fig. 3.7).

![Fig. 3.7: Comparison of MRD for surface-based (left) and elevated (right).](image)

### 3.4.3. Pairwise differences

The goal of this work is to find the factor that is different between initiation and null points, so it is natural to consider the pairwise difference between an initiation point
and the null points associated with it. All initiation and null points that do not have a “mate” with valid data are discarded at this stage. The same initiation point could be paired with as many as eight null points and differences were computed for each resulting pair. Absolute differences are transformed to a common scale to facilitate comparison of the discriminatory ability of each parameter. This transformation accounts for the relative magnitude of values being compared (a difference in CIN of 10 J kg\(^{-1}\) is much more significant when the raw values are -5 and -15 than when they are -90 and -100) and accounts for the possibility of one or both parameters being zero or negative. For this transformation the difference for initiation points is defined as

\[
I = \frac{\text{init} - \text{null}}{\max(\text{abs}(\text{init}), \text{abs}(\text{null}))}
\]

and for null points it is defined as

\[
N = \frac{\text{null} - \text{init}}{\max(\text{abs}(\text{init}), \text{abs}(\text{null}))}.
\]

Variables that are negative by convention, specifically CIN and omega, will have positive normalized differences when initiation values are less negative (smaller in magnitude) than null values. Variables that are always of one sign will have values of \(I\) and \(N\) that are always between -1 and 1. For variables that have meaningful values of either sign, \(I\) and \(N\) are bound by -2 and 2. For both types of variables, the distributions of \(I\) and \(N\) are symmetric about 0 with respect to each other. The above formulation is still vulnerable to both initiation and null values being zero. In these situations the normalized difference is set to 0.
Because not all the null points could be paired with an initiation point (due to the MUCAPE criterion), the number of pairs considered in this step of the analysis is a bit less than the number of null points at each range. The number of pairs ranges from 315,000 at 60 km to 340,000 at 180 km. The box-and-whisker plots of $I$ and $N$ for each parameter are shown in Figs. 3.8-3.9, where the number following $I$ or $N$ corresponds to the range (60, 120, or 180 km).

![Box-and-whisker plots](image)

Fig. 3.8: Normalized pairwise difference distributions of CAPE and CIN at 60, 120, and 180 km.
Fig. 3.9: Normalized pairwise difference distributions at 60, 120, and 180 km for remaining parameters.
Like the raw values, the discriminatory ability of each parameter increases with greater separation between initiation and null points. Also, the best-performing parameters still appear to be maximum omega, Δz*, convergence, MLCAPE, and MLCIN. One attribute of the convergence distributions that is worth noting is that even though the medians are separated by a considerable amount, the area of overlap between the boxes is quite large. Plotting a histogram of the values of $I$ for 0-LFC convergence at 120 km shows that the distribution is bimodal (Fig. 3.10). This bimodality is common to both convergence parameters at all distances. The larger peak is the positive one, and more than half of the values are positive, which is why the median is significantly positive. However, there are a lot of negative values as well, which is why the third quartile is not higher. The bimodality means that convergence values at initiation and null points are not very well correlated, suggesting a field that is noisy rather than smoothly varying. This enhances the possibility of getting unrepresentative raw or normalized difference values for convergence if the RUC places a convergent boundary incorrectly, even if the error is fairly slight.
The ultimate goal of this work is to provide insight on which of the relevant factors for convection is most often the one missing from cases of initiation failure. In order to provide this insight, the frequency of normalized differences that are “significantly” positive or negative for each parameter are cataloged in order to see how many of the initiation/null pairs could be correctly identified by various combinations of parameters. To “correctly identify” the pair of points is to be able to find parameters with significant differences of a sign consistent with initiation at the initiation point rather than the null point. Differences were determined to be “significant” if they were outside the overlapping part of the boxes on the box-and-whisker plots. Mathematically, the significance threshold is $T = \min\{q_{3,\text{init}}, q_{3,\text{null}}\}$, where $q_{3,\text{init}}$ and $q_{3,\text{null}}$ are the upper
quartile for the initiation and null distributions, respectively. This threshold is determined for each variable at each range. If the absolute value of a normalized difference is greater than \( T \) for that variable, then the sign of the normalized difference is cataloged. Variables with better separation in the distributions will have a greater number of significant differences, and a greater percentage of them should be of the same sign.

The first step in finding the parameters offering the best discrimination of initiation and null environments is to rank the parameters based on the percentage of significant differences that are of the same sign using the entire dataset. This ranking is shown in Tables 3.6-3.8.
Table 3.6: Ranking of parameters by significant differences over the entire dataset at the 60 km range.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>% of significant differences of same sign</th>
<th>Abs(# Pos. differences – # Neg. differences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum omega</td>
<td>56.15</td>
<td>22133</td>
</tr>
<tr>
<td>MUCAPE</td>
<td>55.94</td>
<td>21280</td>
</tr>
<tr>
<td>MLCIN</td>
<td>55.71</td>
<td>14843</td>
</tr>
<tr>
<td>MLCAPE</td>
<td>55.59</td>
<td>19848</td>
</tr>
<tr>
<td>LCLCAPE</td>
<td>55.46</td>
<td>19342</td>
</tr>
<tr>
<td>Sfc convergence</td>
<td>55.21</td>
<td>18377</td>
</tr>
<tr>
<td>0-LFC convergence</td>
<td>55.05</td>
<td>17735</td>
</tr>
<tr>
<td>Δz*</td>
<td>54.92</td>
<td>17221</td>
</tr>
<tr>
<td>SBCAPE</td>
<td>54.67</td>
<td>16276</td>
</tr>
<tr>
<td>MUCIN</td>
<td>53.05</td>
<td>10259</td>
</tr>
<tr>
<td>MRD</td>
<td>52.54</td>
<td>8437</td>
</tr>
<tr>
<td>H_{LFC}</td>
<td>52.49</td>
<td>8263</td>
</tr>
<tr>
<td>ACBL lapse rate</td>
<td>52.11</td>
<td>6915</td>
</tr>
<tr>
<td>SBCIN</td>
<td>51.40</td>
<td>3814</td>
</tr>
<tr>
<td>Subcloud shear</td>
<td>50.71</td>
<td>2273</td>
</tr>
<tr>
<td>Ht of max omega</td>
<td>50.42</td>
<td>1346</td>
</tr>
<tr>
<td>ACBL shear</td>
<td>50.36</td>
<td>1152</td>
</tr>
</tbody>
</table>
Table 3.7: Ranking of parameters by significant differences over the entire dataset at the 120 km range.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>% of significant differences of same sign</th>
<th>Abs(# Pos. differences – # Neg. differences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum omega</td>
<td>59.56</td>
<td>39885</td>
</tr>
<tr>
<td>MUCAPE</td>
<td>59.01</td>
<td>37154</td>
</tr>
<tr>
<td>MLCAPE</td>
<td>58.90</td>
<td>36587</td>
</tr>
<tr>
<td>LCLCAPE</td>
<td>58.89</td>
<td>36547</td>
</tr>
<tr>
<td>MLCIN</td>
<td>58.66</td>
<td>24809</td>
</tr>
<tr>
<td>0-LFC convergence</td>
<td>58.41</td>
<td>34202</td>
</tr>
<tr>
<td>Δz*</td>
<td>57.95</td>
<td>31294</td>
</tr>
<tr>
<td>Sfc convergence</td>
<td>57.56</td>
<td>30013</td>
</tr>
<tr>
<td>SBCAPE</td>
<td>57.03</td>
<td>27649</td>
</tr>
<tr>
<td>H_{LFC}</td>
<td>54.84</td>
<td>18106</td>
</tr>
<tr>
<td>MUCIN</td>
<td>54.75</td>
<td>17737</td>
</tr>
<tr>
<td>MRD</td>
<td>54.53</td>
<td>16830</td>
</tr>
<tr>
<td>ACBL lapse rate</td>
<td>53.68</td>
<td>13394</td>
</tr>
<tr>
<td>SBCIN</td>
<td>51.52</td>
<td>4320</td>
</tr>
<tr>
<td>Ht of max omega</td>
<td>51.48</td>
<td>5167</td>
</tr>
<tr>
<td>Subcloud shear</td>
<td>51.10</td>
<td>3815</td>
</tr>
<tr>
<td>ACBL shear</td>
<td>50.03</td>
<td>116</td>
</tr>
</tbody>
</table>
Table 3.8: Ranking of parameters by significant differences over the entire dataset at the 180 km range.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>% of significant differences of same sign</th>
<th>Abs(# Pos. differences – # Neg. differences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum omega</td>
<td>61.70</td>
<td>51919</td>
</tr>
<tr>
<td>LCLCAPE</td>
<td>61.00</td>
<td>47985</td>
</tr>
<tr>
<td>MUCAPE</td>
<td>60.83</td>
<td>47005</td>
</tr>
<tr>
<td>MLCAPE</td>
<td>60.66</td>
<td>46049</td>
</tr>
<tr>
<td>0-LFC convergence</td>
<td>60.04</td>
<td>42746</td>
</tr>
<tr>
<td>Δz*</td>
<td>59.79</td>
<td>41422</td>
</tr>
<tr>
<td>MLCIN</td>
<td>59.72</td>
<td>27933</td>
</tr>
<tr>
<td>SBCAPE</td>
<td>58.60</td>
<td>35345</td>
</tr>
<tr>
<td>Sfc convergence</td>
<td>58.46</td>
<td>34621</td>
</tr>
<tr>
<td>H_{LFC}</td>
<td>56.52</td>
<td>25468</td>
</tr>
<tr>
<td>MUCIN</td>
<td>55.91</td>
<td>22781</td>
</tr>
<tr>
<td>MRD</td>
<td>55.89</td>
<td>22713</td>
</tr>
<tr>
<td>ACBL lapse rate</td>
<td>54.62</td>
<td>17253</td>
</tr>
<tr>
<td>Ht of max omega</td>
<td>52.22</td>
<td>7907</td>
</tr>
<tr>
<td>SBCIN</td>
<td>51.10</td>
<td>3088</td>
</tr>
<tr>
<td>Subcloud shear</td>
<td>50.83</td>
<td>2855</td>
</tr>
<tr>
<td>ACBL shear</td>
<td>50.25</td>
<td>849</td>
</tr>
</tbody>
</table>
It is likely that several of the parameters in the above tables are closely related to other parameters near them in the list. For example, all the CAPE and CIN values should be correlated, since they are all derived from the same basic procedure applied to the same sounding. Likewise, omega and convergence (especially 0-LFC mean convergence) are dynamically linked via the continuity equation, so when one is favorable the other should be as well – they are largely redundant. Also, since mixing ratio deficit is an integrated quantity, and the depth of integration depends on \( \Delta z^* \), they are related. And of course \( H_{\text{LFC}} \) is derived from the height of maximum omega and \( \Delta z^* \), so they will be related.

Some correlations noted in the data that may be a little less intuitive are ACBL lapse rate with \( \Delta z^* \) (0.49) and 0-LCL shear with CIN (-0.42). Both correlations are likely related to the presence of inversions. The \( \Delta z^* \)-ACBL lapse rate connection can be reasoned by considering the typical “loaded gun” sounding with a strong elevated mixed layer (EML). If the LFC is within the EML, the low-level parcel must be forced through a significant depth to reach the LFC. When it does, the environmental lapse rate in the layer above the LFC is very large. Conversely, if the low-level parcel has high enough \( \theta_e \) to be just warmer than the nose of the inversion, its LFC will be much lower, and the lapse rate in the layer above the LFC will also be much lower.

The relationship between CIN and 0-LCL shear is less intuitive, since one is a thermodynamic variable and one is a kinematic variable that is normalized to account for differences in LCL height. The theory proposed for this relationship is that the base of an inversion is typically a layer with enhanced vertical shear since static stability inhibits mixing of momentum across the layer. If the parcel level is below the inversion and the
LCL is above the base of the inversion, the 0-LCL layer will have increased wind shear. The inversion also explains the higher CIN.

To account for the interrelatedness of the parameters, the procedure for determining which ones are most useful is to start the list with the one that is most effective for all the data, in this case maximum omega. Then the same procedure that was applied to get Tables 3.6-3.8 is applied to cases in which maximum omega to find the next important parameter. This is done recursively until no more useful parameters are left. When one parameter is controlled for, other parameters strongly linked with it should be largely controlled for as well. This is more effective than simply comparing the effectiveness of the parameters based on all the data, as it is difficult to say for certain which one of a set of related parameters is really best.

Implementation of this procedure at each range determined that the three best parameters are maximum omega, MLCAPE, and Δz*. After those were accounted for, the next best parameter was 0-LFC convergence, which is directly related to omega, which we have already used. This suggests that all the truly important factors have been accounted for.

With the list of parameters narrowed to three, it is possible to evaluate how often each one is the one that is different when the other two are neutral. These results are presented both as conditional probabilities and total number of occurrences, as the number of cases with the other two parameters both neutral was not necessarily the same (Table 3.9). Specifically, the number of cases with two of the three parameters neutral decreased with range. Since lower values of omega and Δz* favor initiation, the conditional probability evaluated is the probability of a significant negative difference
given that the other two parameters are neutral, whereas for MLCAPE it is the probability of a significant positive difference given the same condition. From these calculations, maximum omega was the most significant parameter. The conditional probabilities of a significant difference in the “correct” direction are slightly higher at the 180 km range, and slightly lower at the 60 km range, though the relative importance of each factor is the same. This suggests that it is simply the model detecting fewer differences between points that are closer together, which is unsurprising.

Table 3.9: Conditional probabilities of “correct” significant differences (count of such occurrences in parentheses) at each range.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>60 km</th>
<th>120 km</th>
<th>180 km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max omega</td>
<td>30.46 (19645)</td>
<td>35.57 (20409)</td>
<td>39.9 (20463)</td>
</tr>
<tr>
<td>MLCAPE</td>
<td>28.41 (17587)</td>
<td>32.23 (17049)</td>
<td>34.62 (15554)</td>
</tr>
<tr>
<td>Δz*</td>
<td>28.51 (17559)</td>
<td>31.85 (16576)</td>
<td>34.14 (14869)</td>
</tr>
</tbody>
</table>

Parameter values were also collected at initiation and null points an hour before the time of initiation. Examination of these values showed that the only measurable difference between the two times (determined by using pairwise differences between the \( t_0 \) and \( t_1 \) data) is a slight increase in upward motion and convergence at the initiation points.

3.5. Discussion

One of the interesting differences between surface-based and elevated initiations is the MRD parameter. In addition to the changes described in section 3.4.2., there are
also changes in the correlations between MRD and other parameters that may offer insight. In surface-based cases, it is most correlated with Δz* (0.60) and MUCAPE (0.40), though in elevated cases it is most correlated with Δz* (0.69) and MUCIN (-0.62). In surface-based cases, the best correlation with CIN was only -0.23 for MLCIN, and the correlation to MUCAPE in elevated cases was only 0.25. The common thread is Δz*, which makes sense as MRD is an integrated quantity, and the upper limit of integration depends on the LFC, so a deeper layer would increase the total deficit. The correlation with CAPE can best be explained by considering a case where both the lifted parcel and the environment are saturated above the LFC. In this case, the only factor controlling the integrand in the MRD formulation above the LFC is the temperature excess of the parcel, and greater buoyancy will produce a greater difference. The lack of discrimination in surface-based cases may be a result of increasing CAPE favoring initiation, while increasing Δz* inhibiting it. For elevated cases, MUCIN and Δz* are highly correlated (-0.85) and the relationship between MUCIN and MRD likely works through changing the depth of integration, measured by Δz*. In this case, less negative CIN and lower Δz* both favor initiation, resulting a more consistent trend in MRD.

There are two possible reasons for a high Δz*, either there is a deep well-mixed layer with little CAPE or CIN (typical of continental convection), or there is a significant depth of inhibition between the initial parcel level and the LFC, either above or below the LCL. The former case would lead to high Δz* with low CIN values and low correlation between CIN and Δz*. In the second case, Δz* and CIN should be strongly correlated. It is certainly possible to have both a high LCL and significant inhibition. Grant (1995) found typical 850 hPa dewpoint depressions in cases of elevated severe thunderstorms to
be about 2°C, suggesting that the first situation is rarely applicable to elevated storms. We know in cases with CIN the environment is not saturated for two reasons: 1) if it was, it would have a higher $\theta_e$, thus becoming the “most unstable” parcel; 2) the environment mixing ratio would be higher than the parcel mixing ratio from parcel level to LFC, which would tend toward positive correlation between MUCIN and MRD, though a negative correlation is observed. This suggests that above the most unstable parcel the environment becomes drier and also potentially warmer, which suggests that potential instability may play some role in the development of elevated convection when CIN is present. In fact, soundings associated with elevated storms throughout the literature (Grant 1995; Banacos and Schultz 2005; Holgan et al. 2007) show both nearly saturated conditions at the level of the most unstable parcel and potential instability above this level. These results appear to indicate that when elevated convection is possible, the most useful set of features to look for in combination is relatively low MUCIN, potential instability, and upward vertical motion to enable the release of the potential instability.

The results also indicate that 0-LFC mean convergence is a more useful predictor of initiation than simply surface convergence, as expected. While the magnitude of maximum omega was the most effective measure of lift, 0-LFC convergence is second best. It has more significant correlation with omega, as one would expect from the continuity equation. While surface convergence is certainly not applicable to elevated convection, neither convergence variable offers much insight there, likely since the circulation is sloped and not well represented at a single model gridpoint.

The parameters consistently identified as most important in this work are maximum omega, $\Delta z^*$, and CAPE. All of the CAPE parameters are highly correlated,
and when one of them is selected as the best parameter, the others are usually not far behind, so whether MUCAPE or MLCAPE is selected as “best” may not be tremendously significant. Both of them are primarily measures of buoyancy, though MLCAPE also accounts for dilution, which MUCAPE does not. Since SBCAPE is either equivalent or inferior to MUCAPE (depending whether the convection is surface-based or elevated), and it lacks the ability of MLCAPE to account for dilution, there is not a situation in which SBCAPE is the best one to use for evaluating the potential for thunderstorm initiation.

Relating these parameters back to the four basic factors indicates that lift is the most important single factor in determining where thunderstorms will initiate. Physically, it makes sense that lift would be most important factor. In addition to its role as a trigger for thunderstorm initiation, upward motion also serves a role in preconditioning the atmosphere. Persistent updrafts along a boundary act to deepen the moist boundary layer, making it more suitable for subsequent updrafts to reach their LFC (Ziegler et al. 1997).

Since both lift and inhibition were measured, it is apparent from these results that initiation failures due to “insufficient” lift are more commonly due to differences in lift rather than differences in inhibition. Also, buoyancy appears to be a more important factor than dilution, although both MLCAPE and Δz* have a relationship to dilution. The criterion that all points used in the analysis have positive MUCAPE already controlled for buoyancy to a limited degree. While this may result in a slight underestimation of the importance of buoyancy, cases where thunderstorms failed to initiate due to an absence of buoyancy are trivial.
Since deep convection is parameterized in the model used to compute the environmental parameters, there may be some concern about the values for maximum omega in this study actually representing the effects of the convective parameterization being triggered at the gridpoint chosen to represent initiation (it is an initiation point after all). It appears unlikely that this is a significant influence on the results for a number of reasons. First, the vertical domain for finding the maximum omega extends to at most 100 mb above the LFC, so the highest midlevel updraft values should not be sampled. Also, if the convective parameterization was triggered, omega values should increase significantly above the LFC and the maximum omega value should occur at the upper bound of this domain. This should result in significant differences in the height of maximum omega compared to null points where the convective parameterization is not active. It should also result in $H_{LFC}$ values significantly above 1. However, neither of these is true about the dataset in general, as height of maximum omega shows essentially no difference between initiation and null points, and the median $H_{LFC}$ value is very near 1. Additionally, only a very slight increase in the omega value is noted in the hour prior to initiation, which is not consistent with a change in the status of the convective parameterization. Though it is impossible to rule out the possibility of the convective parameterization coming into play in a few cases, it seems reasonable to conclude that the differences in omega are legitimate variations in the environment rather than artifacts of the model convective parameterization.

The interrelatedness of the parameters used here shows that the initiation of deep convection is a complex process that cannot easily be boiled down to a single number. A value of a certain parameter, such as omega, that might support initiation in one
environment may not in another, since the distance to the LFC or some other parameter is different. Also, a wide range of thermodynamic environments are capable of initiating deep convection, which makes the creation of parameters that apply to all thunderstorms difficult. Defining parameters that use layers dependent on the parcel LCL or LFC to make them adaptable to different thermodynamic environments ends up introducing unintended interdependencies, and a parameter intended to measure one thing becomes dependent on something else. For example, the mixing ratio deficit and ACBL lapse rate are mainly intended to measure dilution or the feedback between buoyancy and dilution, though both of them are sensitive to $\Delta z^*$, since the location of the LFC determines what layer is considered in evaluating them. When using parameters defined in such a way, it is important to keep these interdependencies in mind.

The goal of this study was to identify which parameters (and ultimately processes) are most important to the initiation of thunderstorms in real-world environments. The relevance to forecasting is helping distinguish the parameters that are useful from the ones that are not, as well as identifying relationships between different parameters that may be relevant to their interpretation. It is beyond the scope of this work to determine the best algorithm to forecast thunderstorm initiation using these parameters. However, some suggestions can be offered. The way to translate the pairwise difference approach to something computable on a grid is most likely through the evaluation of differences from a neighborhood mean. This study can offer some guidance on an appropriate radius over which to compute the spatial mean. These values could be used as “interest fields” or as input to a number of decision-making techniques, including decision trees and logistic regression.
3.6. Conclusions

The goal of this work is to determine which of the four basic factors (lift, inhibition, buoyancy, and dilution) that regulate convection is most often responsible for initiation or non-initiation. This is done using parameters derived from 20-km RUC-II analysis near a large sample of thunderstorm initiation points and null points. Analysis of the raw parameter values showed that a wide range of environments are capable of initiating convection and as a result there are no “magic numbers” that effectively discriminate initiation and null points. Neither subcloud shear nor ACBL shear show any meaningful differences between initiation and null environments. However, the orientation of the shear with respect to a possible initiating boundary was not considered, and this is likely important for the shear in both layers.

Separately examining surface-based and elevated initiations shows that maximum omega and Δz* are useful discriminators for both types. Convergence is useful for surface-based cases, but not for elevated cases. In surface-based cases the mixed-layer parcel provides the most useful CAPE and CIN values, but for elevated cases only the most unstable parcel is applicable.

Analysis of pairwise differences between initiation and null points shows that lift is the most important single factor, and that the maximum upward motion in the column (no higher than 100 mb above the LFC) is the most effective way to quantify lift. The two next most effective parameters are MLCAPE (measuring buoyancy) and Δz* (measuring inhibition). Beyond these three parameters, there is not much additional information to be gained in remaining parameters. Even though none of those parameters primarily related to dilution, that does not mean that it can be ignored. The primary value of MLCAPE
over MUCAPE or SBCAPE is that it attempts to include the effects of dilution, and the distance a parcel has to travel ($\Delta z^*$) influences the amount of dilution a parcel can experience below the LFC. Greater lift facilitates initiation through preconditioning as well as triggering, so it makes sense that lift would be the most significant factor and dilution would be of secondary importance.
CHAPTER 4: SUMMARY OF KEY FINDINGS

In this study, thunderstorms were identified and tracked during the years 2005-2007 over the central United States. The purpose of this study was to examine thunderstorm initiation through two primary avenues. The first was analysis of the spatiotemporal distribution of thunderstorm initiation points and the interannual changes in that distribution. The primary findings of this are: 1) initiation is favored over mountainous areas in the domain during the warm season; 2) there were significantly fewer initiation points in the 2007 warm season than in other years; and 3) the amount of precipitation in a region is not necessarily well-correlated to the number of initiation points as precipitation also depends on the longevity and organization of the convection.

The second focus of the analysis was to compare the atmospheric environment in which thunderstorms initiate to environments seemingly conducive for convection that do not initiate thunderstorms. This comparison focused on the four basic factors that govern convection (buoyancy, dilution, lift, and inhibition), quantifying them with a set of parameters computed from 20-km RUC-II analysis data. There is a wide range of environments found to support thunderstorms, and none of the parameters considered provided a threshold value that could consistently discriminate initiation and null points. Analysis of pairwise differences between initiation and null points showed that the magnitude of maximum upward motion, MLCAPE, and Δz* represent the best set of parameters to discriminate initiation and null points. Of these, maximum omega is the most effective, which indicates that lift is most frequently the factor that is the difference between initiation and non-initiation.
References


APPENDIX A: STRATFILTER ALGORITHM

The motivation for identifying and attenuating stratiform precipitation is that w2segmotionll frequently detects large stratiform areas, and occasionally lumps nearby convection into them, which is undesirable for our purposes. By reducing the reflectivity values in stratiform regions below the threshold used by w2segmotionll (30 dBZ in our case), these areas should no longer be detected. The approach taken to make the convective-stratiform distinction is similar in principle to the approach used by Biggerstaff and Listemaa (2000).

The fundamental difference between convective and stratiform precipitation is the magnitude of the vertical velocity, and this can be inferred from the three-dimensional reflectivity structure. In convective precipitation, strong localized updrafts create a column of high reflectivity, characterized by large horizontal gradients and small vertical gradients. In stratiform precipitation, vertical velocities are weaker and more uniform, resulting in a large shield of precipitation with similar intensity. In such cases, horizontal gradients are generally small and vertical gradients are large since the maximum reflectivity is closely tied to the melting level. In some cases it was noted that horizontal gradients in the composite reflectivity can be significant near the edge of stratiform regions. The reflectivity 3 km above the level of the maximum is very uniform in stratiform precipitation, though in convective precipitation the individual updrafts still lead to large horizontal gradients.

The algorithm to distinguish stratiform and convective precipitation makes use of four pieces of information at each pixel: 1) the merged composite reflectivity (MCR) value of the pixel, 2) the horizontal gradient of the MCR, 3) the vertical reflectivity
gradient, and 4) the horizontal reflectivity gradient computed 3 km above the level of maximum reflectivity at that pixel. The horizontal gradients are determined by applying a sixth-order centered-difference method to the east-west and north-south directions. The length of the vector made by these components is the horizontal gradient at that pixel. The vertical gradient is simply the difference between the maximum reflectivity in the column and the reflectivity 3 km above the level of the maximum, divided by the distance between them. To account for storm tilt, a 9x9 pixel neighborhood around the pixel in question is searched to obtain the value 3 km above the level of the maximum, consistent with the method used by Zipser and Lutz (1994).

Each pixel in the MCR above 25 dBZ is given a “final” score between zero and one where zero indicates definitely stratiform and one indicates definitely convective. Pixels less than 25 dBZ are not considered. First, if a pixel in the MCR has a reflectivity of 50 dBZ or greater or a horizontal gradient within the MCR of 6 dBZ/km or greater, the pixel is given a “final” score of one (convective). If a pixel has a vertical gradient of 5 dBZ/km or greater, however, the pixel is given a “final” score value of zero (stratiform). Any pixels that don’t fit these criteria are given a final score comprised of weighted scores for the above three gradient criteria where

\[
\text{final score} = 0.2\cdot \text{Hscore} + 0.4\cdot \text{H}_{\text{max+3} \text{score}} + 0.4\cdot \text{Vscore}
\]

and each score corresponds to the three gradients above, respectively. The Hscore for a pixel is set to zero if the horizontal gradient in the MCR is less than 1 dBZ/km, one if the horizontal gradient is greater than 3 dBZ/km, and linearly interpolated between zero and one between gradient values of 1 and 3 dBZ/km. The \(\text{H}_{\text{max+3} \text{score}}\) is set in the same fashion for the horizontal gradient at three kilometers above the maximum reflectivity.
level. The Vscore is set to zero if the vertical gradient between the height of the maximum reflectivity and three kilometers above this level is greater than 4 dBZ/km, one if the vertical gradient is less than 1 dBZ/km, and linearly interpolated between zero and one between gradient values of 4 and 1 dBZ/km (Fig. A.1). The final score for each pixel of the MCR is then smoothed (to account for issues with the radar’s sampling).

Finally, the values in the MCR are altered using a fuzzy logic approach (Fig. A.2). Pixels with a final score less than 0.25 are classified as “definitely stratiform” and assigned the “stratiform value”, which is either 20 dBZ or the reflectivity 2 km above the level of maximum reflectivity, whichever is less. Pixels with a final score greater than 0.55 are classified as “definitely convective” and retain their original value (the convective value). Remaining pixels are assigned a reflectivity that is a weighted average of its convective value and its stratiform value, where the weight for each is the truth value for that classification. Reflectivity clusters are then identified from this altered MCR image by the w2segmotionll algorithm (Fig. A.3).
Fig. A.1: Chart depicting the score values for the horizontal (both in the MCR and at 3 km above max) and vertical gradients.

Fig. A.2: Chart depicting the convective and stratiform truth values over the range of final score values.
Fig. A.3: (a) An example merged reflectivity composite mosaic (MCR) valid at 00:01 UTC 11 July 2006. (b) The stratiform-filtered MCR corresponding to (a). (c) Reflectivity clusters identified by w2segmotionll (white regions) overlaid on the filtered MCR in (b).
APPENDIX B: ThOR TRAINING AND VERIFICATION

B.1. Training

Determining appropriate tracking parameters for ThOR and verifying the skill of tracks it produced was a two-part process. All tracks for both training and verification are based on clusters detected by w2segmotionll with a minimum size of 50 km$^2$ and minimum and maximum reflectivity of 30 and 70 dBZ, respectively. Automated algorithms considered only the locations of the centroids of the clusters, while the committee of three meteorologists responsible for creating the human tracks had access to both the centroids and the cluster boundaries, as well as the merged composite reflectivity the clusters were derived from. While there are inherent ambiguities in thunderstorm tracking, particularly concerning instances of storm mergers or splits, the human tracks are designed to represent the “best practices” in tracking, and are used as the reference for “correct” tracks. Since ThOR is allowed to find a continuation of the track from $t_0$ to $t_1$ or from $t_0$ directly to $t_2$ (skipping $t_1$), the human tracks are also allowed to skip one time. More than one time may be skipped within a track, as long as consecutive points on the track are no more than 12 minutes apart.

For the training part, the committee selected a few storms from a variety of events to track. The choice to only track a few features from each event rather than consider all the tracks from the event was made so that more of the storm mode – storm speed space could be sampled while keeping the workload of the committee manageable. The list of cases used for this is in Table B.1.
Table B.1: List of events used for ThOR training. Start and end times are in UTC.

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Region</th>
<th>Type of Storm</th>
</tr>
</thead>
<tbody>
<tr>
<td>20:26 9 Jul 2005</td>
<td>01:40 10 Jul 2005</td>
<td>West TX</td>
<td>Multicell</td>
</tr>
<tr>
<td>18:00 13 Jul 2005</td>
<td>19:43 13 Jul 2005</td>
<td>E OK, W AR</td>
<td>Ordinary cells</td>
</tr>
<tr>
<td>01:00 13 Jan 2006</td>
<td>08:00 13 Jan 2006</td>
<td>AR</td>
<td>Squall line</td>
</tr>
<tr>
<td>00:55 9 Mar 2006</td>
<td>10:05 9 Mar 2006</td>
<td>TX, OK, AR</td>
<td>Supercell + squall line</td>
</tr>
<tr>
<td>17:40 11 Mar 2006</td>
<td>00:00 12 Mar 2006</td>
<td>OK, MO</td>
<td>Supercells</td>
</tr>
<tr>
<td>00:10 13 Mar 2006</td>
<td>07:10 13 Mar 2006</td>
<td>OK, KS, MO</td>
<td>Supercells</td>
</tr>
<tr>
<td>19:35 2 Apr 2006</td>
<td>01:40 3 Apr 2006</td>
<td>N AR, MO, IL</td>
<td>Supercell + squall line</td>
</tr>
<tr>
<td>21:20 12 Apr 2006</td>
<td>00:20 13 Apr 2006</td>
<td>AR</td>
<td>Backbuilding multicell</td>
</tr>
<tr>
<td>18:40 16 May 2006</td>
<td>21:05 16 May 2006</td>
<td>IA, MO</td>
<td>Ordinary cells, weak multicell</td>
</tr>
<tr>
<td>23:05 28 Feb 2007</td>
<td>05:15 1 Mar 2007</td>
<td>KS, MO</td>
<td>Supercell</td>
</tr>
<tr>
<td>19:05 17 Apr 2007</td>
<td>22:35 17 Apr 2007</td>
<td>TX</td>
<td>Squall line</td>
</tr>
<tr>
<td>05:05 24 Apr 2007</td>
<td>08:20 24 Apr 2007</td>
<td>KS, NE, IA</td>
<td>Multicell</td>
</tr>
<tr>
<td>21:05 24 Apr 2007</td>
<td>03:20 25 Apr 2007</td>
<td>S TX</td>
<td>Supercell</td>
</tr>
</tbody>
</table>

There were five parameters we intended to set with the training tracks: *narrbound*, *motiontime*, *R₀*, *a*, and *offset*. Descriptions of these parameters are given in Table B.2. The NARR 0-6 km mean wind is used as the motion estimate for the storm until the track gets long enough that its observed motion can be used as the motion estimate. For a length of time at the beginning of the track set by *narrbound*, only the NARR is used. Between *narrbound* and 30 minutes, the motion estimate is a weighted average of the NARR mean wind and the observed motion. Beyond 30 minutes, the motion estimate is entirely based on the observed motion of the track over a period of time set by *motiontime*. These two parameters were treated as independent variables, as they affect the projected positions, which in turn affect the distance and angle errors. Tested values of *narrbound* were 0 minutes, 10 minutes, and 20 minutes, and tested
values of \textit{motiontime} were 30 minutes, 45 minutes, and 60 minutes. The other three variables were to be derived from the distance and angle errors.

An algorithm very similar to the tracking algorithm of ThOR was constructed to follow the human tracks using each possible combination of \textit{narrbound} and \textit{motiontime} and report the minimum search radius and change in bearing needed to create each segment of the track (or the search radius and angle that would incorrectly extend a track when it should terminate), as well as the $u$ and $v$ motion components used to generate the projected position. The first step in the analysis was to identify the combination that best discriminated the correct and incorrect continuations. Though differences between the various combinations of \textit{narrbound} and \textit{motiontime} were small, \textit{narrbound} was chosen to be 10 minutes, and \textit{motiontime} was chosen to be 45 minutes based on slightly better discrimination of “good” and “bad” points. The complete list of parameter values chosen is in Table B.2.

Table B.2: List of parameters derived from ThOR training.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Narrbound}</td>
<td>Length of time when motion is estimated by NARR only</td>
<td>10 minutes</td>
</tr>
<tr>
<td>\textit{Motiontime}</td>
<td>Amount of position history used to estimate motion beyond \textit{narrbound}</td>
<td>45 minutes</td>
</tr>
<tr>
<td>$R_0$</td>
<td>Search radius at one timestep (5 minutes) ahead</td>
<td>$9.04 \text{ km} + 142.7 \text{ s} * v \text{ (km s}^{-1})$</td>
</tr>
<tr>
<td>$a$</td>
<td>Amount to increase search radius after a skipped time</td>
<td>3 km</td>
</tr>
<tr>
<td>Offset</td>
<td>Amount to widen cone</td>
<td>Cone not used</td>
</tr>
</tbody>
</table>
Having settled on *narrbound* and *motiontime*, the next step was to determine an appropriate search radius. There are three reasons that the actual position could differ from the projected position. The first is the storm deviating from the motion estimate, which could be due to storm propagation (i.e., supercells or squall lines that follow the propagation of the cold pool rather than the mean wind) or NARR mean wind being a poor estimate at the beginning of a track (this is most likely if the storms are not surface-based, thus not being influenced by the wind below the level of their inflow parcels). The magnitude of this error (measured as a distance) increases with time. The second is simply random variation in the location of the centroid around a straight path, which in our case also includes changes in the size and/or shape of the clusters output by w2segmotionll. This type of error should be relatively constant with time. Both of these potential sources of error need to be allowed for in our search radius. The third reason for error in the estimated position is that the cluster in question actually belongs to a different track, and the search radius should not allow for this type of error. This type of error is also assumed to be constant with time. The goal is to pick a search radius at both 5 minutes and 10 minutes (through $R_0$ and $a$) that allows for the first two sources of error and not the third. Analysis of the radius data output by the training program for 5 minute projections indicated a trend for correct continuations to have increasing error as storm speed increased, with little trend in the errors for incorrect continuations, many of which were very large (>30 km). The formula for the search radius was derived using Fisher’s linear discriminant (FLD; Wilks 2011), and the resulting best line gave the search radius as a function of velocity: $R_0(V) = 9.04km + 142.7s * V$, where $V$ is the velocity in km s$^{-1}$. 
The “bad” points with errors greater than 30 km were removed prior to calculating the FLD in order to satisfy the assumption of similar covariance matrices between the two groups. This is justified since there were no “good” points at such a large range, so those points are not helpful in determining the search radius. A scatterplot of the data is shown in Fig. B.1.

Fig. B.1: Search radii for correct continuations (gray +) and incorrect continuations (black squares) and search radius function derived from FLD (black line) based on 5-minute data.
The data for 10 minute projections are only useful when a time needs to be skipped, which is rather infrequent in both human and ThOR tracks. However, it is important that ThOR have a search radius at 10 minutes that is large enough to pick up a correct continuation of the track, and not so large that it jumps to a different object. To train this search radius, every point was projected forward both 5 minutes and 10, mimicking a skip at every possible point along the track. A similar procedure was followed to determine the appropriate search radius. The line returned from FLD had a similar slope to the line for 5-minute data, with values that were slightly larger. Rather than make formulas more complicated than necessary, it was decided to simply add 3 km to the search radius if the time between consecutive clusters is 10 minutes rather than 5 minutes (Fig. B.2). The final formula for search radius is:

$$R(V, dt) = R_0(V) + a^*(dt - 5)/5$$

where $R_0(V)$ is defined as above, $a = 3$ km, and $dt$ is the difference in minutes between the time of the previous cluster and the time of the potential continuation.

It is worth noting that for both 5-minute and 10-minute data, most of the correct continuations have errors of 5 km or less at essentially all velocities. This is the magnitude of error one would expect due to factors such as storm propagation away from the mean wind or slight wobbling of the centroid. The correct continuations with large error are mostly instances where w2segmotionll merged or split clusters, causing the centroids to “jump”. This effect was most significant in larger mesoscale convective systems (MCSs) where there are large contiguous areas of high reflectivity. The search thresholds chosen are dependent on the nature of the objects being tracked. If our
identifications could be free of this “jumping” then we may have selected different search parameters.

Fig. B.2: Search radii for correct continuations (gray +) and incorrect continuations (black squares) based on 10-minute data. Black line is the 5-min FLD function plus 3 km.

Once the search radius was set, the decision regarding the angle of the search cone could be made, since the formulation for the cone was \( \theta = \tan^{-1}\left(\frac{R}{D}\right) \), where \( \theta \) is the half-cone angle, \( R \) is the search radius as defined above, and \( D \) is the distance the cluster
was projected to travel since its last known location. The idea of the cone is that storms should not change their direction of travel very sharply, resulting in a small angle between the line connecting the previous cluster and the projected cluster and the line from the previous cluster to the actual cluster. Clusters that may fall within the search radius and have a large angle are likely incorrect continuations. An exception to this would be slow-moving storms, when the motion is small and possibly erratic. The idea for some kind of directional constraint was proposed by Johnson et al. (1998) as a way to limit incorrect continuations in the Storm Cell Identification and Tracking (SCIT) algorithm. The goal was to see if an offset needed to be added to the formula to make the cone wider. What was found, however, is that very few incorrect continuations were within the search radius and outside the cone, but quite a number of correct continuations fell into that category (most associated with cluster splits/mergers). As a result, the cone was discarded entirely. The lack of incorrect continuations that could be eliminated by the cone only was attributed to the w2segmotionll centroids having lower spatial density than SCIT cells, making the cone less useful.

It should be noted that the chosen tracking parameters are dependent on the objects being tracked, in this case clusters generated by w2segmotionll from composite reflectivity with a particular scale size and dBZ range. Changes to the method of identification would likely change the choice of tracking parameters. The most important attributes of the identifications to the tracking would be the density of centroids (related to the size of the features) and the behavior of the identifications when storms merge or split.
B.2. Verification

Once the tracking parameters were selected, the quality of ThOR tracks could be assessed. A two-part approach was used to obtain a good picture of the overall quality of the ThOR tracks. Johnson et al. (1998) verified the SCIT algorithm by comparing SCIT tracks to human tracks and assessing how many times it made the correct association from one time to the next. There are some flaws with this method, such as lack of specificity, overestimation of skill, and the labor-intensive nature of the manual tracking (Lakshmanan and Smith 2010). Failure to make the correct association could be due to the algorithm failing to make any association at all, choosing the wrong object, or picking up a different object when the original object is no longer detected and the track should end. Lakshmanan and Smith (2010) propose using a set of descriptive statistics from a large sample of tracks to identify these errors and compare different tracking algorithms without requiring human tracks. Specifically, they propose using median duration, the standard deviation of VIL along the track, and the RMSE of the track compared to a linear fit. The latter two may be overly sensitive for short tracks, so they are only averaged over all tracks with duration greater than the median duration. The standard deviation of VIL and deviation from a straight path are used to detect if the algorithm is switching to a different storm during the track, and duration is used to detect if the algorithm regularly misses associations along the tracks. The intent is not to produce specific numbers that correspond to “good” or “bad” performance, nor does it require knowledge of what the “correct” tracks are. Instead the idea is that, in general, a better algorithm should produce longer, straighter tracks, and properties of the storm should be similar along the track.
Our verification used the contingency table approach with a somewhat small dataset of human tracks as well as the aggregate statistics approach on a larger dataset of tracks. For both steps, the ThOR tracks were compared with a benchmark tracking algorithm to provide context for the verification statistics. This benchmark algorithm is essentially a “poor man’s tracking”, simply choosing the cluster nearest to the projected location at each timestep, as long as that nearest cluster is within 12 km of the projected location. It uses the NARR mean wind as the motion estimate for all times, and it does not skip times. For all events used in verification, both ThOR and the benchmark algorithm were run on the same set of clusters.

Table B.3: List of events used for human tracking component of ThOR verification. Start and end times are in UTC.

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Radar</th>
<th>Storm Mode</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00 23 Jun 2003</td>
<td>02:00 23 Jun 2003</td>
<td>KUEX</td>
<td>Supercells</td>
<td>Very slow</td>
</tr>
<tr>
<td>21:30 7 Jul 2005</td>
<td>23:30 7 Jul 2005</td>
<td>KUDX</td>
<td>Multicell</td>
<td>Slow</td>
</tr>
<tr>
<td>20:00 15 Nov 2005</td>
<td>22:00 15 Nov 2005</td>
<td>KNQA</td>
<td>Supercells and line segments</td>
<td>Fast</td>
</tr>
<tr>
<td>02:00 06 Jun 2008</td>
<td>04:00 06 Jun 2008</td>
<td>KTLX</td>
<td>Squall line</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

The four events chosen for manual tracking each consisted of a two-hour window of clusters based on data from a single radar (Table B.3). The events are intended to represent different speeds and storm modes. The size of the human track dataset is limited as manually tracking thunderstorms is labor-intensive and time-consuming, and the workload of the committee needed to be kept manageable. All clusters were used for tracking, although those that could not be matched with another cluster within 10 minutes
either way were discarded, as neither automated algorithm wrote out “tracks” containing only one point.

To assess the accuracy of the tracks, the human tracks were paired with one and only one ThOR track with which it had at least two consecutive matches. In the case of multiple potential matches, the ThOR track with the most points in common was paired with the human track, and the other ThOR tracks remain available for pairing with another human track. Once the pairing is done, each point along all tracks is counted as a hit, miss, or false alarm. For paired tracks, hits occur when the tracks contain the same cluster at the same time, false alarms occur when ThOR has extra points at either end of the track, and misses occur when the human track has extra points at either end of the track, or both exist at the same time and contain different clusters. If both tracks skip at a given time, nothing is counted. All points along a ThOR track that is not paired with a human track are counted as false alarms, and all points along a human track that is not matched with a ThOR track are counted as misses. The full list of possible outcomes is shown in Table B.4. This method is better than a simple percent correct approach since it scores a missed association in the middle of a track differently than an error on the first or last cluster on a track. It also gives some indication as to whether misses or false alarms are the larger source of error.

From the total hits, misses, and false alarms for all events, probability of detection (POD), false alarm rate (FAR), and critical success index (CSI) were computed. In this case, the statistics are computed as:
\[ POD = \frac{\text{hit}}{\text{hit} + \text{miss} + \text{miss}_1 + \text{miss}_2 + \text{miss}_3}; \]

\[ FAR = \frac{\text{falsealarm} + \text{falsealarm}_1 + \text{falsealarm}_2}{\text{hit} + \text{falsealarm} + \text{falsealarm}_1 + \text{falsealarm}_2}; \]

\[ CSI = \frac{\text{hit}}{\text{hit} + \text{miss} + \text{miss}_1 + \text{miss}_2 + \text{miss}_3 + \text{falsealarm} + \text{falsealarm}_1 + \text{falsealarm}_2}. \]

Table B.4: Full list of possible outcomes at each time along a ThOR or human track.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>Clusters match within a track</td>
</tr>
<tr>
<td>Miss</td>
<td>Human exists when ThOR skips</td>
</tr>
<tr>
<td>Miss1</td>
<td>ThOR track starts too late</td>
</tr>
<tr>
<td>Miss2</td>
<td>ThOR track ends too early</td>
</tr>
<tr>
<td>Miss3</td>
<td>Human and thor exist and have different clusters</td>
</tr>
<tr>
<td>FalseAlarm</td>
<td>Thor exists when human skips</td>
</tr>
<tr>
<td>FalseAlarm1</td>
<td>ThOR track starts too early</td>
</tr>
<tr>
<td>FalseAlarm2</td>
<td>ThOR track keeps going too late</td>
</tr>
<tr>
<td>CorrectNeg</td>
<td>Clusters match outside of a track</td>
</tr>
</tbody>
</table>

The benchmark tracks are handled in the same manner as the ThOR tracks, and verification statistics for the two should be directly comparable. The results shown in Table B.5 indicate that ThOR matches the human tracks slightly better than the benchmark tracks.

Table B.5: Track-by-track verification statistics for ThOR and the benchmark.

<table>
<thead>
<tr>
<th>Method</th>
<th>POD</th>
<th>FAR</th>
<th>CSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThOR</td>
<td>0.889</td>
<td>0.108</td>
<td>0.803</td>
</tr>
<tr>
<td>Benchmark</td>
<td>0.849</td>
<td>0.106</td>
<td>0.771</td>
</tr>
</tbody>
</table>
The aggregate statistics suggested by Lakshmanan and Smith (2010) were also computed for all three sets of tracks from these four events. One difference between our statistics and those suggested by Lakshmanan and Smith (2010) is that maximum reflectivity was used instead of VIL in the calculation of standard deviation. In this case, “better” automated tracks are more similar to the human tracks. Results for this sample are shown in Table B.6. Standard deviations are also shown for linearity and mismatch errors to show that the differences between the three algorithms are not significant (p-values from a Student t-test are around 0.3). The median duration of the benchmark tracks is considerably lower than that of the other two. Overall, the aggregate statistics of the ThOR tracks are nearly identical to those of the human tracks, suggesting good agreement.

Table B.6: Aggregate statistics for manually tracked events.

<table>
<thead>
<tr>
<th>Method</th>
<th>Duration</th>
<th>Linearity</th>
<th>Mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (s)</td>
<td>Mean (km)</td>
<td>St. dev. (km)</td>
</tr>
<tr>
<td>Human</td>
<td>1238</td>
<td>2.61</td>
<td>1.62</td>
</tr>
<tr>
<td>ThOR</td>
<td>1240.5</td>
<td>2.59</td>
<td>1.34</td>
</tr>
<tr>
<td>Benchmark</td>
<td>941.5</td>
<td>2.38</td>
<td>1.19</td>
</tr>
</tbody>
</table>

The real value of the aggregate statistics approach is the performance of ThOR can be evaluated on a large sample of tracks that would be prohibitively labor-intensive to manually track. Such a sample was taken from the events in 2005 covering different times of year, different storm modes, and different storm motions. The statistics for these events are shown in Table B.7 and they look very similar to the statistics from the smaller
sample of manually tracked events. This similarity suggests that the result from those events is reliable. The values for median duration and mean linearity error are of comparable magnitude to the values found for better-performing algorithms in Lakshmanan and Smith (2010), which further supports the claim that ThOR does a good job tracking these features.

Table B.7: Aggregate statistics for events with only automated tracks.

<table>
<thead>
<tr>
<th>Method</th>
<th>Duration</th>
<th>Linearity</th>
<th>Mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (s)</td>
<td>Mean (km)</td>
<td>St. dev. (km)</td>
</tr>
<tr>
<td>ThOR</td>
<td>1191</td>
<td>2.59</td>
<td>1.34</td>
</tr>
<tr>
<td>Benchmark</td>
<td>894</td>
<td>2.47</td>
<td>1.35</td>
</tr>
</tbody>
</table>