UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IDENTIFYING PRECURSORS TO STRONG LOW-LEVEL ROTATION WITHIN NUMERICALLY SIMULATED SUPERCELL THUNDERSTORMS: A DATA MINING APPROACH

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN METEOROLOGY

By

DEREK HAROLD ROSENDAHL
Norman, Oklahoma
2008

IDENTIFYING PRECURSORS TO STRONG LOW-LEVEL ROTATION WITHIN NUMERICALLY SIMULATED SUPERCELL THUNDERSTORMS: A DATA MINING APPROACH

A THESIS APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

_____
Dr. Kelvin K. Droegemeier, Chair

_____
Dr. Amy McGovern, Co-Chair

_____
Dr. Rodger A. Brown

_____
Dr. Howard B. Bluestein

ACKNOWLEDGMENTS

I am truly grateful to those who have assisted in this research effort and for the incredible support of so many others who have helped bring me to where I am today. I first would like to thank my advisor and committee chair Dr. Kelvin Droegemeier for giving me this opportunity and for the support and encouragement he has provided as well as for being both a mentor and a friend. I would also like to thank my co-committee chair Dr. Amy McGovern for bringing her vast knowledge of computer science and data mining to the project. This was a highly cross-disciplinary, collaborative effort and could not have been a success without her involvement. Thanks also go to Dr. Rodger Brown, a committee member, who provided guidance and advice over every aspect of the project. I also thank committee member Dr. Howard Bluestein for the advice he provided.

Additional assistance was provided by students working under Dr. McGovern who helped with data mining logic and code development. These students include Adrianna Kruger, Meredith Beaton, Nathan Hiers and David Gagne II. Technical assistance also was provided by computer science department technician Jim Summers and the University of Oklahoma (OU) supercomputing center technician Brett Zimmerman.

I am grateful to the Center for Analysis and Prediction of Storms (CAPS) family at OU and the center for Collaborative Adaptive Sensing of the Atmosphere (CASA) for providing the funding for this project which was supported by the

Engineering Research Centers Program of the National Science Foundation under award number 0313747. Thanks also go to the rest of the OU School of Meteorology faculty and staff for their help over the past many years.

The numerical simulations of this project were generated at the Pittsburgh Supercomputing Center and I would like to especially thank David O'Neal for his technical assistance in setting up and running simulations on the supercomputers. Dr. Ming Xue, Dr. Fanyou Kong and Dr. Yunheng Wang all provided very helpful guidance for learning how to successfully run the Advanced Regional Prediction System (ARPS) model and Dr. Ed Adlerman helped in the development of simulation initial environments.

I am also deeply indebted to Dr. Alan Czarnetzki, my advisor and mentor at the University of Northern Iowa (UNI). Dr. Czarnetzki is one of the most gifted professors I have ever known, having the unique ability to adjust to anyone's knowledge level and help them understand any concept or problem, no matter now difficult. He also happens to be an outstanding human being with a great understanding of those things that are most important in life. Having the opportunity to be his student research assistant early in my undergraduate career was an extremely valuable experience that truly energized my interest and passion for the field of meteorology.

I also am grateful to the other professors and staff I interacted with at UNI along with the internship opportunities I acquired there. I especially want to thank

Todd Shea, Glenn Lussky and the others at the National Weather Service in La Crosse, WI as well as Craig Johnson at KWWL in Waterloo, IA.

I would like to thank the many friends I've made along the way, especially the Campbell Hall gang from UNI, for keeping life exciting and fun. I also would like to thank all of my family for their love and support, especially my grandparents, Mert and Elsie Rosendahl and Galin and Wilma Barth, as well as my siblings, Kari (and husband Roger), Jodi and Todd. A special thanks also to my many supportive aunts and uncles, Wally and Merlene, Phyl and Stan, Marc and Kathy, Gary and Chris, Jerry and Jackie, Ken and Bette, my Godparents Don and DiAnn and Wayne and Arletta, and the many others in my extended family.

My father- and mother-in-law, Bob and Marcia Thornburg, have given constant love and support and I am very thankful to have them in my life. With me living so far away from my own family they have always done everything they could to make sure I knew that I had family close by, which I have sincerely appreciated.

It is hard to say the right words to describe the important role my parents, Harold and Marlys Rosendahl, have played in my life. They have completely supported me in everything I have done and have taught me more than they could ever know. They instilled in me a strong sense of self discipline and the desire to work hard in everything I do. Growing up, my parents constantly encouraged me to believe there was nothing I could not accomplish if I set my mind to it. They always believed in me, even in those times when I may have doubted myself. It is for these reasons and so many more that I have become the person I am today. I am forever

grateful for the life lessons they taught me, which could never have been taught in a classroom or found in a textbook. The completion of this thesis is as much a credit to them as it is to me. I could not have done this without them.

And to the most important person in my life, my wonderful and beautiful wife Becca, I am continually amazed at your love, patience and kindness. You have been by my side every step of the way and I truly am blessed to have you in my life. Your encouragement and support has been unsurpassed. I love you so much.

Finally, I thank God for guiding me through life's many crossroads to bring me to where I am today.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 3

CHAPTER 4

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

ABSTRACT

Despite considerable progress made in recent decades in the observation, modeling, and theoretical understanding of tornadoes, warning and forecasting their occurrence remains a considerable challenge. Quantitative statistics clearly show that warning probability of detection (POD) and lead time have plateaued in recent years, with false alarm ratio (FAR) remaining relatively constant, principally because existing surveillance radars and hazardous weather detection methodologies suffer from fundamental limitations that allow key meteorological quantities and associated features to go undetected. New advances will be required if substantial improvements in warning and forecasting accuracy are to take place.

One promising avenue is the use of a data assimilation procedure, in *real-time*, which is capable of increasing the number of meteorological quantities available for use in detection algorithms. Applying algorithms to these real-time, gridded analyses should provide a considerable advantage over existing techniques, which, in the case of tornadoes, mostly utilize radar radial velocity and reflectivity data in their native spherical-polar coordinates. However, using assimilated data rather than direct observations will necessitate the development of a new suite of algorithms able to operate on regular grids and accommodate retrieved fields. The performance of these algorithms will depend upon their ability to identify storm features and feature interrelationships prior to the development of a severe weather event (e.g., tornado).

This study provides an initial framework for identifying important features and feature interrelationships, intended for future hazardous weather detection algorithms, which herald the development of strong low-level rotation within deep convective storms. Numerical simulations were used instead of observational data because simulations provided all meteorological fields within a 3-D gridded structure, analogous to future assimilated analyses, and could generate a large number of storms under controlled conditions that utilize varying initial background environments. Assimilated data sets based upon observations could not be used because too few of these data sets exist, and it would be difficult to verify the retrieved fields of those available. The low-level rotation qualifier was used because tornadic vortices can not be resolved with the model grid spacing employed here, though as computational advances foster higher resolution simulations, the methodology may be extended to include smaller scale vortices.

A total of 1168 numerically simulated storms were generated within initial environments characteristic of supercell storms and were categorized by whether they produced strong, weak or no low-level rotation. A computational data mining procedure was developed to search the gridded fields for meteorological precursors, occurring in repeatable patterns, that lead to the development of strong low-level rotation. An analysis was performed on storms producing "strong" and "no" low-level rotation and a separate analysis was performed on storms producing "strong" and "weak" low-level rotation.

Our results identified sets of precursors, in the form of meteorological quantities reaching extreme values in a particular temporal sequence, unique to storms producing strong low-level rotation. Statistical analyses were performed on the sequences (termed rules) to identify their significance and the highest rated rules consisted of the same meteorological quantities with small variations in temporal ordering. This implied that the *order* in which quantities reached extreme values was less significant than the requirement that the quantities simply *reach* extreme values. With this in mind, frequency distributions of quantity occurrence in the top rated rules were generated, identifying the most important quantities reaching extreme values prior to the development of strong low-level rotation. The top five quantities identified in storms exhibiting "strong" and "no" low-level rotation were: maximum in vertical vorticity stretching below 2 km, minimum in baroclinic vertical vorticity generation below 2 km, minimum in vertical vorticity stretching below 2 km, minimum vertical perturbation pressure gradient force above 2 km and maximum vertical perturbation pressure gradient force below 2 km. The second analysis group (storms with "strong" and "weak" low-level rotation) generated rules with comparable quantities but less statistical significance. The large number of rules identified by this study should prove useful in the development of algorithms for anticipating strong low-level rotation in real-time 3-D gridded assimilated analyses.

Simulations with finer grids could be used to extend this concept to the tornado scale.

CHAPTER 1

INTRODUCTION

One of the most societally important and scientifically challenging problems – effectively predicting an impending tornado with sufficient lead time and accuracy to provide adequate warning to mitigate life and property loss – has improved significantly over the past several decades. This improvement can be attributed to factors such as a greater understanding of tornadic storms, the integration of weather radar products and other observations into the warning decision process, enhanced watch/warning dissemination procedures and a greater awareness and response by the general public (e.g., Doswell et al. 1999; Moller 2001; Simmons and Sutter 2005). Despite the progress made, statistical measures show that tornado prediction still poses a considerable challenge to forecasters and is far from being a solved problem.

This challenge is illustrated in Figure 1.1, which shows nationwide tornado warning verification statistics of probability of detection (POD), false alarm ratio (FAR) and lead time during the period 1986-2007 as well as National Weather Service (NWS) goals through 2012. POD represents the likelihood that a warning was issued for an observed tornado and FAR represents the non-occurrence of a tornado following an issued warning (Wilks 2006, pp. 260-265). Lead time denotes the time from which a tornado warning was issued to the time a tornado first entered the warned area. In terms of the standard 2x2 contingency table (see Figure 1.2), POD and FAR are defined by the following:

$$POD = \frac{a\ (\text{hits})}{a\ (\text{hits}) + c\ (\text{misses})}, \tag{1.1}$$

$$FAR = \frac{b\ (\text{false alarms})}{a\ (\text{hits}) + b\ (\text{false alarms})}, \tag{1.2}$$

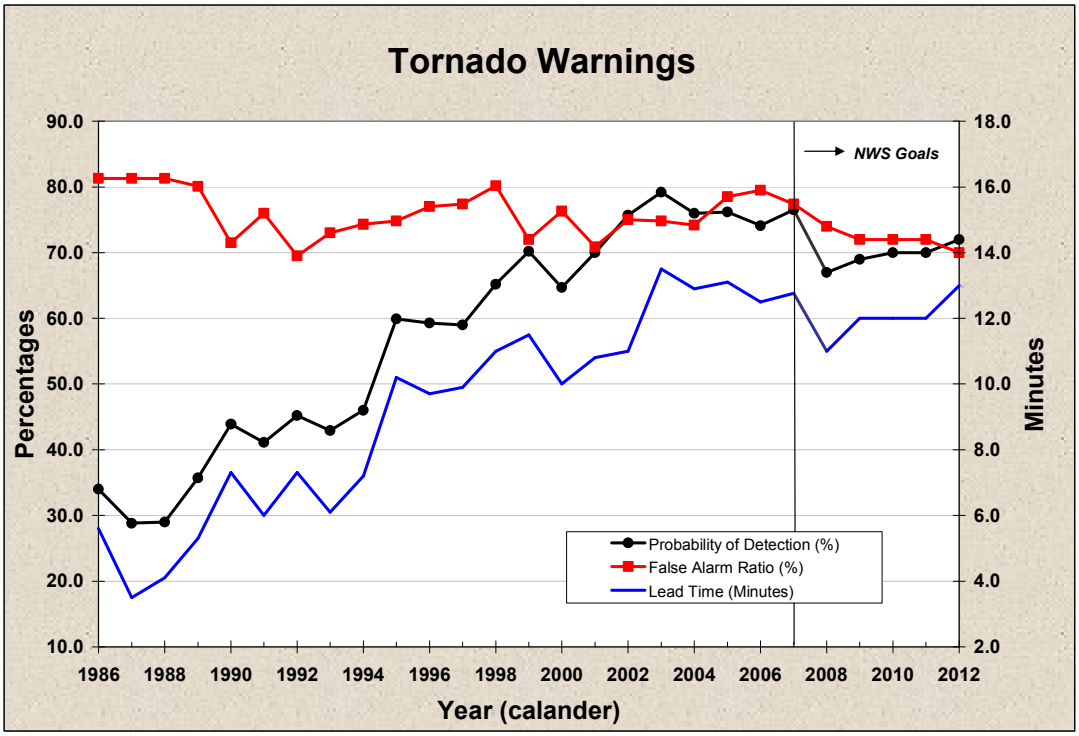


Figure 1.1: Nationwide tornado warning verification statistics from 1986-2007 as well as NWS goals for new storm-based warnings beginning in 2008: Probability of detection (black line with circles), false alarm ratio (red line with squares) and lead time (blue line) with future goals (same with dotted lines). [Data courtesy of B. MacAloney II, National Weather Service Performance Branch, 2008.]

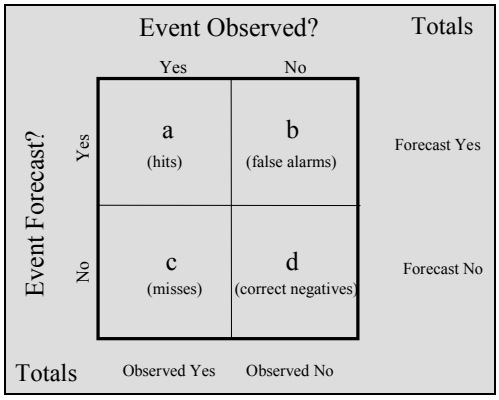


Figure 1.2: Standard 2x2 contingency table for observations and forecasts of a given event or set of events. [From Wilks (2006).]

The improvements over the past two decades are evident in Figure 1.1, with POD and lead time more than doubling during that period, resulting in current (2007) values of 76.5% and 12.8 minutes, respectively. These indices, however, appear to have reached a fundamental limit that is evident in the slight downward trend during the past four years and minimal change in future NWS goals. POD and lead time are even expected to *decrease* with the implementation of new storm-based warnings in 2008 and are not anticipated to re-establish their current values until beyond 2012. This implies that tornado prediction is approaching a maximum in capability using current understanding and technology and that new advancements will be required in both areas if significant improvements beyond NWS stated goals are to be achieved.

This assertion is further supported by FAR, which remained relatively constant over the past 20 years, oscillating around 75% with a current value of approximately 77%. Because FAR never exhibits substantial improvement during this period and is not expected to improve significantly in the future, new

advancements will be required to improve upon current capabilities in which roughly three out of every four tornado warnings are false alarms.[1] Thus, new technologies must be developed to improve tornado prediction with the goal of decreasing FAR while simultaneously increasing POD and lead time. That objective serves as an underlying motivation for this project.

Before attempting to develop a method for improving tornado prediction, however, an understanding is needed regarding the fundamental limit apparently being reached in POD and lead time and the ongoing high values in FAR. One influencing factor is the difficulty of improving all three performance measures simultaneously because of their mutual relationship. Improving POD to 100% simply requires warning on all possible storms, which also might improve lead time but would drastically increase the number of false alarms and drive FAR to extreme levels. Conversely, improving FAR to near 0% would require not issuing warnings on virtually any storms, or warning on only those storms for which a tornado has been confirmed. This approach would result in a nonexistent lead time and very low POD. Therefore, improving these performance measures is a difficult task requiring an appropriate balance of advancing all three in the proper directions.

Another factor influencing tornado warning performance involves the observing platforms utilized by forecasters to monitor the atmosphere. Doppler weather radar, with its ability to surveil the internal structure of a storm, has become

[1] The extent to which a relatively high false alarm rate translates into greater threat to the public via increased apathy to warnings continues to be debated and may have weaker linkage than previously suggested (Barnes et al. 2007; Harold Brooks, 2007, Personal Communication).

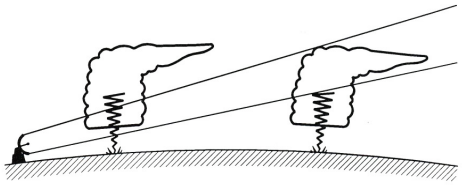
the standard tool for detecting and predicting tornadoes[2], especially because of the NWS Weather Surveillance Radar-1988 Doppler (WSR-88D) network across the United States (Crum and Alberty 1993). The WSR-88D arguably has played the most important role in improving tornado warning performance since its nationwide inception in the 1990's (e.g., Polger et al. 1994; Bieringer and Ray 1996; Simmons and Sutter 2005) and has been credited with providing an estimated 45% reduction in expected tornado fatalities and a 40% reduction in expected injuries per year (Simmons and Sutter 2005).

Despite these improvements, inherent limitations in the WSR-88D system impact its ability to detect important storm features and therefore restrict its tornado detection and prediction capabilities. One such limitation is the radar's inability to observe the full 3-dimensional (3-D) wind, temperature, pressure and each of the separate water substance fields. Observations are limited to precipitation intensity and the radial component of the wind in regions where sufficient scattering particles are present, resulting in only a partial representation of storm structure.

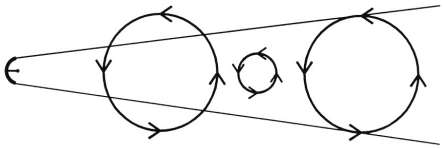
Another limitation is the radar horizon problem depicted in Figure 1.3(a). With a goal of covering the entire nation in the most cost effective manner, each radar in the WSR-88D network is required to collect data out to long ranges. With increasing distance from the radar, the beam center increases in altitude due to Earth's curvature and therefore can overshoot the lower and middle regions of a storm at

[2] For this study, the phrase "predicting tornadoes" refers to short-term, non-dynamical forecasting often referred to as nowcasting.

longer ranges. Consequently, important low- and mid-level pre-tornadic storm structure can be missed.



(*a*)



(*b*)

Figure 1.3: Schematic diagrams depicting (*a*) the radar horizon problem and (*b*) the aspect ratio problem. [From Burgess et al. (1993).]

A third limitation is the aspect ratio (or beam spreading) problem depicted in Figure 1.3(b). The ability of a WSR-88D to detect vortices in the radial wind field is dependent upon the ratio of the radar beam width to the vortex size (Burgess et al. 1993). If the beam width is much larger than the vortex diameter, the vortex cannot be detected. The center vortex in Figure 1.3(b) illustrates this scenario, which typically occurs for tornadoes sampled by the WSR-88D at longer range. The vortex on the right represents another instance in which the circulation may be detected but its characteristics would change depending upon the center location of the beam relative to the vortex center (see Wood and Brown 1997). The vortex on the left has sufficient size compared to the beam width and therefore would be detected with only

slight variations in its characteristics based on beam position. The aspect ratio limitation becomes worse with increasing range from the radar as the beam spreads.

A fourth limitation is the relatively long revisit times of the radar beam for any given region of a storm. The WSR-88D completes 360º scans across 9-14 elevations throughout the atmosphere, which typically requires four to six minutes (OFCM 2006). This can be too lengthy for many tornadoes that develop and dissipate in only a few minutes.

Because of these limitations, forecasters are encouraged to incorporate additional information into their warning decision process to gain a better understanding of the storm environment as well as the storm itself. The additional information may come in the form of observations from surface stations, satellites, radar wind profilers, lightning detection networks or storm spotters and may be used to identify important features such as shallow surface boundaries, enhanced near-ground storm-relative helicity, strong low-level convergence or characteristics associated with the updraft or rear flank downdraft (WDTB 2002). As with the WSR-88D, though, this additional information suffers from limitations because each observing platform is confined to observing specific quantities at finite time intervals and locations and thus cannot provide a complete characterization of relevant features. Therefore, even when utilizing all available data, key storm structure may still go undetected, thus allowing features indicative of the onset of tornadogenesis to be missed.

From this discussion, one can conclude that the primary cause of the asymptotic behavior exhibited by POD and lead time, and the ongoing high values of FAR in tornado warning statistics, appear related to the inability of observing platforms to provide a complete assessment of conditions within tornadic storms and their near-storm environments, particularly in the boundary-layer. The inclusion of new observing systems, such as the WSR-88D, has allowed forecasters to distinguish more features and therefore improve tornado detection and lead time; however, without a complete view of the most relevant meteorological quantities at sufficiently high temporal and spatial resolution, key storm features and morphology are being overlooked, resulting in the current limits in POD and lead time. Additionally, the difficulties in distinguishing tornadic from non-tornadic storms produce a relatively high degree of uncertainty in the prediction process that keeps FAR larger than desired.

Therefore the best way to improve POD, lead time and FAR, and our understanding of tornadoes, appears to be the development of a single, real-time analysis of the atmosphere, having both fine spatial and temporal resolution, that contains all relevant meteorological fields throughout the entire troposphere. Such an analysis already is being developed within the data assimilation community and is producing promising results (e.g., Tong and Xue 2005; Hu and Xue 2007; Brewster et al. 2007; Xue et al. 2007). Indeed, the availability of such assimilated analyses in real time appears practicable within a few years (Ming Xue, 2007, personal communication).

Data assimilation takes observations from individual platforms – in all of their varying geometries and dependent variables – and integrates them into a single gridded analysis in which unobserved quantities are retrieved and all quantities in the analysis are dynamically and thermodynamically consistent. Even though the assimilation process tends to decrease the spatial resolution of the raw observational data, the gridded analysis appears to be superior to using solely individual observing platforms because it retrieves additional meteorological fields at very fine spatial and temporal resolution that satisfy the governing equations or other constraints with minimal error relative to the true observations (Robert Fritchie, 2007, personal communication). The high resolution fields are capable of resolving previously unobserved storm structure and, therefore, have the potential to improve tornado warning performance and our understanding of tornadogenesis by providing additional information not currently attainable. Additionally, when hazardous weather detection algorithms are applied to the assimilation analyses, new observing systems can be added to the assimilation process, thereby improving the analyses, without requiring a change to detection and anticipation algorithms that utilize them.

Incorporating the newly resolved tornadic storm features into the warning decision process will first necessitate the development of a more sophisticated method for identifying and tracking the significant features which, in turn, will require a greater understanding of the features and feature interrelationships that are most influential to the development of a tornado. Yet one important question remains: What will be the most effective approach for identifying these important

storm features and their interrelationships and quantifying their relative importance to tornadogenesis?

The most straightforward approach would be to examine one or two tornadic storm cases using WSR-88D and/or mobile radar data as well as any other available observations in an attempt to identify all important features present and distinguish which generate the most influential precursors to tornadogenesis. This technique, however, lacks the data necessary to provide a full representation of all meteorological fields surrounding and within a storm, particularly near the ground. That is, it suffers from the very limitations we seek to overcome because it uses data from observing platforms that give rise to the existing limited warning performance. Additionally, the findings would only be relevant to environments similar to the specific cases studied.

A second approach would be to examine one or two high resolution numerically simulated storms in search of significant features that herald tornado development. This technique would have difficulty resolving actual tornadic vortices due to computational limitations but still holds the tremendous advantage of providing all meteorological quantities at each grid point in the model and therefore addresses the data insufficiency problem. Indeed, this is the approach taken by many numerical modelers (e.g., Wicker and Wilhelmson 1995; Adlerman et al. 1999) and although valuable insight has been produced, the results can neither be generalized nor applied operationally since, again, the results are only relevant to environments similar to the chosen cases.

A superior approach, therefore, would go beyond analyzing a few cases and instead incorporate a large number of numerically simulated storms, perhaps on the order of hundreds, spanning an extensive parameter space of possible simulated "tornadic" environments. This technique provides both the high resolution meteorological fields and the broad parameter space necessary for identifying repeated storm-feature patterns linked to "tornadogenesis" that occur across a broad range of simulated storms. Initially, the approach would be restricted to investigating the development of strong low-level rotation rather than a tornado due to the model's inability to resolve tornadic vortices under current computational limitations. This approach, however, does have the potential of determining whether a universal set of precursors to significant low-level rotation exist in high resolution data fields which can be directly applied to the assimilated gridded analyses of the future since the analyses will contain the same 3-D gridded domain structure as the numerical simulations. And with future computational advances, the method can be extended to include simulated tornadic vortices. Therefore, this is the approach utilized in this study.

Because there exists a vast number of observed tornadic storm environments across different storm types, confining the parameter space to the supercell storm spectrum is deemed most appropriate for this initial study because supercells are the most fundamental observed tornadic storm type and contain a considerable set of identifiable features. If no significant precursors to strong low-level rotation can be found within simulated supercell storms, then it is unlikely any will be found in other,

less structured, simulated storms capable of producing strong low-level rotation such as squall lines. The parameter space of this study therefore attempts to cover the full spectrum of numerically simulated supercell storms.

Analyzing such a large number of storm cases requires a procedure superior to the standard manual approach of inspecting each case individually and interrelating storm characteristics across all cases (e.g., Droegemeier et al. 1993; Adlerman and Droegemeier 2005). Automated data mining (e.g., Hand et al. 2001) therefore is applied to computationally sort through the large number of numerically simulated data sets to extract useful storm feature information that otherwise would be too difficult and time consuming to obtain by hand. This is the first time for a data mining technique to be applied in this specific manner.

Therefore *the overall goal of this study is to apply automated data mining techniques to hundreds of numerically simulated supercell storms – across a full spectrum of supercell environments – in an attempt to identify repeated patterns among storm features that herald the development of strong low-level rotation.* The results will be directly applicable to the real-time, 3-D gridded assimilated analyses to be produced in the future because they will provide knowledge on the previously unobserved influential supercell storm features available in the new high resolution 3-D gridded data fields. These storm features can be incorporated into the development of new detection and anticipation algorithms capable of improving the warning performance measures of POD, FAR and lead time once computational advances

foster the extension of this study's methods towards the identification of a developing tornado.

The remainder of the thesis is organized as follows. Chapter 2 provides a discussion of the relevant literature associated with this study, including a review of supercell structure and evolution. The methodology is described in Chapter 3, with a discussion of the storm simulations, preliminary and final methods for identifying relevant storm features, and an overview of the data mining technique utilized. Chapter 4 presents results and Chapter 5 provides a summary as well as limitations of the current study with suggestions for future work.

CHAPTER 2

LITERATURE REVIEW

This chapter contains a review of the relevant literature on supercell storms. The first section provides an overview of key structural features of a supercell storm discovered through observational and numerical modeling studies. The second section contains a summary of the hypothesized origins of supercell rotation at both mid- and low-levels.

2.1 Supercell Storm Structure

Early research on the structure and evolution of severe thunderstorms began to uncover characteristic differences between average storms and especially intense storms, i.e., storm motion typically followed mean low- to mid-tropospheric winds for average storms (Byers and Braham 1949) while especially intense storms moved to the right of the mean winds (Byers 1942; Newton and Katz 1958). Following the advent of weather radar in the 1950's, a low-level radar reflectivity signature was discovered in association with strong, violent storms. This signature was first documented by Stout and Huff (1953) (see Figure 2.1) and consisted of a discrete region of reflectivity having an appendage on the right, rear flank extending roughly 90 degrees to the right of storm motion. The signature was later termed a "hook echo" by van Tassell (1955) due to its hook-like appearance (Figure 2.2). Fujita (1958) analyzed the evolution of these hook echoes within radar reflectivity data and

determined they were associated with storm rotation. His findings were confirmed by subsequent Doppler radar studies using wind velocity data to identify storm regions of strong horizontal wind shear and rotation (e.g., Donaldson 1970; Lemon et al. 1975; Ray et al. 1975; Ray 1976; Brandes 1977a; Burgess et al. 1977; Lemon 1977; Barnes 1978a,b).

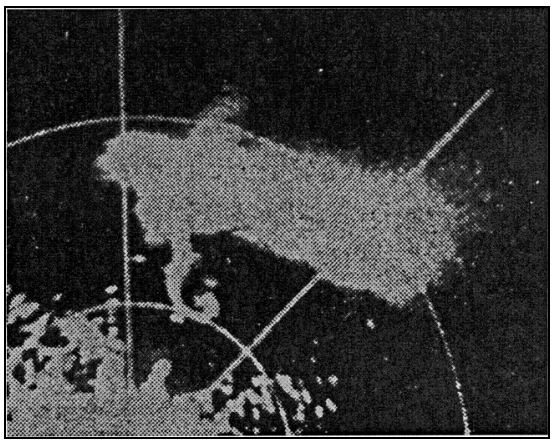


Figure 2.1: Radar reflectivity of first documented hook echo. This reflectivity structure was associated with a tornadic storm near Champaign, IL on 9 April 1953. [From Stout and Huff (1953).]

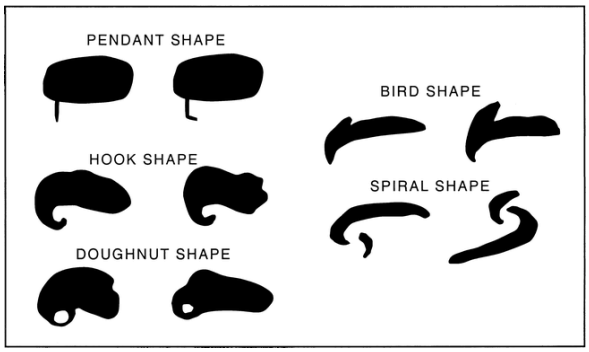


Figure 2.2: Illustration of the variation in hook echo shapes observed in radar reflectivity data. [From Fujita (1973).]

The low-level hook echo was found to be part of a 3-dimensional (3-D) reflectivity structure common to especially intense severe storms (Figure 2.3). The strong reflectivity region centered near the hook echo's right rear flank extends vertically to the upper levels of the storm where it slants horizontally to form an echo overhang (e.g., Browning 1964; Marwitz 1972a). Directly beneath this upper level echo overhang is a region of minimal reflectivity associated with a storm's strong updraft and identified as the weak echo region (WER) (Chisholm 1973; Lemon 1977). When the updraft is especially strong, the WER can penetrate the upper-level reflectivity region (i.e., will be completely surrounded by moderate to strong reflectivity at mid to upper levels) and is called a bounded weak echo region (BWER) or vault (e.g., Browning and Ludlam 1962; Browning and Donaldson 1963; Browning 1964, 1965) The BWER is labeled in Figure 2.3(*b*) and can be seen as an echo-weak hole beneath the upper-level reflectivity max in the corresponding reflectivity horizontal cross-sections (Figure 2.3(*a*)).

Lemon (1980) created a conceptual diagram illustrating how updraft intensity can affect storm structure to produce the characteristic features observed in radar reflectivity (Figure 2.4). Figure 2.4(*a*) illustrates a weaker updraft that is significantly tilted by moderate to strong shear with precipitation generally falling in a symmetric pattern towards the surface. The stronger updraft in Figure 2.4(*b*) is more vertically oriented and capable of producing a WER at mid-levels. Within an intense updraft (Figure 2.4(*c*)), the cloud droplets do not have sufficient time to grow to radar-detectable size until they are in the upper portions of the storm—leading to the

formation of the BWER at midaltitudes. The extension of strong reflectivity to the surface immediately adjacent to the intense updraft leads to the characteristic hook echo structure at low levels.

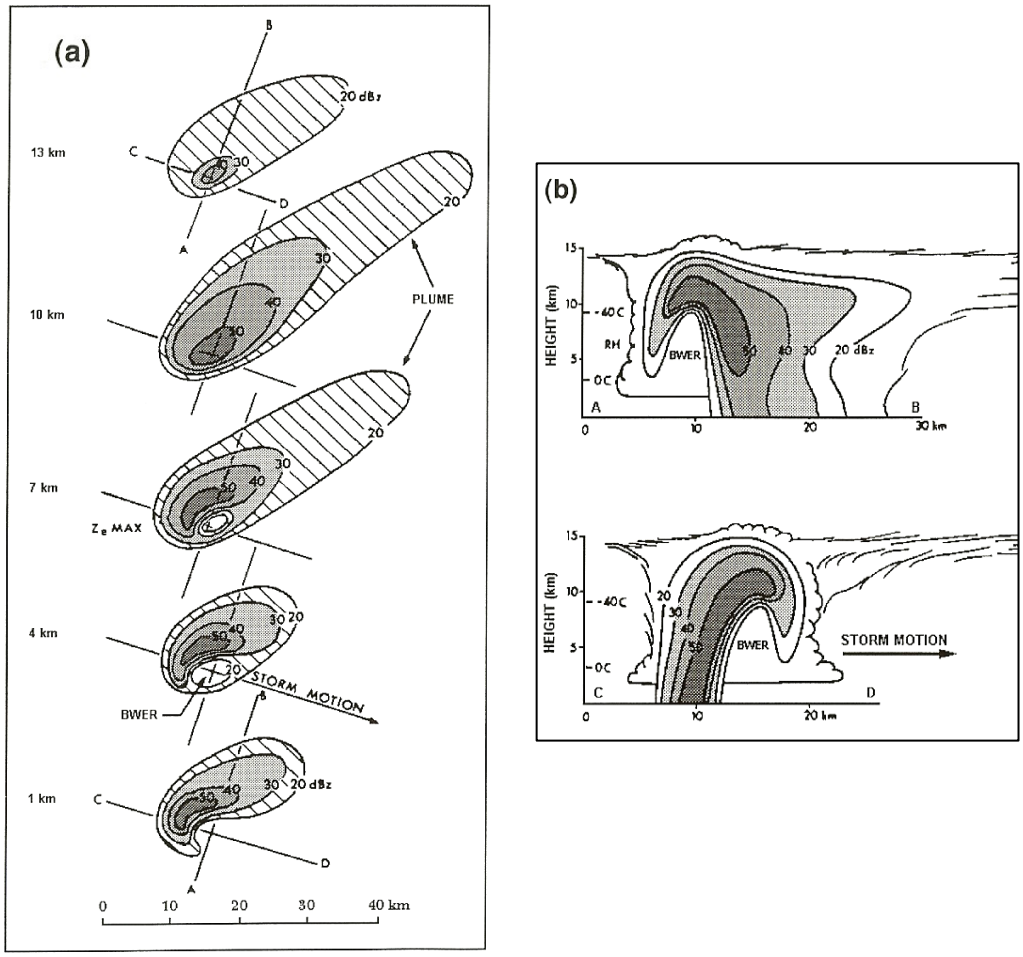


Figure 2.3: Radar reflectivity (dBZ) of a supercell storm looking (*a*) in the horizontal at 1, 4, 7, 10, and 13 km AGL and (*b*) vertically from cross sections A-B (top diagram) and C-D (bottom diagram). [From Chisholm and Renick (1972).]

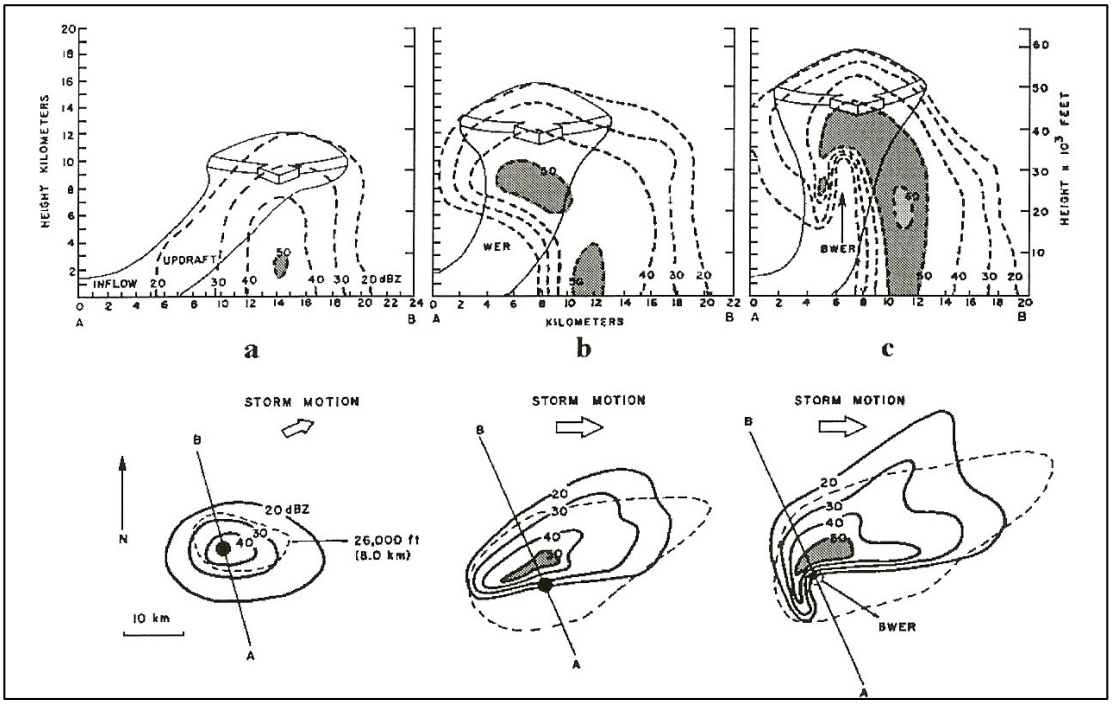


Figure 2.4: Schematic diagram illustrating the structural differences of storms with (*a*) a weak updraft that is tilted by strong wind shear, (*b*) moderate to strong updraft capable of producing a WER and (*c*) intense updraft producing a BWER and low-level hook echo signature. Lower plots provide horizontal cross-section of radar reflectivity (dBZ) and upper plots provide vertical cross-section of reflectivity along the line A-B with updraft outlined. Dashed line in horizontal cross section is the 20 dBZ outline at 8 km. [From Lemon (1980).]

A detailed analysis of the internal wind flow pattern within these strong, long lasting severe thunderstorms was conducted by Browning and Ludlam (1962) and Browning and Donaldson (1963). Their investigations used radar reflectivity data and surface observations to deduce that the storms contained a strong updraft region distinctly separated from a downdraft region. These studies led Browning (1964) to propose a conceptual model of the pertinent characteristics for the unique flow pattern within these especially severe storms (Figure 2.5). In the model, a single, dominant

updraft continuously ingests warm, moist air from the low-level storm inflow. The heavy precipitation "north" of the updraft (far side in his diagram) combines with dry midlevel air which wraps around the front (east) of the updraft to produce a downdraft aided by evaporational cooling. The cool, dry downdraft air reaches the ground and spreads out along the surface where it eventually propagates towards the south and east and converges with the warm moist inflow air, thereby enhancing updraft growth. This process provides a nearly steady-state flow pattern in which the separated updraft and downdraft do not inhibit one another, allowing the storm to sustain itself for a substantial period of time. When a tornado was present it was located along the updraft-downdraft interface near the surface outflow boundary. A storm exhibiting the unique characteristics of this conceptual model was later termed a "supercell" by Browning (1968).

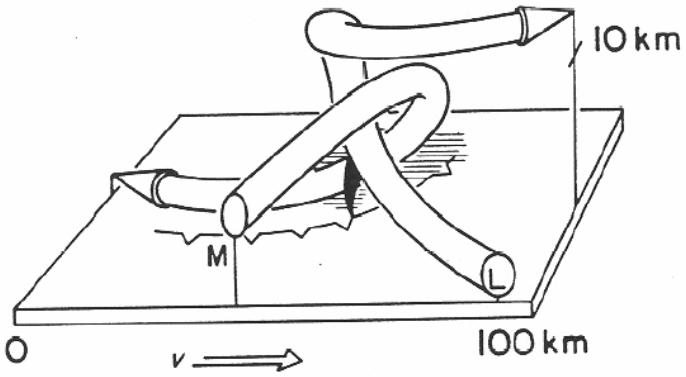


Figure 2.5: Conceptual model of the three-dimensional air flow pattern associated with a right-moving supercell thunderstorm developed by Browning (1964). All components are given relative to storm motion (v) and the far (right) side of the diagram is regarded as north (east). Arrow tube shapes represent airflow originating at low-levels (L) and mid-levels (M). Precipitation reaching the ground is depicted with a hatched area and the gust front of the downdraft outflow is shown as a cold frontal boundary. The typical location of a tornado is also included (shaded black). [From Browning (1964).]

Lemon and Doswell (1979) provided one of the first comprehensive conceptual models of a supercell thunderstorm by examining radar, surface and visual observations from a vast number of contemporary studies (e.g., Marwitz 1972a,b; Brown et al. 1973; Moller et al. 1974; Lemon et al. 1975, 1977, 1978; Burgess et al. 1976, 1977; Lemon 1976; Brandes 1977a,b, 1978; Darkow and McCann 1977; Golden and Purcell 1977, 1978; Nelson 1977; Barnes 1978a,b). Their conceptual model for the low-level structure of a supercell is shown in Figure 2.6 and corresponds with a time period when the storm is undergoing a rapid intensification of its near-ground rotation (tornadogenesis). A strong updraft is located on the warm-inflow (east) side of the hook echo reflectivity appendage with the main precipitation downdraft on the storm's forward side called the forward flank downdraft (FFD) and another downdraft adjacent to the updraft on the storm's rear side called the rear flank downdraft (RFD). Both the FFD and RFD produce outflow boundary gust fronts which merge near the interface between the updraft and RFD in a location most often associated with tornadic development.

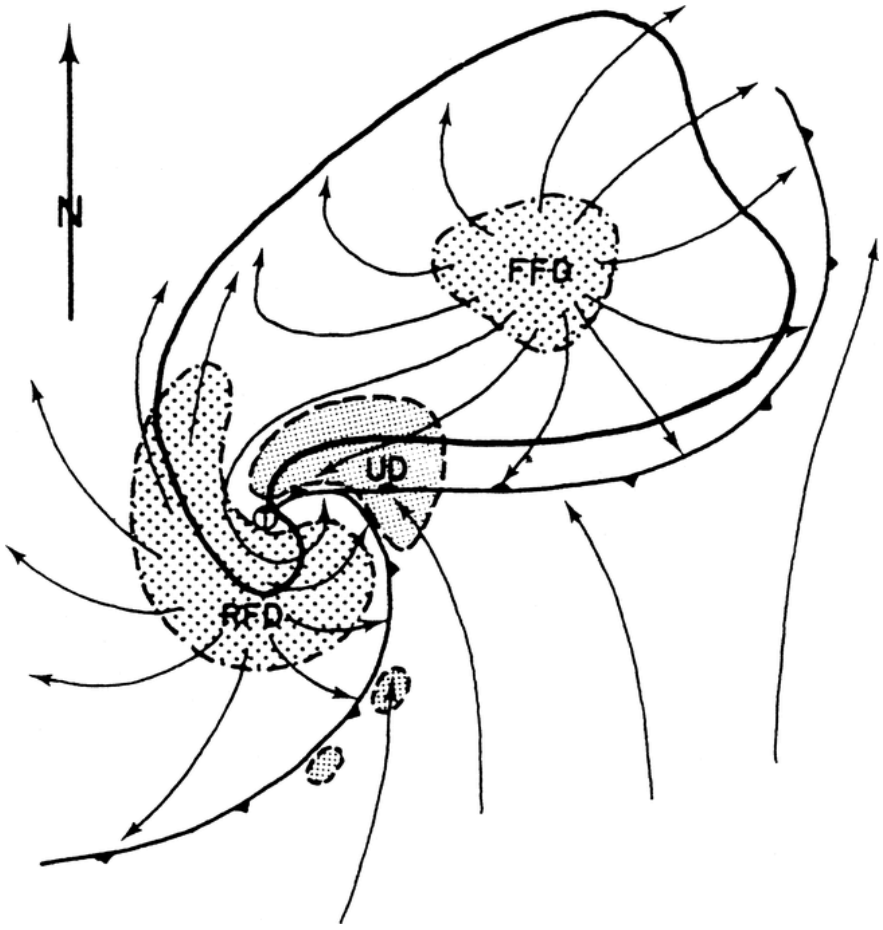


Figure 2.6: Schematic diagram proposed by Lemon and Doswell (1979) for the low-level features associated with a tornadic supercell thunderstorm. The region of observed radar reflectivity is outlined in solid black. Updrafts are finely stippled regions outlined with a dashed line while downdrafts are coarsely stippled regions outlined with a dash-dot line. The dominant updraft (UD), forward flank downdraft (FFD) and rear flank downdraft (RFD) are labeled accordingly. Streamlines (relative to ground) are shown (arrows) and gust fronts are identified using standard frontal symbols. Location of the tornado is shown as a small encircled "T." [From Lemon and Doswell (1979).]

Lemon and Doswell (1979) also provide 3D diagrams corresponding to the evolving supercell storm (Figure 2.7). Figure 2.7(*a*) represents the supercell's developing stages when it is dominated by a strong, cyclonically rotating updraft with an FFD and associated precipitation on its forward flank. The RFD is proposed to

develop at this time when strong mid- to upper-level winds encounter the intense updraft which induces higher pressure on the updraft's upwind side and therefore a downward-directed pressure gradient force. The initiated RFD is then enhanced by evaporative cooling due to entrained dry air and by precipitation drag until it extends to the surface, acquiring cyclonic rotation along its decent in conjunction with the rotating updraft (Figure 2.7(*b*). Eventually, the updraft and RFD form a split mesocyclonic structure with a stronger, isolated region of rotation developing along their interface (Figure 2.7(*c*)). The FFD and RFD gust fronts begin to occlude at this time and the low-level storm structure coincides with the characteristic features shown in Figure 2.6. As the RFD surface outflow surges beneath the updraft, the warm, moist inflow air is eventually cut off causing the updraft and center of rotation to weaken as a new updraft and center of rotation develop near the intersection of the occluding gust fronts (Figure 2.7(*d*)). This occlusion process repeats itself as new updrafts and centers of rotation are established in a cyclic process, thus supporting a supercell's long lifespan (e.g., Burgess et al. 1982).

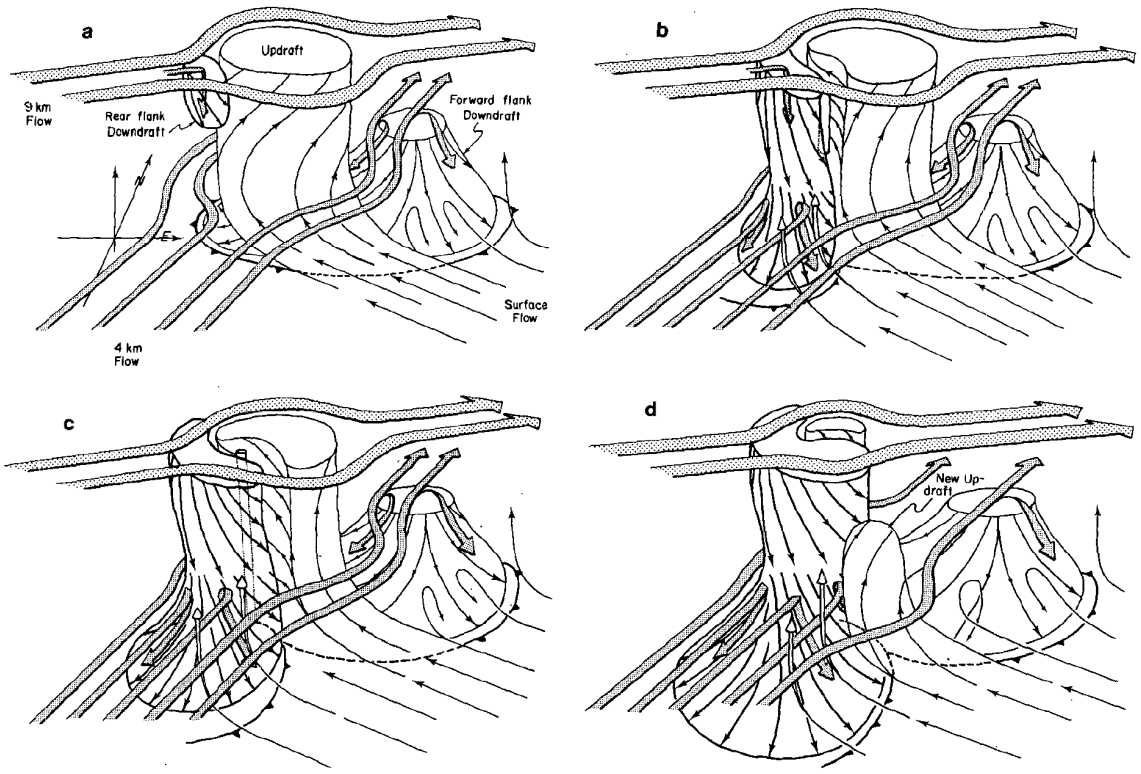


Figure 2.7: Three-dimensional schematic model of an evolving supercell storm proposed by Lemon and Doswell (1979) for (*a*) early stages dominated by a strong updraft with developing RFD, (*b*) RFD development to the surface, (*c*) split updraft-RFD mesocyclonic structure with strong vortex centered at their interface (time period associated with low-level features in Figure 2.6) and (*d*) occluding FFD and RFD gust fronts supporting the development of a new updraft and center of rotation. Storm relative winds are shown as arrows of various sizes and gust fronts are shown with frontal symbols. Updraft, RFD and FFD structures are labeled accordingly in (*a*) and new updraft development is labeled in (*d*). [From Lemon and Doswell (1979).]

The general supercell conceptual model proposed by Lemon and Doswell (1979) has remained essentially unchanged over the years, confirmed in both observational studies (e.g., Brandes 1984a,b; Hane and Ray 1985; Johnson et al. 1987; Dowell and Bluestein 1997, 2002; Beck et al. 2006) and numerical modeling studies (e.g., Klemp and Rotunno 1983; Rotunno and Klemp 1985; Wicker and

Wilhelmson 1995; Adlerman 1999).[3] The only notable exception has been the discovery of a unique downdraft (called the occlusion downdraft) located near strengthening low-level rotation when the RFD and FFD are rapidly occluding (Klemp and Rotunno 1983). Klemp and Rotunno (1983) proposed that this occlusion downdraft develops when strong low-level rotation dynamically induces lower pressure, generating a downward directed pressure gradient force. Subsequent observational studies (e.g., Brandes 1984a,b; Hane and Ray 1985; Wakimoto et al. 1996) and numerical modeling studies (e.g., Rotunno and Klemp 1985; Wicker and Wilhelmson 1995; Adlerman et al. 1999) support this conclusion. Figure 2.8 provides an illustration from Wicker and Wilhelmson (1995) of an occlusion downdraft forming in conjunction with rotationally induced lower pressure (Figure 2.8(*a*)) which then merges with the main RFD region to form one continuous downdraft (Figure 2.8(*b*)).

[3] The observational studies and numerical modeling studies listed are not meant to be exhaustive lists but rather a few selected examples among the large number of studies supporting the supercell conceptual model of Lemon and Doswell (1979).

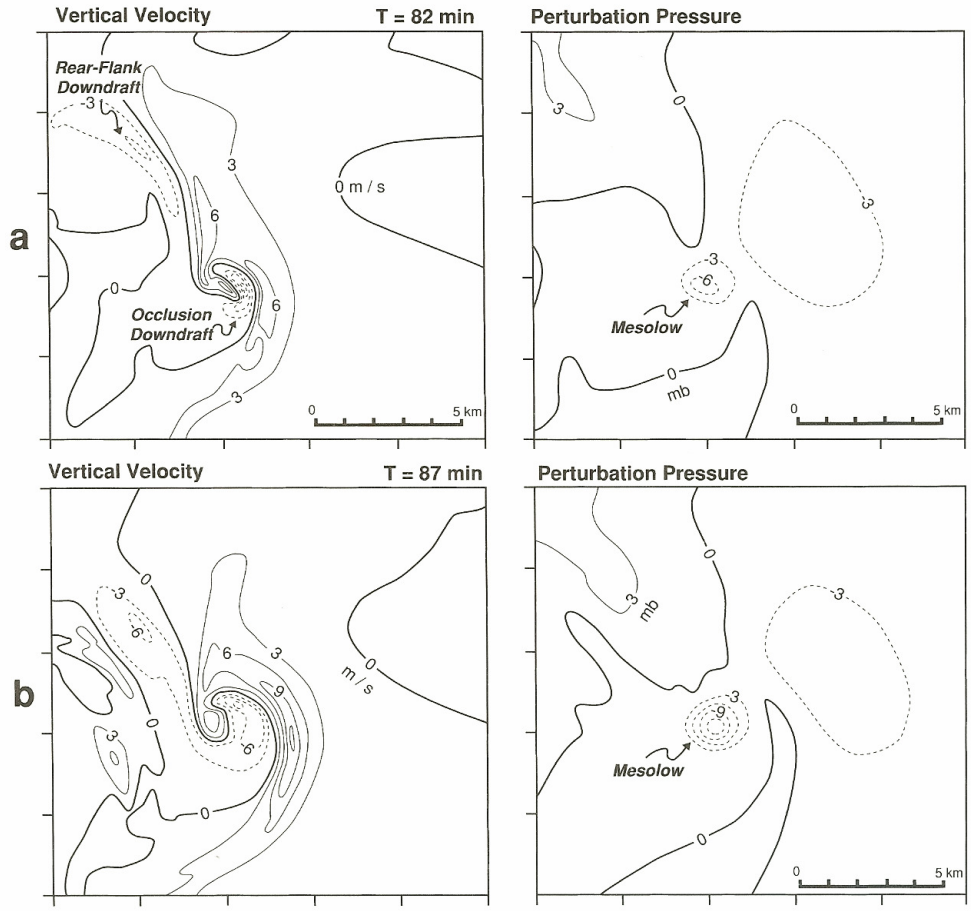


Figure 2.8: Developing occlusion downdraft from numerical simulation of Wicker and Wilhelmson (1995). (*a*) Occlusion downdraft forms in region of lower pressure dynamically induced by strong rotation (rotation not shown) and (*b*) a later time when the occlusion downdraft and RFD merge to form one continuous downdraft. Vertical velocity (left column) and perturbation pressure (right column) are plotted at 3 m s^{-1} and 3 mb intervals respectively with positive (negative) values shown as solid (dotted) lines. [From Wakimoto (2001); Adapted from Wicker and Wilhelmson (1995).]

Because the occlusion downdraft typically merges with or forms within the RFD, questions have been raised whether the RFD formation hypothesis of Lemon and Doswell (1979) and the occlusion downdraft hypothesis of Klemp and Rotunno (1983) are in conflict with one another (e.g., Klemp 1987; Wakimoto 2001). The former identifies the mid-level downdraft as being responsible for the development of

low-level rotation as the descending air approaches the surface while the latter is the reverse with low-level rotation being responsible for initiating the downdraft. In a review on hook echoes and RFD's, Markowski (2002) concludes there should be no conflict between the two hypothesized mechanisms when the RFD and occlusion downdrafts are considered two distinct entities having two separate forcing mechanisms. After the RFD forms at mid- to upper-levels it descends to the surface and aids in the development of low-level rotation (see discussion in next section). Once low-level rotation is strong enough it dynamically induces lower pressure in its vicinity and leads to the development of the occlusion downdraft. Therefore, the occlusion downdraft can be considered a "rapid, small-scale *intensification* of the RFD" having a separate forcing mechanism (Markowski 2002).

Thus far, the discussion has centered on the structure and evolution of a supercell storm's general features. Most of these unique features, however, can be attributed to the defining characteristic of a supercell storm which is storm rotation. Therefore, the next section provides an overview of the origins of storm rotation at various levels in a supercell.

2.2 Development of Supercell Rotation

The following discussion focuses on the mechanisms responsible for supercell rotation and is separated into the development at midlevels (3-5 km AGL) and from low levels (approximately 100 – 1000 m) to near the surface. Within the discussion, other supercell storm attributes, such as storm splitting and storm propagation, will be

briefly described to build the context necessary for understanding the rotational characteristics of a developing supercell.

2.2.1 Midlevel Rotation

Early work by Rotunno (1981) and Lilly (1982) suggested midlevel rotation resulted from vertical tilting of horizontal vortex lines by an updraft, which would produce cyclonic and anticyclonic vertical vortices straddling the updraft center. Based on these earlier studies, Davies-Jones (1984) demonstrated that an updraft could acquire rotation if the storm-relative horizontal environmental vorticity tilted into the updraft had a streamwise component. In his simplified example (Figure 2.9), the shear vector points to the right (east), producing environmental horizontal vorticity directed towards the north. To represent a storm updraft, an axisymmetric vertical displacement peak of the isentropic surfaces is placed within the flow. This displacement lifts the vortex lines due to the "frozen fluid lines" effect explained by Helmholtz's theorem[4] resulting in a positive and negative vertical vorticity couplet straddling the updraft on the near (south) and far (north) sides respectively.

Whether the updraft acquires net rotation or not depends on the orientation of the storm relative inflow winds. Figure 2.9(a) contains storm relative flow in the shear direction (across vortex lines) resulting in the cyclonic and anticyclonic vortices on the right and left sides respectively of the storm-relative flow as it rises into the updraft. These two vortices are equal in size and strength and therefore the updraft

[4] Helmholtz's theorem states that vortex lines move with the fluid they are in. The theorem applies to inviscid flow where viscous (friction) forces are considered small.

acquires no net rotation. Because the horizontal vortex lines are perpendicular to the flow in this case, the environmental horizontal vorticity is labeled "purely crosswise."

In the other extreme case, Figure 2.9(b) contains storm-relative flow across the shear. Because the flow is parallel to the vortex lines the environmental horizontal vorticity is labeled "purely streamwise." The storm relative flow ascending into the updraft is aligned with the cyclonic vortex while the anticyclonic vortex is on the downstream side of the updraft peak in a region of either less intense updraft or even downdraft. The net result is a cyclonically rotating updraft. Therefore, Davies-Jones (1984) concluded that an updraft within an environment containing horizontal vorticity would acquire net rotation if the storm relative flow produced a streamwise component of the vorticity as it is tilted into the updraft (i.e., is not purely crosswise).

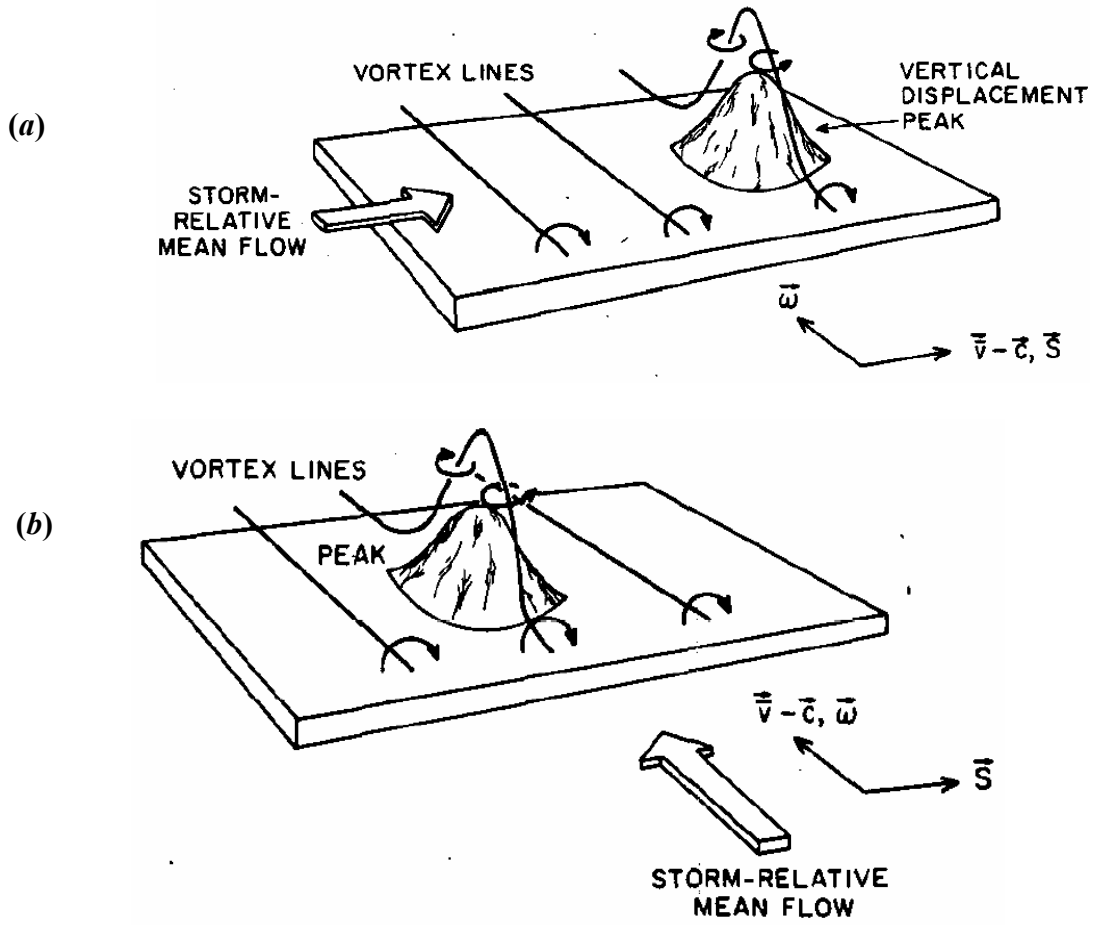


Figure 2.9: Illustration of how an updraft can acquire net rotation in an environment with horizontal vorticity depending on the storm-relative mean flow. Horizontal vortex lines (solid lines) and their orientation (circular arrows and $\vec{w}$) associated with environmental wind shear ($\vec{S}$) are shown as well as the storm-relative mean flow (large white arrow and $\bar{\vec{v}}$ - $\vec{c}$). Updrafts correspond to the "bump" in the isentropic surfaces. (*a*) Purely crosswise horizontal environmental vorticity where storm-relative flow is in the shear direction (across vortex lines) producing no net rotation. (*b*) Purely streamwise horizontal environmental vorticity where storm-relative flow is across the shear (parallel to the vortex lines) producing net cyclonic rotation. [From Davies-Jones (1984).]

The streamwise vorticity concept of Davies-Jones (1984) only partially describes the origin and maintenance of mid-level rotation within a supercell storm because it does not explain the roles of storm splitting or updraft propagation.

Rotunno and Klemp (1982, 1985) provided a rotationally-induced propagation theory which included these additional characteristics. Their theory is best described by initially considering an axisymmetric updraft within unidirectional shear (Figure 2.10(a)). The westerly shear generates environmental horizontal vorticity oriented towards the north (indicated by the horizontal vortex lines and circular arrows). The presence of an updraft tilts the vortex lines into the vertical, producing a cyclonic-anticyclonic vertical vorticity couplet on the updraft's south and north sides respectively with strongest rotation located at mid-levels (e.g., Wilhelmson and Klemp 1978). This initial stage produces an updraft with no net rotation because each vortex rotates in an equal and opposite direction relative to storm inflow (e.g., "purely crosswise" example shown in Figure 2.9(a)).

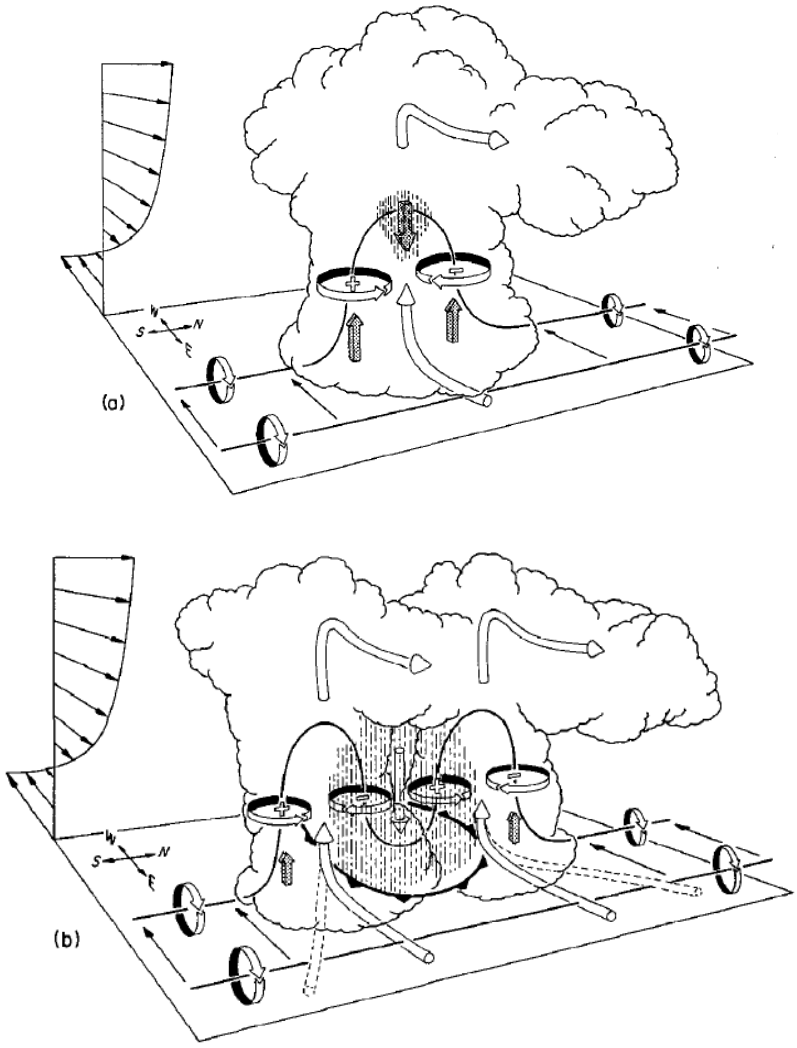


Figure 2.10: Conceptual diagram of the development of storm rotation at midlevels through vortex line tilting of horizontal vorticity generated in unidirectional westerly wind shear. (*a*) Initial stages as updraft lifts horizontal vortex lines generating a vertical vorticity couplet straddling the updraft center. (*b*) Splitting stage when downdraft splits initial updraft into two updrafts propagating toward regions of enhanced updraft growth on their outer flanks caused by rotationally induced lower pressure at mid-levels as well as the advancing downdraft outflow at the surface. Solid black lines represent vortex lines with orientation provided by circular arrows as well as plus and negative signs for vertical components. Thin black arrows on wind profile (southwest corner) and at the surface provide environmental winds and storm-relative airflow is given by cylindrical arrows with dashed cylindrical arrows representing how the airflow orientation evolves with time. Shaded arrows represent upward or downward directed forcings that initiate upward or downward motion. Hatched area corresponds to precipitation and cold front line indicates boundary of downdraft outflow propagating across the surface. [From Klemp (1987).]

As precipitation begins to accumulate above the updraft center, a downdraft forms, splitting the updraft and forcing the two vertical velocity maximums towards the mid-level vortices on the north and south flanks (Figure 2.10(b)). The two vortices assist this propagation by enhancing upward motion in their periphery by dynamically inducing low pressure and generating an upward directed vertical pressure gradient force. The leading edge of the downdraft's cool outflow also enhances updraft growth near the outer flanks as it travels along the surface away from the downdraft. A new pair of vortices is formed when vortex lines are tilted downward in the downdraft. These vortices do not generate the same updraft enhancement as the outer vortices because they are in a region of either downdraft or less intense updraft. They also are not able to ingest the same warm, moist air available to the vortices on the outer flanks. Therefore the southern (northern) updraft propagates to the right (left) of the shear vector due to the enhanced updraft growth on its right (left) flank. By propagating across the shear, the inflow into the southern (northern) updraft contains streamwise (antistreamwise) horizontal vorticity (Davies-Jones 1984) and acquires cyclonic (anticyclonic) rotation. This process is self-sustaining because the southern (northern) rotating updraft continues its across-shear propagation towards the enhanced updraft growth of the leading vortex which always remains slightly ahead of and to the right (left) of the central updraft location. The updrafts continue to rotate as streamwise (antistreamwise) environmental vorticity is continuously tilted into the vertical and stretched by the updraft.

The mirror image, counter-rotating storms described by this theory, however, are rarely observed in the real atmosphere. In fact, Davies-Jones (1986) later found that roughly 98% of radar detected, strongly rotating storms rotated in the same direction (cyclonically)[5]. To explain the discrepancy, Rotunno and Klemp (1982, 1985) replaced the unidirectional wind shear with a wind shear vector rotating clockwise with height at lower levels which is more representative of environments producing strongly rotating supercell storms (Maddox 1976)[6]. A turning shear vector impacts storm evolution because high pressure perturbations form on the upshear side of the updraft while low pressure perturbations form on the downshear side with magnitudes proportional to updraft speeds (i.e., strongest at mid-levels)[7]. Figure 2.11 provides an illustration of how the pressure distributions change within a storm depending on the orientation of the wind shear vector. The original unidirectional wind shear case is shown in Figure 2.11(a) with high and low perturbation pressure regions oriented in the same location at all altitudes due to the wind shear vector pointing in the same direction throughout the vertical. Because pressure perturbations are strongest at mid-levels (due to stronger vertical velocities there) an upward (downward) directed pressure gradient force is generated at lower levels on the east (west) side of the updraft, enhancing (suppressing) updraft growth there. The vortices

[5] Davies-Jones (1985) analyzed strongly rotating storms from radar data and found 143 rotated cyclonically while only 3 rotated anticyclonically.
[6] Maddox (1976) analyzed a large number of wind profiles near tornadic storms and produced a composite wind sounding representative of the environments.
[7] Perturbation pressure was found to be proportional to the dot product between the wind shear vector and the vertical velocity horizontal gradient vector.

present on the north and south flanks are affected equally by this process so neither is preferentially enhanced.

When the vertical wind profile rotates clockwise with height (Figure 2.11(b)), the wind shear vector and related high and low perturbation pressure regions also rotate clockwise with height. The orientation of the high and low pressure regions result in an upward (downward) directed pressure gradient force on the south (north) side of the updraft, enhancing (suppressing) storm inflow into the cyclonic (anticyclonic) vortex. In terms of the splitting storms, this enhances the cyclonic, right-moving storm and inhibits the anticyclonic, left moving storm which results in the predominance of cyclonically rotating, rightward moving storms. When significant turning of the shear vector is present, the initial storm may be so dominated by the cyclonically rotating flank that the storm does not split and no left mover is produced (e.g., Lilly 1983; Klemp 1987; Davies-Jones et al. 2001).

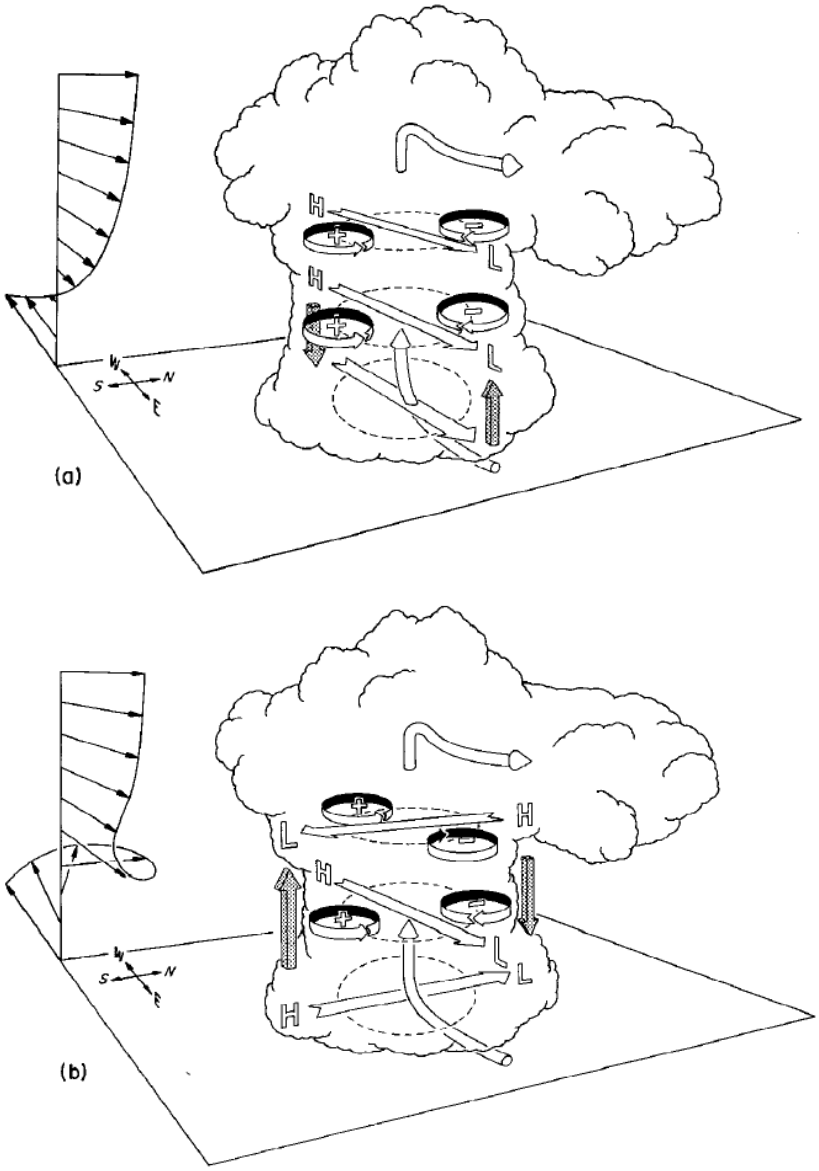


Figure 2.11: Pressure perturbation distribution resulting from updraft interaction with environmental wind shear. (*a*) Unidirectional wind shear with high and low pressure perturbation located in the same position throughout the vertical causing an upward (downward) directed perturbation pressure gradient force on the downshear (upshear) side. (*b*) Wind shear vector rotating clockwise with height with high and low pressure perturbations also rotating clockwise with height resulting in an upward (downward) directed perturbation pressure gradient force on the right (left) flank in the region of cyclonic (anticyclonic) vertical vorticity. Flat (shaded) arrows represent horizontal (vertical) perturbation pressure gradient force with regions of high (H) and low (L) perturbation pressure labeled accordingly. Refer to Figure 2.10 for all other storm feature and symbol definitions. [From Klemp (1987).]

Figure 2.12 provides a numerically simulated example of storm evolution differences between an environment with a rotating wind shear vector compared to unidirectional shear. The unidirectional shear example generates mirror image right moving and left moving storms while the rotating shear vector case contains a dominant rightward propagating storm with a well-defined low-level hook echo structure.

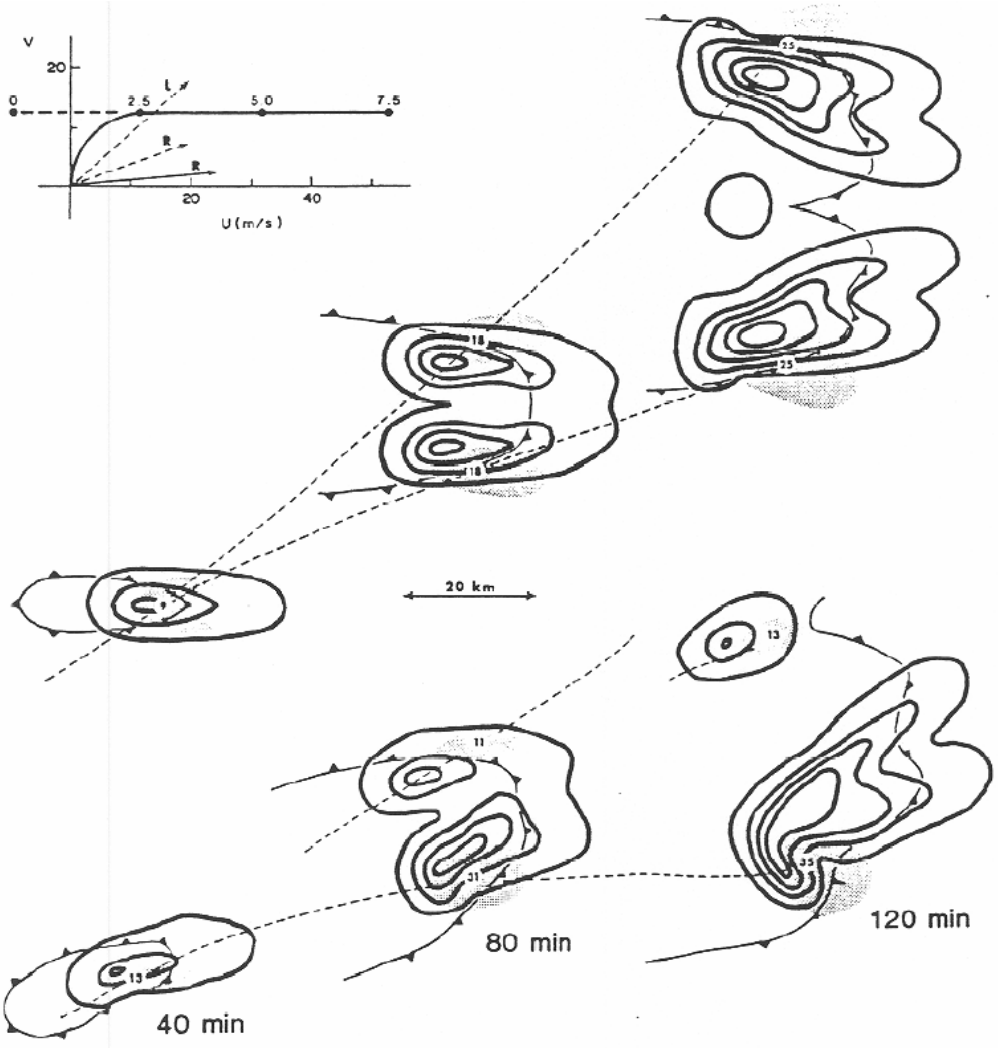


Figure 2.12: Horizontal cross section of the evolution of numerically simulated rainwater fields within storms generated in an environment having unidirectional wind shear (top) and with a rotating wind shear vector (bottom). Plotted is rainwater contours at 1.8 km (solid lines) at 2 g kg^{-1} intervals, cool outflow boundary at surface (cold front barbed line), updraft at 4.6 km (lightly shaded region) with max updraft velocity labeled in m s^{-1}, and storm track for each cell (dashed lines). Storms are given at 40 minute increments starting at 40 minutes into the simulation. Hodograph (upper left corner) shows vertical wind profile for each environment in which both contain speed shear from 2.5 km to 7.5 km and the unidirectional wind shear (rotating wind shear) case contains speed shear (directional shear) from the surface to 2.5 km as shown by the dashed (solid) line. Storm motion is also plotted on the hodograph corresponding to the left (L) and right (R) moving storms for the unidirectional wind shear case (dashed vector) and the right moving storm (R) for the rotating wind shear case (solid vector). [From Klemp (1987).]

2.2.2 Low-Level to Near Ground Rotation

After a rotating updraft becomes established at mid-levels, a supercell storm develops low-level rotation through processes different from those responsible for mid-level rotation. Using numerical simulations Klemp and Rotunno (1983) and Rotunno and Klemp (1985) identified a process capable of generating low-level rotation that was associated with the baroclinity along a supercell's FFD gust front. As precipitation forms in the storm's forward flank, evaporatively cooled air descends in the FFD and spreads out across the surface after it reaches the ground. The leading edge of the cool outflow (called the FFD gust front) advances to a quasi-steady position adjacent to, and parallel with, the low-level storm-relative inflow (Figure 2.13). The horizontal temperature and density gradient along the gust front baroclinically generates horizontal vorticity oriented towards the updraft. As the inflow air travels a considerable distance through this baroclinic zone it acquires horizontal vorticity which is subsequently tilted to the vertical and stretched while entering the updraft. The Klemp and Rotunno studies conclude that the resulting low-level cyclonic rotation is responsible for dynamically inducing lower pressure and a downward directed pressure gradient force (Figure 2.14(a)). This force generates the occlusion downdraft (discussed in previous section) and strengthens the RFD outflow causing a rapid occlusion of the FFD and RFD gust fronts and formation of a new updraft and center of low-level rotation farther to the east (Figure 2.14(b)).

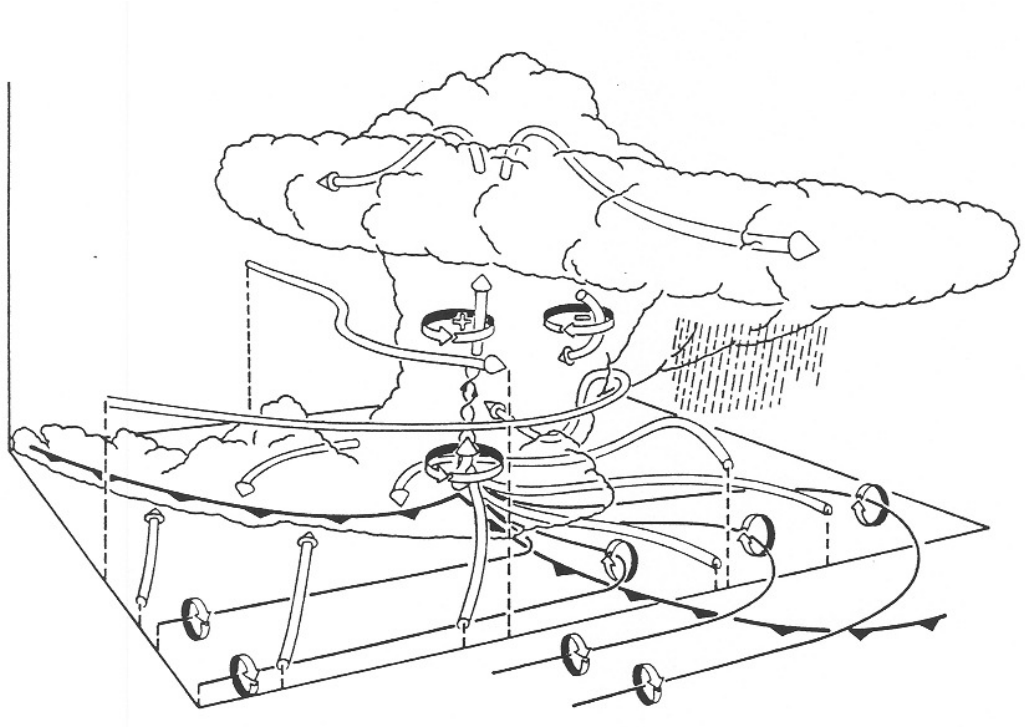


Figure 2.13: Three-dimensional schematic of a numerically simulated supercell storm in unidirectional (westerly) wind shear as viewed from the southeast (see Figure 2.10 for cardinal directions) during a period of intensifying low-level rotation. Cylindrical arrows represent storm-relative flow, black lines correspond to low-level vortex lines with direction of horizontal and vertical rotation given by circular, ribbon arrows. Cold front barbed line represents outflow boundaries from the FFD (east of storm) and RFD (south of storm). [From Klemp (1987).]

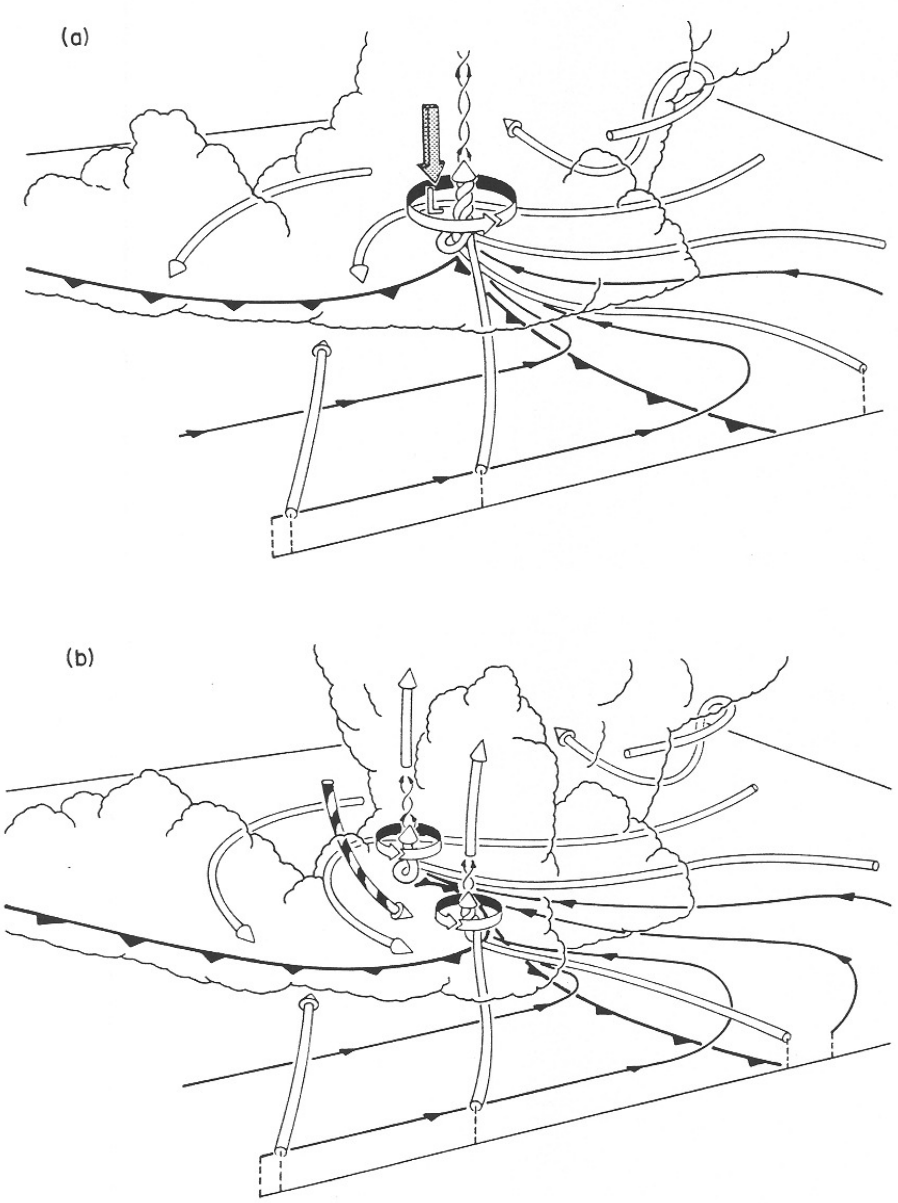


Figure 2.14: Enlarged view of low-level features from numerically simulated storm shown in Figure 2.13 corresponding with (*a*) time period of intensifying low-level rotation (same time depicted in Figure 2.13) as dynamically induced low pressure generates a downward directed pressure gradient force and (*b*) about 10 minutes later when the occlusion downdraft forms, helping to intensify RFD outflow at the surface which cuts off the original center of rotation and initiates a new center of rotation to the east near intersection of the occluded FFD and RFD gust fronts. All symbols and features are the same as Figure 2.13 except for the addition of a representation of the downward directed pressure gradient force (solid arrow), the intensifying RFD (striped cylindrical arrow), and the orientation of the low-level vortex lines (shown as small arrows on the vortex lines). [From Klemp (1987).]

Thus far, observational studies have been unable to confirm that the baroclinic zone along a supercell's FFD gust front is an important factor in generating strong low-level rotation (see review of FFD observational studies in Shabbott and Markowski 2006). Therefore additional observational studies will be needed to either confirm or disprove this hypothesis.

Even if low-level rotation can be generated in the previously explained process it is still insufficient in explaining the generation of significant rotation close to the ground. Davies-Jones (1982a,b) concluded that an "in, up, and out" circulation driven by forces aloft would be unable to produce vertical vorticity close to the ground in a sheared environment with minimal background vertical vorticity.[8] Parcels entering the updraft at very low levels do not acquire significant vertical vorticity until they have ascended 1-2 km because horizontal vorticity must first be tilted to the vertical and then subsequently stretched by the updraft. Generating significant vertical vorticity near the ground would require an abrupt upward turning of the streamlines (and coinciding vortex lines), strong pressure gradients, and large vertical velocities adjacent to the ground. These conditions may exist once a tornado has been established due to the strong vertical velocity horizontal gradients associated with the tornado updraft but they can not be used to explain the initiation of the strong rotation itself.

[8] This conclusion assumes eddies are too weak to transport vertical vorticity downward against the flow which has been subsequently verified (e.g., Rotunno and Klemp 1985; Walko 1993).

Therefore, Davies-Jones (1982a,b) proposed that the RFD was responsible for producing strong rotation at the surface. He excluded baroclinic vorticity generation and found that vertical vorticity could be generated when horizontal vorticity (generated by westerly shear) was tilted downward by the downdraft, resulting in a north-south oriented cyclonic-anticyclonic vortex pair (see illustration in Figure 2.17(a) in upcoming discussion). With the RFD located in a storm's rear flank, generally south of the updraft (see Figure 2.6), the vertically oriented cyclonic vortex is positioned so positive vertical vorticity can flow out from the downdraft across the surface and be ingested into the updraft where it is enhanced by vertical stretching. Walko (1993) replicated this process in a simplified numerical simulation in which a heat sink (representing an RFD) was placed southwest of a heat source (representing an updraft) within westerly shear. An intense vortex was generated at low-levels along the left edge of the downdraft-induced cold pool underneath the updraft.

Davies-Jones and Brooks (1993) analyzed the effect baroclinity might have on descending air parcels in the RFD and found that vertical vorticity of the parcels can transition from anticyclonic to cyclonic vorticity while still descending. Figure 2.15 illustrates this process. As air descends, vortex lines are turned downward but with less inclination than the trajectories of the flow because of the addition of southward directed horizontal vorticity that is being continuously generated by the baroclinity within the hook echo. When the descending air approaches the surface and begins to turn the vortex lines back towards the horizontal, the addition of the baroclinic vorticity allows the vortex lines to actually tilt into the vertical before the air reaches

the surface. The cyclonic vorticity then passes from the downdraft to the updraft where it is amplified further through stretching.

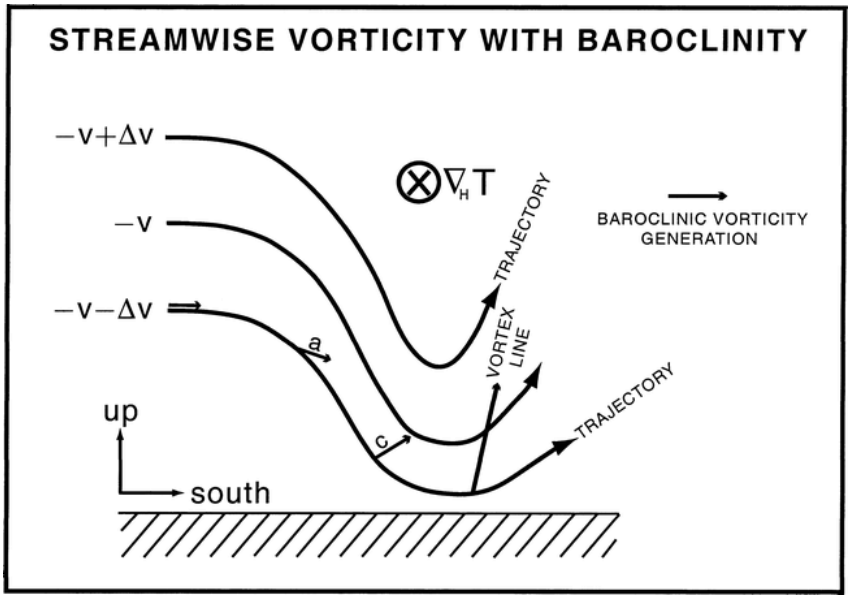


Figure 2.15: Illustration showing how cyclonic vertical vorticity can be generated even before descending air reaches the ground when baroclinic effects within the hook echo region are included. The environment contains horizontal streamwise vorticity that is tilted downward in the RFD. Air parcel trajectories are plotted as curved arrows while vortex lines are given as straight arrows. The temperature gradient direction is into the page. The baroclinic vorticity generation direction is to the right (south). [From Davies-Jones and Brooks (1993).]

Trajectory analyses in subsequent studies indicated that air parcels entering the region of strong low-level rotation *did* pass through the RFD and transition from anticyclonic to cyclonic vertical vorticity but this transition did not occur until *after* the parcels started their ascent into the updraft (e.g., Trapp and Fiedler 1995; Wicker and Wilhelmson 1995; Adlerman et al. 1999). The difference between these results and those of Davies-Jones and Brooks (1993) are considered minor due to the fact

that each study identifies a time tendency of increasing positive vertical vorticity for parcels descending through the RFD (Markowski 2002).

A new hypothesis on the origin of near-ground rotation within a supercell storm was recently introduced by Straka et al. (2007). They examined supercell observational studies and found the same general structure in most storm rear flanks (Figure 2.16). Assuming an eastward-moving storm, this structure consists of a strong vertical velocity gradient along an eastward-moving gust front associated with the intersecting rear-flank downdraft outflow and storm updraft inflow region. Vortex lines along the gust front rise with positive vertical vorticity on the northern (left) side of the rear flank downdraft outflow, become horizontally oriented moving southward along the gust front and then extend downward with negative vertical vorticity on the southern (right) side of the outflow. Therefore cyclonic vorticity is found on the left side of the eastward-moving RFD gust front and anticyclonic vorticity on the right side.

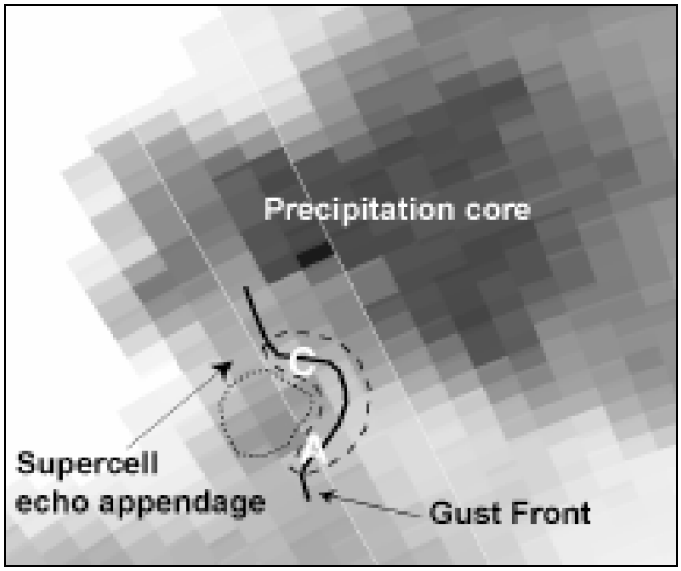


Figure 2.16: Schematic example of an arching vortex line associated with a supercell's RFD gust front. Gray shading represents radar reflectivity with more intense regions shaded darker. The black solid line represents a vortex line that is being lifted vertically between regions of cyclonic (C) and anticyclonic (A) vertical vorticity (i.e., line should be regarded as extending vertically above the two dimensional background plot). Low-level updrafts are outlined with a dashed line while low-level downdrafts are outlined in a dotted line. Other supercell features are labeled accordingly. [From Straka et al. (2007).]

While Davies-Jones (1982a,b) and Walko (1993) attribute this cyclonic-anticyclonic vertical vorticity couplet to downward tilting of environmental horizontal vorticity (Figure 2.17(a)), Straka et al. (2007) argue that the vortex lines actually tilt upwards and that the vertical vorticity couplet is located near the updraft-downdraft interface along the RFD gust front. According to their hypothesis, this vortex line arching structure can be generated in either one of two processes. The first, shown in Figure 2.17(b), requires a downdraft containing stronger downward motion at its center capable of generating a toroidal circulation on the outer flanks of the downdraft (time "t1"). With time, the vortex rings are advected downward with the downdraft air and advected towards the updraft by the horizontal wind where they

subsequently tilt upwards (time “t2”). As the leading edges of the vortex rings enter the updraft (time “t3”), they are rapidly tilted upwards resulting in an arching vortex line pattern with cyclonic vertical vorticity on the north (left) side of the updraft-downdraft interface and anticyclonic vertical vorticity on the south (right) side.

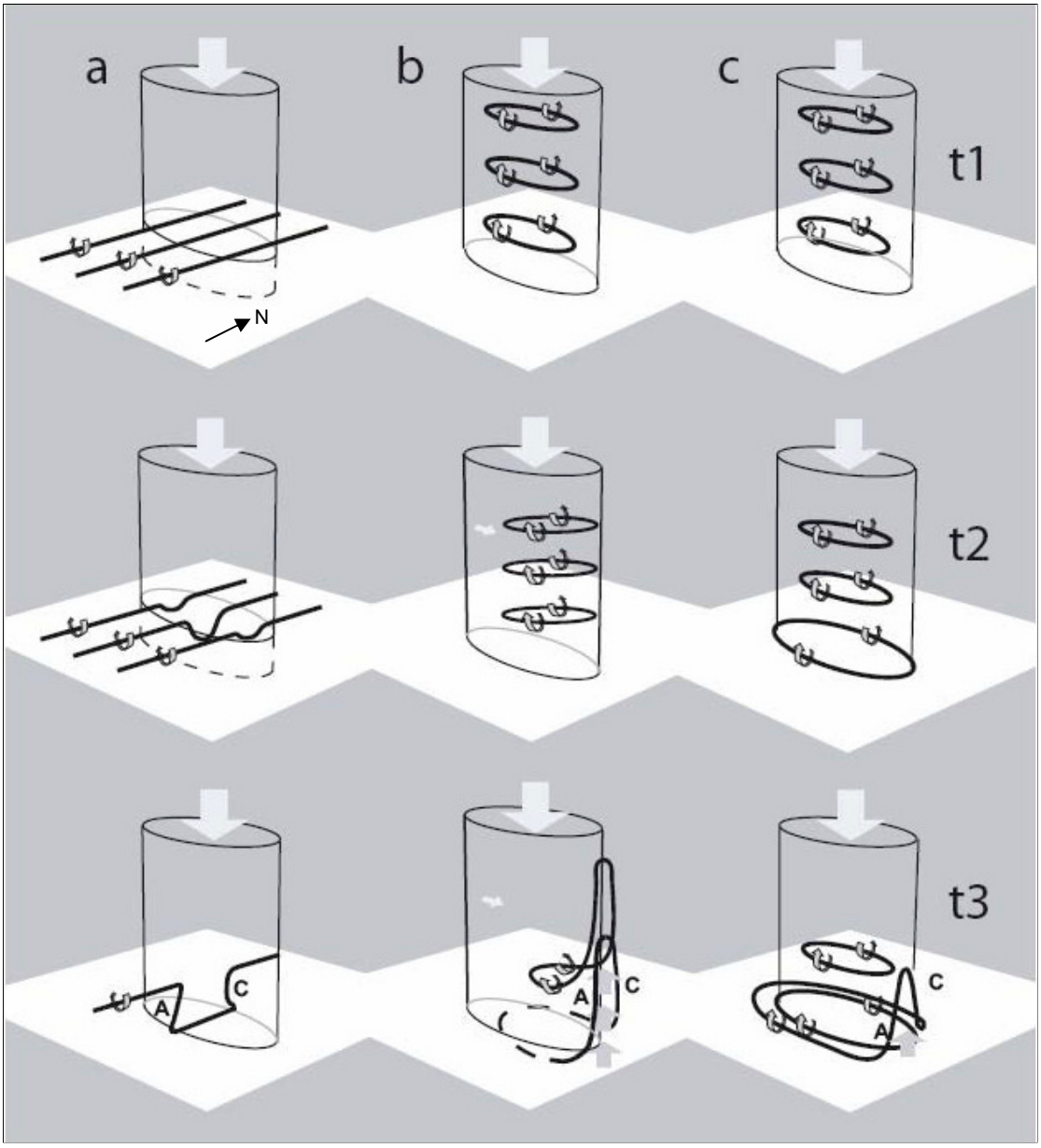


Figure 2.17: Schematics of three hypotheses for the RFD's role in generating significant vertical vorticity and rotation near the surface. (*a*) The hypothesis of Davies Jones (1982a,b) and Walko (1993) in which the RFD tilts existing environmental horizontal vorticity downward to create a cyclonic and anticyclonic vertical vorticity couplet. And the hypotheses of Straka et al. (2007) in which the RFD generates its own toroidal circulation that descends with the downdraft and (*b*) is advected towards the updraft by the horizontal wind or (*c*) expands as it approaches the surface, both processes resulting in the upward tilting of vortex lines with a cyclonic and anticyclonic couplet. Elongated cylinders represent an RFD and vortex lines are given as solid black lines with orientation show with circular ribbon arrows. Cyclonic (C) and Anticyclonic (A) vertical vorticity centers are labeled accordingly and time moves forward from t1 to t3. [From Straka et al. (2007).]

The other process capable of producing this same arching vortex structure (Figure 2.17(c)) begins with the same downdraft-induced toroidal circulation (time "t1") but this time only the downward advection of the vortex rings is required. A rear-to-front horizontal flow is not necessary. As vortex rings are advected downward they also spread out in the horizontal as they approach the ground, in the same manner as the downdraft outflow (time "t2"). Therefore the vortex rings near the surface will advance into the updraft and be rapidly tilted into the vertical generating the arching vortex line pattern and corresponding vertical vorticity couplet (time "t3"). The increased intensity of the cyclonic vortex compared to the anticyclonic vortex is attributed to its close proximity to the strongest portion of the low-level updraft which increases its vertical stretching. The anticyclonic vortex may also be in a region of updraft but it is usually located somewhere along the trailing RFD gust front a greater distance away from the central, dominant updraft (e.g., Davies-Jones et al. 2001; Markowski 2002). Both of the hypothesized processes of Straka et al. (2007) therefore are capable of explaining the low-level rotation characteristics in a supercell's rear-flank.

CHAPTER 3

METHODOLOGY

This chapter describes the methodology for identifying interrelated, repeated patterns among storm features that herald the development of strong low-level rotation within numerically simulated supercell storms. Section 3.1 describes the numerical simulations generated for this study including an overview of the model configuration and parameter space design. Section 3.2 describes the preliminary method for identifying relevant storm features and its limitations, and Section 3.3 describes the final methodology, including a discussion about the automated data mining technique.

3.1 Numerical Simulations

A numerical model is used to generate a large number of simulated supercell storms covering a spectrum of supercell environments. This section describes the numerical model configuration as well as the parameter space design.

3.1.1 Numerical Model Configuration

The numerical simulations are produced using Version 5.2 of the Advanced Regional Prediction System (ARPS), a three-dimensional, nonhydrostatic model developed for storm scale numerical weather prediction (Xue et al. 2000, 2001, 2003). The computational grid has uniform horizontal spacing of 0.5 km within a 100 x 100 x 19.5 km^3 domain. The vertical grid is stretched via a hyperbolic tangent

function over 50 levels, with the grid spacing varying from 50 m at the surface to 750 m near the top of the domain. The lowest scalar gridpoint therefore is located at a height of 25 m above ground, which hereafter is identified for convenience as either the near-ground or surface level. A horizontally homogeneous environment is constructed from an input sounding (discussed in Section 3.1.2) and the environment is then perturbed by an ellipsoidal thermal bubble to initiate convection. In an attempt to keep the dominant storms sufficiently far away from the lateral boundaries throughout the entire simulation, automated domain translation is applied roughly equal to the vector horizontal motion of the 0-6 km vertical velocity averaged over five minute intervals. This translation is weighted by the vertical velocity magnitude and therefore preferentially moves the domain in the direction of the most intense simulated updraft regions. The model is integrated for three hours with history files saved every 30 seconds. Approximately 160 hours of CPU time is required to run a single simulation with roughly 118 of those hours dedicated to I/O. The history files range in size from 10 to 70 MB resulting in an average simulation data dump size of 21 GB. A summary of key model parameters is shown in Table 3.1.

Table 3.1: Key ARPS model parameters used in the numerical simulations.

Parameter	Symbol	Value
Horizontal grid spacing	Δx, Δy	0.5 km
Vertical grid spacing	Δz	50 m $\leq \Delta z \leq$ 750 m
Large time step	Δt	1.0 s
Small time step	$\Delta \tau$	0.5 s
Initial thermal perturbation:		
Magnitude	$\Delta \theta$	4.0 K
Horizontal Radius	x_r, y_r	10.0 km
Vertical Radius	z_r	1.5 km
Height of center (AGL)	z_c	1.5 km
Coriolis parameter	f	0.0 s^{-1}
Nondimensional surface drag coefficient	C_d	0.0
Divergence damping coefficient	a	0.05
Fourth-order horizontal mixing coefficient	$K_{4,h}$	3.125 x 10^8 m^4s^{-1}
Fourth-order vertical mixing coefficient	$K_{4,v}$	3.125 x $10^3 \leq K_{4,v} \leq$ 1.58x10^8 m^4s^{-1}
Microphysics	Ice microphysics	
Grid stretching function	Hyperbolic tangent	
Lateral boundary conditions	Radiation	
Top boundary conditions	Rigid with Rayleigh sponge layer	
Horizontal advection	Fourth-order	
Vertical advection	Second-order	
Turbulence parameterization	Anisotropic 1.5-order TKE	

3.1.2 Parameter Space Design

Analytic soundings are generated to specify the horizontally homogeneous background environment used in the numerical simulations. The thermodynamic profiles of these soundings are based upon those used by Weisman and Klemp (1982), which have been the predominant profiles used in previous numerical studies (e.g., Weisman and Klemp 1984; Brooks and Wilhelmson 1993; Droegemeier et al. 1993; Adlerman et al. 1999; Adlerman and Droegemeier 2003, 2005) despite the notably tropical character of the moisture profile. A more advanced technique for

producing a thermodynamic profile is available (McCaul and Weisman 2001), but its complexities are considered unnecessary for the goals of this study.[9]

Figure 3.1 illustrates the Weisman and Klemp (1982) thermodynamic profile variations, in which adjustment of the surface mixing ratios shifts the lifted condensation level and level of free convection to different heights, thus changing the buoyancy available to a rising parcel of air. This available buoyancy is typically measured by the convective available potential energy (CAPE), which is defined as the vertically integrated positive thermal buoyancy of an air parcel (Weisman and Klemp 1982) and is calculated using

$$CAPE = g \int \frac{\theta(z)_{parcel} - \theta(z)_{env}}{\theta(z)_{env}}, \tag{3.1}$$

where g is the acceleration due to gravity, $\theta(z)_{parcel}$ is the potential temperature of the rising parcel of air by dry and then moist-adiabatic ascent, and $\theta(z)_{env}$ is the potential temperature of the environment. The integration is performed only within vertical levels where the parcel is positively buoyant (i.e., $\theta(z)_{parcel} > \theta(z)_{env}$).

For this study, five thermodynamic profile variations with surface mixing ratios of 13, 14, 15, 16 and 17 g kg^{-1} are used, resulting in CAPE ranging from 1284-3407 J kg^{-1}, respectively (or 686-2338 J kg^{-1} when virtual temperature correction and

[9] The thermodynamic profiles from McCaul and Weisman (2001) allow alterations of many individual parameters such as CAPE, lifting condensation level, level of free convection, and the shape of the buoyancy profile whereas in Weisman and Klemp (1982), only the surface mixing ratios were variable with no option to alter separately the other thermodynamic parameters. However, because the present study is concerned with analyzing storm features and not relationships between storm behavior and environmental parameters, the latter, more general, method is considered a suitable choice (Eugene McCaul Jr., 2006, personal communication). Widespread use of the Weisman and Klemp (1982) profiles in previous studies also makes its use valuable here, although limitations are noted such as its nearly tropical structure with no midlevel dry layer or low-level inversion.

water loading are applied). This represents a range of environments from relatively small to large available buoyancy.

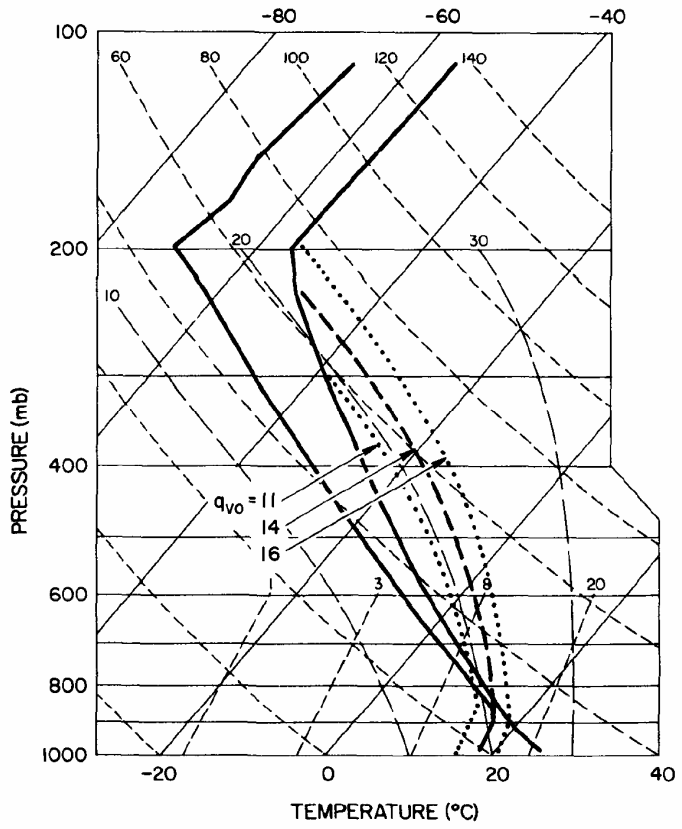


Figure 3.1: Skew-T/Log-P diagram with isotherms (tilted solid lines), dry adiabats (short dashed lines) and moist adiabats (long dashed lines) depicting the Weisman and Klemp (1982) method for constructing thermodynamic profiles. Temperature (heavy solid line on right side) and dew point temperature (heavy solid line on left side) profiles are plotted along with ascending parcel from the surface for surface layer mixing ratios (q_{vo}) of 14 g kg^{-1} (heavy dashed line) and 11 and 16 g kg^{-1} (heavy dotted lines as shown). [From Weisman and Klemp (1982).]

Applied to the five thermodynamic profiles are a set of vertical wind profiles appropriately chosen based upon previous parameter studies of simulated supercell storms that used a broad range of hodograph shapes and shear magnitudes (e.g., Droegemeier et al. 1993; Adlerman and Droegemeier 2005). The chosen wind

profiles are summarized in Figure 3.2 and range from half circle hodographs of radius 8-30 m s^{-1} with turning depth 4-10 km to quarter circle hodographs of radius 8-35 m s^{-1} with turning depth 1-5 km to quarter circle with tail hodographs of radius 10-25 m s^{-1}, turning depth 1-3 km and tail length through 10 km of 10-40 m s^{-1}. Straight, three-quarter and full circle hodographs are not included in this study because the half, quarter and quarter circle with tail hodographs are thought to more closely resemble observed tornadic supercell proximity hodographs (e.g., Maddox 1976).

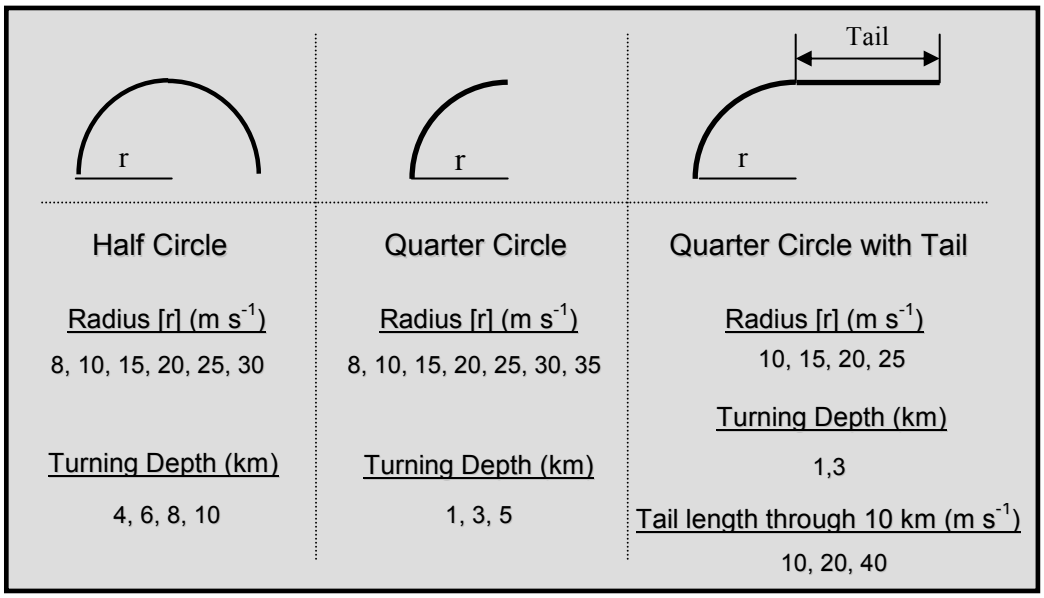


Figure 3.2: Schematic summary of hodograph shapes and shear magnitudes used in this study.

The set of soundings resulting from the combined wind and thermodynamic profiles are further narrowed using a specified cross-section of bulk Richardson number (BRN) and 0-3 km storm relative helicity (SRH) to produce a final parameter space thought to most likely represent appropriate environments for supercell storms.

The BRN provides a general representation of the effects of buoyant instability and vertical wind shear and is defined by

$$BRN = \frac{CAPE}{0.5\left(\overline{u}^2 + \overline{v}^2\right)}, \tag{3.2}$$

where CAPE is defined by equation 3.1 and $\overline{u}$ and $\overline{v}$ are the components of the difference between the density-weighted mean winds over the lowest 6 km and the lowest 500 m and therefore represent a measure of the ambient environmental wind speed shear. The BRN can be used, in general, to categorize storm type where modeled supercell storms are likely to occur when $5 \leq BRN \leq 50$ and modeled multicell storms are likely to occur when $BRN > 35$ (Weisman and Klemp 1982, 1986). Due to the obvious overlap between supercell and multicell storms as well as the inability of the BRN to account for hodograph shape and magnitude of storm-relative winds (e.g., Droegemeier et al. 1993), the parameter SRH is also used to narrow the final parameter space.

SRH provides a good measure of the rotational potential of an updraft and is defined as minus twice the signed area swept out by the storm-relative wind vector between two levels on a hodograph (Davies-Jones et al. 1990). In an observational study using proximity soundings Davies-Jones et al. (1990) found 0-3 km SRH of 157 $m^2\ s^{-2}$ as an approximate lower threshold for mesocyclone development and SRH of 280, 330 and 530 $m^2\ s^{-2}$ for weak, strong and violent tornadoes, respectively. By comparison, the numerical modeling study of Droegemeier et al. (1993) estimated 0-3

km SRH of 250 m^2 s^{-2} as a suitable threshold for the development of supercells containing persistent rotating updrafts.

The storm motion vector required in the SRH calculation is estimated using an adjustment to the 0-6 km mean wind as proposed by Davies and Johns (1993). Their study analyzed a large data set (242 cases) of strong and violent mesocyclone-induced tornadoes from Johns et al. (1990) and found that proximity sounding-derived 0-6 km mean wind speed > 30 kt (15.5 m s^{-1}) resulted in an average storm motion 20° to the right and 85% the strength of the 0-6 km mean wind. With sounding-derived 0-6 km mean wind speed ≤ 30 kt, average storm motion was 30° to the right and 75% the strength of the 0-6 km mean wind. Adlerman and Droegemeier (2005) applied this method in their numerical modeling study for estimating SRH and found it to be a reasonable approximation to SRH values derived from actual simulated supercell storm motions.

The final parameter space chosen for this study therefore consists of initial environments with BRN ≤ 50 and SRH ≤ 750 m^2 s^{-2}. This combination has the potential to generate a large variety of numerically simulated supercell storms, some capable of producing strong low-level rotation. Although some initial environments may contain unrealistically large CAPE and wind shear combinations or have the potential to generate non-supercell storms, this parameter space has the advantage of providing a broad spectrum of supercell storm environments.

Figure 3.3 shows the 256 individual model environmental profiles that fall within the aforementioned parameter space requirements, with the diamond, square

and triangle symbols representing hodographs having shapes of a half circle, quarter circle, and quarter circle with tail, respectively. Each of the environments also is listed in table format in Appendix A, along with their derived parameters of CAPE (using both T only and T_v with water loading)[10], BRN shear (denominator of equation 3.2), BRN (using both versions of CAPE), and SRH (0-1 km and 0-3 km). The naming scheme for the environments listed in Appendix A and used throughout this study consists of the Weisman and Klemp thermodynamic profile surface mixing ratio in g kg^{-1}, listed first as either WK13, 14, 15, 16, 17, followed by the shape of the hodograph as either a half circle (half), quarter circle (qtr) or quarter circle with tail (qtr_tail). Next is the radius of the hodograph in m s^{-1} leading with the letter "r" (i.e., 8, 10, 15, 20, 25, 30, 35), followed by the tail length through 10 km, for the quarter tail hodographs only, in m s^{-1} leading with the letter "t" (i.e., t10, t20, t40). Lastly, the turning depth, in km, of the half and quarter circle portions of each hodograph is listed (i.e., 1, 3, 4, 5, 6, 8, 10 km).

For example, the experiment identifier "WK14_half_r15_4km" refers to the environment characterized by a 14 g kg^{-1} surface mixing ratio and a half circle hodograph of radius 15 m s^{-1} that turns through 4 km. An example of a more complex experiment identifier is "WK14_qtr_tail_r15_t20_3km" which refers to the environment characterized by a 14 g kg^{-1} surface mixing ratio and a quarter circle

[10] CAPE is calculated from within the model and therefore includes the effects of the vertical grid which discretizes the temperature and moisture profiles. This will only minimally affect CAPE values compared to those calculated directly from the analytic equations used to create the soundings because of the relatively small vertical grid spacing used, especially at lower levels.

hodograph of radius 15 m s^{-1} that turns through 3 km with a tail of length 20 m s^{-1} from 3 km through 10 km.

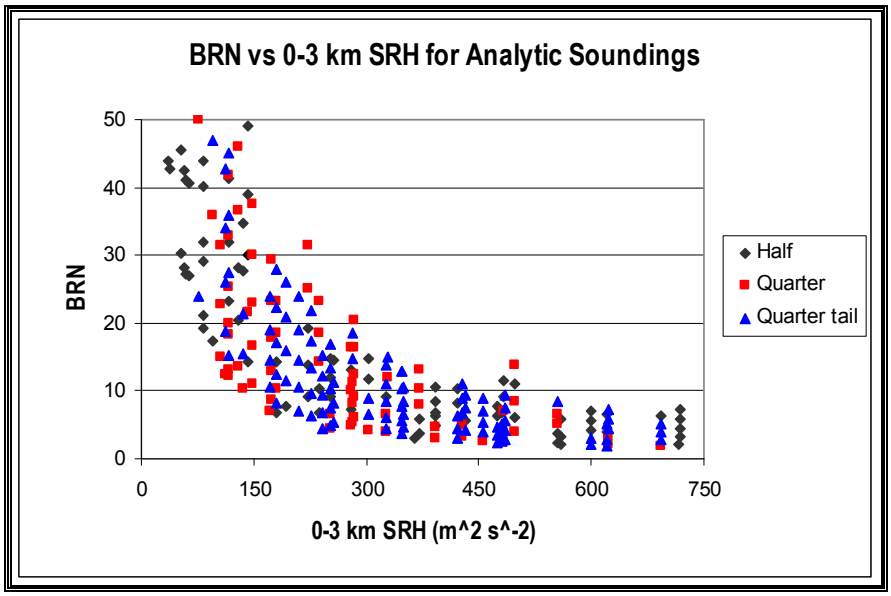


Figure 3.3: Parameter space used to specify the horizontally homogeneous initial model environment for the storm simulations. Analytic soundings contain BRN ≤ 50 and 0-3 km SRH ≤ 750 m^2 s^{-2}. BRN is calculated using CAPE with the virtual temperature correction and water loading and 0-3 km SRH is calculated using the method of Davies and Johns (1993). The 256 analytic soundings are identified by the shape of their hodograph as either a half circle (black diamond), quarter circle (red square) or quarter circle with tail (blue triangle).

3.2 Preliminary Data Mining Methodology

To achieve the principal goal of this research, namely, identifying interrelated, repeatable patterns of behavior among storm features, a technique was needed to search the large number of numerical simulations to identify relevant supercell features, and their mutual relationships in time and space, influential in the development of strong low-level rotation. A preliminary approach utilizing a high-

level feature recognition technique was developed but deemed too complex for current mining capabilities. Nevertheless, it is described here, along with the limitations restricting its application, to set the context for describing the methodology eventually chosen, which is a simplified version of the original approach. Advances in spatio-temporal data mining eventually should provide a framework for applying the method initially crafted.

3.2.1 High-Level Feature Recognition Technique

The approach initially pursued to identify important storm features involved identifying, tracking and interrelating three-dimensional high-level features associated with a typical simulated supercell storm. These high-level features, which henceforth also will be referred to as objects, include updraft, downdraft (rear flank, forward flank and occlusion), rainwater, hail core, gust front (rear flank and forward flank), mesocyclone (mid- and low-level), bounded weak echo region, hook-like signature and others. Analyzing the evolution of each of these objects was thought to be useful in defining a sequence of events among them which herald the development of strong low-level rotation. A brief overview of how this approach might be carried out is given below, followed by a discussion in the next section of why it is difficult to implement and why, after preliminary work, an alternate approach was adopted.

The first step in applying the object-based data mining strategy requires identifying each of these objects in the model dependent variable fields (e.g., wind components) and associated derived quantities (e.g., vorticity, divergence). This

requires establishing a set of criteria to define each object in an appropriate manner. It then is necessary to examine how each object evolves in time and how all objects interrelate to one another. To accomplish this daunting task, each object is assigned a number of attributes that also need to be tracked. These attributes include physical size, overall 3-D shape and location within a storm as well as object intensity or amplitude.

After identifying all objects and their attributes, each must be quantitatively interrelated, both temporally and spatially. Once this has been done for all storms, the various relationships and sequences can be compared to identify whether a unique set occurs sufficiently often, and in the same general way, so as to become practical predictors of strong low-level rotation.

3.2.2 Limitations of the High-Level Feature Recognition Technique

Although the approach just outlined is conceptually straightforward and would appear to be promising, its practical implementation is quite difficult. The ability to first identify each of the relevant objects within the model output is an exceedingly difficult task. Basic objects (i.e., updraft, rainwater and hail core) are relatively easy to identify because they can be distinguished by assigning thresholds to dependent model fields such as rainwater mixing ratio, upward vertical velocity, etc. Basic objects originating from the same meteorological fields but existing at different altitudes also can be identified relatively easily. For example, the mid- and low-level mesocyclone could be identified as vertical vorticity $\geq 0.01\ s^{-1}$ over some

minimum depth and area extending from 3 to 9 km in altitude for the mid-level mesocyclone, and from the near surface to 2 or 3 km altitude for low-level mesocyclone.

The difficulty arises when objects originating from the same meteorological fields exist at the same altitudes. For example, various downdrafts within a supercell (i.e., rear flank, forward flank and occlusion) can coexist at the same time and height and in some cases can physically overlap, e.g., the rear flank and occlusion downdrafts (e.g., Wicker and Wilhelmson 1995). Because these downdrafts are associated with different dynamical mechanisms and therefore may play different roles in the development of strong low-level rotation, it is advantageous to separate them and track their individual characteristics. A strict set of criteria therefore is required to assure proper identification of these downdrafts. Although it may be possible to create a simplified version of such a classification scheme, difficulties arising in the ensuing steps make further refinement ineffectual.

The second major difficulty with the high-level feature approach concerns quantifying and tracking the attributes of each object. If it *were* possible to properly identify all relevant objects, then their size could easily be distinguished from their 3-D volume and/or 2-D area at specified height levels; however, quantifying their location within a storm, and their 3-D shape and intensity, is more difficult.

The location of an object may be specified in relation to the center of the storm and storm movement, the former possibly associated with the dominant updraft. Specifying a central location for each object could be done by using either a

mean weighted average, 3-D maximum value or 2-D maximum value at a specified altitude for the meteorological quantity used to identify it. For example, a location attribute for a specific object could take the following form: Object X is centered 1 km away from the center of storm A and 140° away from storm A's motion vector. However, keeping a constant storm center anchor point throughout storm lifetime is quite difficult, as is maintaining a center point for each object, because storms evolve with time, split or merge with other storms, etc.

Specifying the structural shape of an object poses problems as well because it is difficult to effectively quantify and store 3-D shapes and their variations through time in a way that is useful for analysis. For example, the typical hook echo shape of the low-level rainwater or derived reflectivity fields within a supercell is relatively simple to visualize by the human observer, but quantifying that shape in a computational manner and tracking its variations with time is extremely difficult.

Quantifying the intensity or amplitude of an object also is difficult, mainly owing to the various options available. A maximum value or mean value can be used from either within the entire 3-D volume of an object or at a specific horizontal level. The size of an object also could represent a partial measure of intensity as well as the rate at which the object is changing with time. Again, an overall attribute identification and tracking scheme could be attempted but due to further difficulties in the required data mining approach described below, such a strategy was not pursued.

The most important difficulty of the high-level feature approach is the requirement of a data mining procedure capable of searching through all object

attributes for each simulated storm and identifying specific object interrelationships deemed influential to the development of strong low-level rotation. Not only would the data mining method need to search through temporal variations within the attributes of each object, such as those that can be represented by a single value at each time step (i.e., size, volume, intensity), but the method also would need to sort through spatial changes such as how the objects are moving and changing shape within the storm and how they are evolving in relationship with one another. The significant complexities associated with this type of data mining procedure made it necessary to reformulate and simplify the overall methodology of this study – creating an approach that is still capable of achieving the primary goal of identifying precursors to the development of strong low-level rotation within supercell storms but within reach of data mining approaches currently available. The next section discusses the revised methodology.

3.3 Methodology Employed in this Study

The methodology actually used in this study is conceptually similar to the high-level feature recognition method described in Section 3.2, except that it bypasses the major limiting requirement of a data mining technique capable of both spatially and temporally interrelating 3-D objects. The revised methodology scales 3-D objects down to a more tangible metadata set of maximum and minimum meteorological quantities at individual grid points while still maintaining the ability

to identify influential precursors to strong low-level rotation within a simulated supercell storm.

The first section below describes the method by which the new technique extracts and defines meteorological quantities, and the second section explains the storm tracking scheme used. The third section then discusses the data mining method used to sort through all of the extracted quantities in search of influential precursors.

3.3.1 Meteorological Quantity Extraction Technique

Many important attributes of simulated supercell storms can be captured by extracting time series of maximum and minimum values of relevant meteorological variables. A schematic of the extraction process is shown in Figure 3.4. Because supercell storms typically contain a dominant updraft with strongest vertical velocities at mid-levels (e.g., Doswell and Burgess 1993; Rotunno 1993), the upward or positive vertical velocity (w) at 4 km altitude is used to identify each storm. A horizontal region at 4 km altitude containing $w \geq 15$ m s^{-1} over at least eight adjoining grid points[11] is used as a supercell storm indicator region (SIR) representing a strong, dominant updraft[12]. The outlining boundary of a single storm is defined by extending outward 5 km in all directions from its SIR. The entire vertical domain encompassed by this 5 km extension is identified as the region of interest for each storm in which meteorological quantities are extracted. The use of a 5 km extension around a

[11] Grid points are separated by $\Delta x = \Delta y = 500$ m. The threshold minimum of eight adjoining grid points can take on any horizontal shape.

[12] These storm indicators are tracked through time and must last for at least 30 minutes to qualify as true storm indicators. This process is discussed in Section 3.3.2.

supercell's mid-level updraft is deemed sufficient to contain its most dynamically important features (e.g., see Figure 2.6) while at the same time limiting the amount of overlap that is likely to occur between neighboring storms. This specified 3-D extraction region of interest hereafter is referred to either as the storm extraction region (SER_{5km}) or simply a storm, with the understanding that it is not intended to represent the entire numerically simulated storm but rather a defined region of interest in association with a specific simulated storm. The "5km" subscript is included to limit any potential confusion that might occur if this region were labeled SER, which is very similar to SIR. The example given in Figure 3.4 of this process is for an idealized horizontally circular SIR (highlighted red) and subsequent horizontally circular SER_{5km}. However, the SIR may take on any horizontal shape as long as it consists of 8 connected grid points, and therefore the 5 km horizontally extended SER_{5km} also can take on any number of shapes.

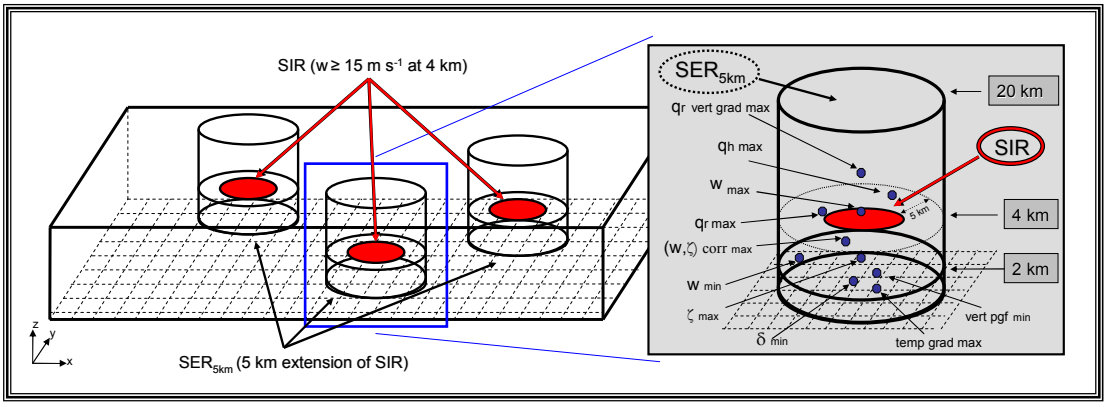


Figure 3.4: Schematic diagram of the meteorological quantity maximum and minimum value extraction process. A supercell storm indicator region (SIR) is identified having $w \geq 15$ m s^{-1} at 4 km altitude and then a 5 km extension of this region is defined, termed the storm extraction region (SER_{5km}), in which maximum and minimum quantities are extracted from various levels (as listed in Table 3.2).

Once the storm boundary (SER_{5km}) has been defined, the maximum and minimum values of 100 meteorological quantities are extracted from both above and below 2 km altitude as well as other locations within the SER_{5km} in each 30 second history file output interval throughout the lifetime of each defined storm. Table 3.2 contains a summary of the extracted quantities as well as the vertical level at which each is defined. Specifying quantities both above and below 2 km allows separation of low-level and mid- to upper-level processes. Vertical velocity at 4 km altitude also is extracted to assist with the storm tracking technique (see Section 3.3.2), and other quantities are extracted from the surface level to assist with defining regions of strong low-level rotation (see Section 3.3.3). Quantities involving a correlation have additional extraction regions from the surface to 8 km altitude and the surface to 12 km altitude to provide additional information corresponding with previous studies (e.g., Droegemeier et al., 1993). This extraction process therefore reduces the large amount of 3-D model output and derived meteorological fields to a single metadata set of 100 meteorological maximum and minimum quantity time series for each simulated storm.

Table 3.2: List of extracted meteorological quantities and their symbol or equation. Maximum (max) and/or minimum (min) identifiers are listed with each quantity as well as the region of extraction defined by the following: at the surface (sfc), from the surface to 2 km, 8 km or 12 km (sfc – 2 km, 8 km or 12km), above 2 km (2 km +) and at 4 km (4 km). Also, associated with the correlation terms are average values calculated from the surface to 8 km (avg sfc – 8 km).

Meteorological Quantity	*Symbol – Equation*
1.) vertical velocity (m s^{-1}) [(max,min); sfc – 2 km, 2 km +, 4 km]	w
2.) vertical velocity horizontal gradient (s^{-1}) [(max); sfc – 2 km, 2 km +]	$\nabla_h(w) = \sqrt{\left(\frac{\Delta w}{\Delta x}\right)^2 + \left(\frac{\Delta w}{\Delta y}\right)^2}$
3.) vertical vorticity (s^{-1}) [(max,min); sfc, sfc – 2 km, 2 km +]	$\zeta = \frac{\Delta v}{\Delta x} - \frac{\Delta u}{\Delta y}$
4.) rainwater mixing ratio (kg kg^{-1}) [(max); sfc – 2 km, 2 km +]	q_r
5.) rainwater mixing ratio horizontal gradient (kg kg^{-1} m^{-1}) [(max); sfc – 2 km, 2 km +]	$\nabla_h(q_r) = \sqrt{\left(\frac{\Delta q_r}{\Delta x}\right)^2 + \left(\frac{\Delta q_r}{\Delta y}\right)^2}$
6.) rainwater mixing ratio vertical gradient (kg kg^{-1} m^{-1}) [(max,min); sfc – 2 km, 2 km +]	$\nabla_v(q_r) = \left(\frac{\Delta q_r}{\Delta z}\right)$
7.) perturbation potential temperature (K) [(min); sfc – 2 km, 2 km +]	$\theta' = \theta - \overline{\theta}$
8.) potential temperature horizontal gradient (K m^{-1}) [(max); sfc – 2 km, 2 km +]	$\nabla_h(\theta) = \sqrt{\left(\frac{\Delta \theta}{\Delta x}\right)^2 + \left(\frac{\Delta \theta}{\Delta y}\right)^2}$
9.) pressure perturbation (Pa) [(min); sfc, sfc – 2 km, 2 km +]	$p' = p - \overline{p}$
10.) pressure perturbation vertical gradient force (m s^{-2}) [(max,min); sfc – 2 km, 2 km +]	$-\left(\frac{1}{\rho}\right)\left(\frac{\Delta p'}{\Delta z}\right)$
11.) horizontal divergence (s^{-1}) [(max,min); sfc – 2 km, 2 km +]	$\delta = \frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta y}$
12.) horizontal wind speed (m s^{-1}) [(max); sfc, sfc – 2 km, 2 km +]	$\overline{V}_h = \sqrt{u^2 + v^2}$
13.) hail mixing ratio (kg kg^{-1}) [(max); sfc – 2 km, 2 km +]	q_h
14.) hail mixing ratio horizontal gradient (kg kg^{-1} m^{-1}) [(max); sfc – 2 km, 2 km +]	$\nabla_h(q_h) = \sqrt{\left(\frac{\Delta q_h}{\Delta x}\right)^2 + \left(\frac{\Delta q_h}{\Delta y}\right)^2}$
15.) hail mixing ratio vertical gradient (kg kg^{-1} m^{-1}) [(max,min); sfc – 2 km, 2 km +]	$\nabla_v(q_h) = \left(\frac{\Delta q_h}{\Delta z}\right)$
16.) vertical stretching term (s^{-2}) [(max,min); sfc – 2 km, 2 km +]	$-\left(\frac{\Delta v}{\Delta x} - \frac{\Delta u}{\Delta y}\right)\left(\frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta y}\right)$

Table 3.2: Continued

Meteorological Quantity	*Symbol – Equation*
17.) tilting term (s^{-2}) [(max,min); sfc – 2 km, 2 km +]	$\left(\frac{\Delta w}{\Delta y}\frac{\Delta u}{\Delta z}\right)-\left(\frac{\Delta w}{\Delta x}\frac{\Delta v}{\Delta z}\right)$
18.) baroclinic vorticity generation vertical term (s^{-2}) [(max,min); sfc – 2 km, 2 km +]	$\left(\frac{1}{\rho^2}\right)\left(\frac{\Delta\rho}{\Delta x}\frac{\Delta p}{\Delta y}-\frac{\Delta\rho}{\Delta y}\frac{\Delta p}{\Delta x}\right)$
19.) baroclinic vorticity generation horizontal terms (s^{-2}) [(max,min); sfc - 2 km]	$\frac{1}{\rho^2}\left[\left(\frac{\Delta\rho}{\Delta y}\frac{\Delta p}{\Delta z}-\frac{\Delta\rho}{\Delta z}\frac{\Delta p}{\Delta y}\right)\hat{i}+\left(\frac{\Delta\rho}{\Delta z}\frac{\Delta p}{\Delta x}-\frac{\Delta\rho}{\Delta x}\frac{\Delta p}{\Delta z}\right)\hat{j}\right]$
20.) baroclinic vorticity generation horizontal (x,y magnitude) term (s^{-2}) [(max); sfc - 2 km]	$\left(\left[\frac{1}{\rho^2}\left(\frac{\Delta\rho}{\Delta y}\frac{\Delta p}{\Delta z}-\frac{\Delta\rho}{\Delta z}\frac{\Delta p}{\Delta y}\right)\right]^2+\left[\frac{1}{\rho^2}\left(\frac{\Delta\rho}{\Delta z}\frac{\Delta p}{\Delta x}-\frac{\Delta\rho}{\Delta x}\frac{\Delta p}{\Delta z}\right)\right]^2\right)^{\frac{1}{2}}$
21.) updraft and vertical vorticity correlation [(max,min); sfc - 2 km, 2 km +, avg sfc - 8 km, sfc - 12km]	$\frac{\sum\left(w_{ij}-\overline{w}_{(w>1)}\right)\left(\zeta_{ij}-\overline{\zeta}_{(w>1)}\right)}{\left(\sum\left(w_{ij}-\overline{w}_{(w>1)}\right)\right)^{1/2}\left(\sum\left(\zeta_{ij}-\overline{\zeta}_{(w>1)}\right)\right)^{1/2}}$ [See Appendix B]
22.) downdraft and vertical vorticity correlation [(max,min); sfc - 2 km, 2 km +, avg sfc - 8 km, sfc - 12km]	$\frac{\sum\left(w_{ij}-\overline{w}_{(w<-1)}\right)\left(\zeta_{ij}-\overline{\zeta}_{(w<-1)}\right)}{\left(\sum\left(w_{ij}-\overline{w}_{(w<-1)}\right)\right)^{1/2}\left(\sum\left(\zeta_{ij}-\overline{\zeta}_{(w<-1)}\right)\right)^{1/2}}$ [See Appendix B]
23.) radar reflectivity (dBZ) [(max); sfc - 2 km, 2 km +]	$ref = 10LOG_{10}\left(Z_{e,rain}+Z_{e,snow}+Z_{e,hail}\right)$ [See Appendix C]
24.) radar reflectivity horizontal gradient (dBZ m^{-1}) [(max); sfc - 2 km, 2 km +]	$\nabla_h(ref)=\sqrt{\left(\frac{\Delta ref}{\Delta x}\right)^2+\left(\frac{\Delta ref}{\Delta y}\right)^2}$
25.) radar reflectivity vertical gradient (dBZ m^{-1}) [(max,min); sfc - 2 km, 2 km +]	$\nabla_v(ref)=\left(\frac{\Delta ref}{\Delta z}\right)$
26.) horizontal Laplacian of radar reflectivity (dBZ m^{-2}) [(max); sfc - 2 km, 2 km +]	$\nabla^2 ref = \frac{ref_{i+1,j}-2ref_{i,j}+ref_{i-1,j}}{(\Delta x)^2}+\frac{ref_{i,j+1}-2ref_{i,j}+ref_{i,j-1}}{(\Delta y)^2}$
27.) updraft and horizontal Laplacian of radar reflectivity correlation [(max,min); sfc - 2 km, 2 km +, avg sfc - 8 km, sfc - 12 km]	$\frac{\sum\left(w_{ij}-\overline{w}_{(w>1)}\right)\left(ref_{ij}-\overline{ref}_{(w>1)}\right)}{\left(\sum\left(w_{ij}-\overline{w}_{(w>1)}\right)\right)^{1/2}\left(\sum\left(ref_{ij}-\overline{ref}_{(w>1)}\right)\right)^{1/2}}$ [See Appendix B]
28.) downdraft and horizontal Laplacian of radar reflectivity correlation [(max,min); sfc - 2 km, 2 km +, avg sfc - 8 km, sfc - 12 km]	$\frac{\sum\left(w_{ij}-\overline{w}_{(w<-1)}\right)\left(ref_{ij}-\overline{ref}_{(w<-1)}\right)}{\left(\sum\left(w_{ij}-\overline{w}_{(w<-1)}\right)\right)^{1/2}\left(\sum\left(ref_{ij}-\overline{ref}_{(w<-1)}\right)\right)^{1/2}}$ [See Appendix B]

3.3.2 Storm Tracking Technique

To ensure that extracted meteorological quantities reflect the changing characteristics of a single evolving simulated supercell storm, a suitable storm tracking scheme is required to keep the SER_{5km} aligned with a simulated storm

throughout its lifetime. The technique implemented in this study is inspired by the National Severe Storms Laboratory (NSSL) Storm Cell Identification and Tracking (SCIT) algorithm (Johnson et al. 1998) with the notable exception that it is based upon vertical velocity instead of radar reflectivity.

In our application of a modified SCIT algorithm, a combination of the SIR and its associated SER_{5km} are used to appropriately track individual storms. Figure 3.5 illustrates the tracking procedure in four typical SIR translation scenarios (single storm, two storms, slitting storms and merging storms).[13] Each scenario is divided into three consecutive history file output dump times (hereafter referred to as output times) where time "t" corresponds with the initial cell alignment and times "t+1" and "t+2" each contain two plots to more thoroughly illustrate how the tracking scheme works. The first of these two plots provides all identified potential SIR's outlined with a solid black line and labeled with a question mark (to indicate they are only potential cells at this point). The black dotted lines correspond with identified cells at the previous output time and the red dot-dashed lines represent the search spaces associated with the previous cells that are used to locate a cell's new position. The search space is a 5 km outward extension of a cell's previous boundary location (e.g., its previous SER_{5km}) and represents all possible locations the cell could have traveled in one 30 second output time interval, including potential SIR evolution[14]. If an

[13] Becoming an officially tracked SIR with a SER_{5km} only gives the SIR an opportunity to be used for defining a storm region in which meteorological quantities are extracted and used in the data mining procedure. The SIR and its SER_{5km} must be tracked for longer than 30 minutes before it fully qualifies.

[14] This search space is an overestimation of the distance a cell could travel in one 30 second output time interval since a 5 km search space allows a cell to move over 150 m s^{-1} and still be associated with the previous cell. This overestimation is used to ensure cells are not missed and to account for any unusual motion that may occur due to automated grid translation of the domain. Test cases also supported this choice.

identified SIR overlaps with any part of this search space, then it is considered a continuation of the previous SIR. If, however, a SIR is located outside of any search spaces, then it is designated a new SIR. The second of the two plots for times "t+1" and "t+2" provide the resulting SIR's that continue to be tracked based on the search space analysis from the first of the two plots for these times. Officially tracked SIR's are labeled by alphabetic letters containing subscripts denoting the number of output times they have been tracked (e.g., A_1, A_2 and A_3 in the three output times of Figure 3.5(a)).

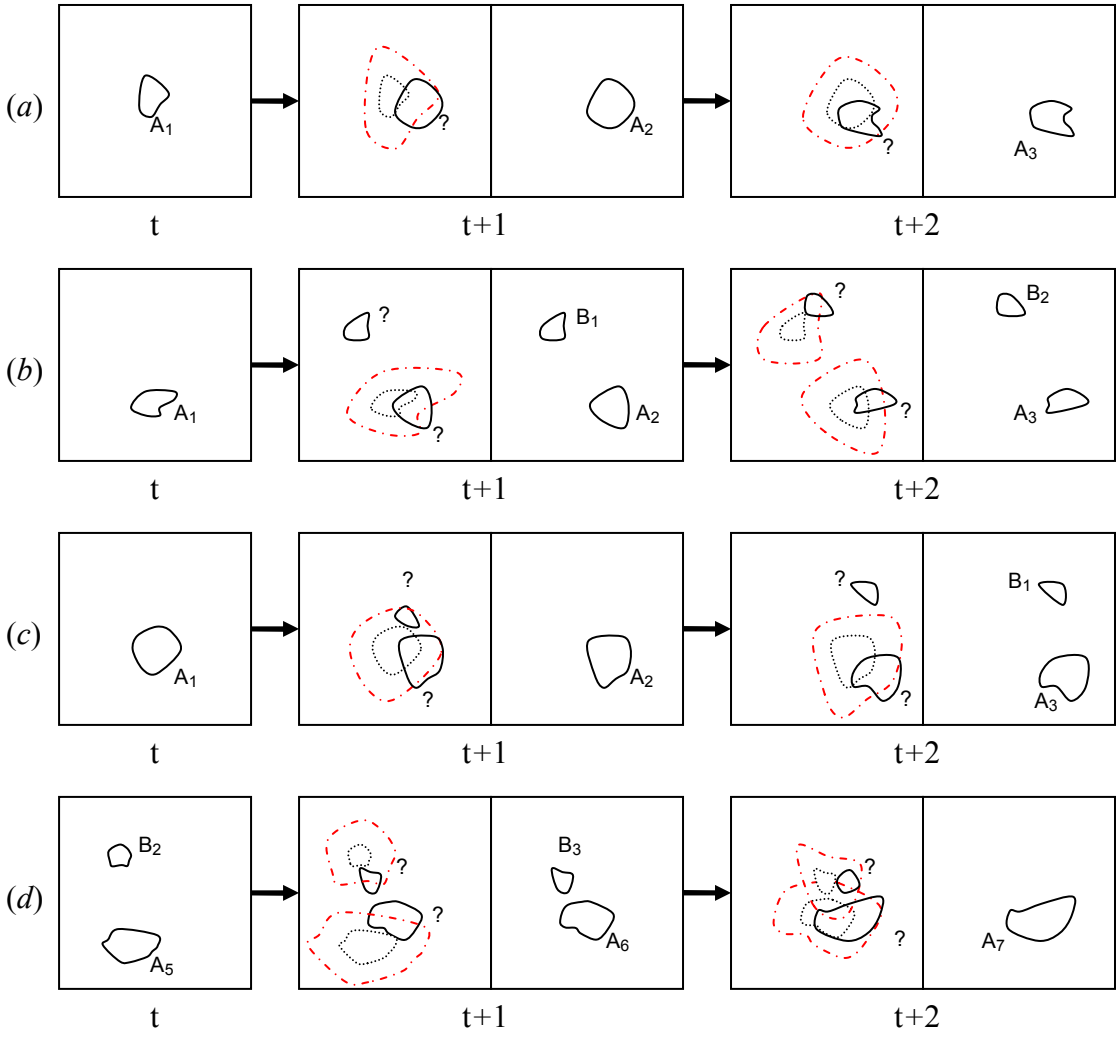


Figure 3.5: Schematic diagram of the technique used to track storm indicator cells of $w \geq 15$ m s^{-1} at 4 km. Examples are shown for tracking (*a*) one cell, (*b*) two cells, (*c*) splitting cells and (*d*) merging cells. Solid black lines indicate identified cell regions for each output time while dotted black lines indicate cells from the previous output time. Dot-dashed red lines represent the search space associated with cells of the previous output time which are used to locate the same cells after one output time interval.

Figure 3.5(a) provides the most general case involving only a single SIR. At output time "t," the SIR is labeled "A_1," which indicates it is the first output time this particular SIR has been tracked. In the first plot of time "t+1," a black dotted line

highlights the previous location of SIR "A_1" along with its 5 km search space extension (red dot-dashed line). The identified potential SIR at "t+1" also is plotted (solid black line labeled with question mark) and is found to be within the search space of SIR "A_1:" therefore, it is defined as the officially tracked continuation of that SIR and is assigned the label "A_2" since this is the second output time it has been tracked. Time "t+2" displays similar characteristics as "t+1" and results in SIR "A_3."

Figure 3.5(b) is the two-cell extension of the single-SIR case. Again, a search space is created in the first plot of time "t+1" for SIR "A_1," only this time, two identified potential SIR's exist. One is located within the search space of "A_1" and therefore is identified as its continuation ("A_2") and the other potential SIR is located outside any search space region and therefore is labeled as a new SIR ("B_1"). Time "t+2" then tracks each SIR further, resulting in "B_2" and "A_3" in the same manner as described previously.

Figure 3.5(c) illustrates the situation in which a single SIR splits into two. After the split, two potential SIR's are located within the search space of the initial SIR "A_1" (see first plot of time "t+1"). When this occurs, the potential SIR's having the largest area encompassed by the search space is identified as the continuation of the previous SIR. In this case, the bottom (or southern) SIR has the largest area within the search space and therefore is labeled "A_2" while the upper (or northern) SIR is not tracked. This restriction is applied to keep closely spaced SIR's from establishing separate SER_{5km}'s that could potentially extract meteorological quantities from the same location, thereby erroneously duplicating the individual storm

metadata. At "t+2," the northern SIR has moved far enough away from the southern SIR to be outside its search space and therefore qualifies as a new SIR ("B_1").

Figure 3.5(d) is another special case involving merging SIR's. In this example, the initial time "t" begins with a northern SIR ("B_2") that has been tracked for 2 output times and a southern SIR ("A_5") that has been tracked for 5 output times. At time "t+1," the two SIR's undergo a normal tracking cycle with each advancing to a new position. At "t+2," however, the search spaces overlap significantly. When overlap occurs and at least one of the SIR's search space is overlapped by more than 25%, then only the SIR tracked for the longest period of time continues to be tracked. This restriction is put into place to limit the amount of incorrectly extracted metadata from SER_{5km}'s with very close proximity. For the example shown, SIR "A_6" has been tracked longer (six output time intervals) than SIR "B_3" (three output time intervals) and therefore the search space of SIR "A_6" is retained and the search space of SIR "B_3" is discarded. Of the two potential SIR's (solid black lines labeled with question marks), the southern SIR has the largest area within the search space of SIR "A_6" and therefore is identified as its continuation, ("A_7"). The northern SIR no longer is tracked.

An additional feature of this tracking scheme is that a SIR can decrease below the intensity criterion of $w \geq 15$ m s^{-1}, e.g., to $w \geq 10$ m s^{-1} at 4 km and still be tracked, uninterrupted. This is useful for analyzing the full lifespan of storms that weaken and then reintensify (i.e., drop below $w \geq 15$ m s^{-1} at 4 km but stay above $w \geq 10$ m s^{-1} for a period of time and then return to $w \geq 15$ m s^{-1}). In the context of the

simple examples in Figure 3.5, if no SIR of $w \geq 15$ m s^{-1} remains in a search space, then SIR's of $w \geq 10$ m s^{-1} are identified (if present) and labeled as continuations of the previous SIR in the manner previously discussed. Once SIR's drop below $w = 10$ m s^{-1} they are no longer tracked.

Every time an official SIR is identified, a 5 km storm boundary extension is defined and all 100 meteorological quantities are extracted (as shown in Figure 3.4). After all SIR's have been tracked throughout an entire simulation, those lasting longer than 30 minutes are defined as "storms" and their extracted meteorological quantity time series metadata qualifies for the data mining analysis procedure described in the next section.

3.3.3 Data Mining Technique

The 100 maximum and minimum time series generated for each storm have the potential to contain a sequence of events unique to those instances when strong low-level rotation develops. Because it is unrealistic to manually search this large metadata set, an automated technique is used to computationally sort through the quantities in search of repeated patterns that might define important precursors of low-level rotation. This automated technique is one example of a broader class of pattern/knowledge discovery techniques generally referred to as data mining (e.g., Hand et al. 2001).

Before explaining the data mining technique, it is first necessary to quantify the notion of strong low-level rotation. The criteria for strong low-level rotation are

as follows[15]: a negative pressure perturbation (i.e., departure from the base state) at the surface of at least -900 Pa (-9 mb) occurring in less than 1000 s accompanied by either an increase in vertical vorticity magnitude at the surface greater than 0.03 s^{-1} in less than 500 s or an increase in surface horizontal wind speed of 5 m s^{-1} in less than 750 s. Further, these criteria must be met within a 600 s time window to ensure they are correlated. A summary of these requirements are stated below with $\Delta p'_{sfc}$, $|\Delta\zeta_{sfc}|$ and Δhws_{sfc} as the surface changes in perturbation pressure, vertical vorticity magnitude and horizontal wind speed respectively. If a storm passes these requirements it is identified as a "Positive" storm.

$$\text{600 s [10 min]} \left\{ \begin{array}{l} \Delta p'_{sfc} \le \text{-900 Pa in } t \le 1000 \text{ s [~17 min]} \\ \left(\begin{array}{l} |\Delta\zeta_{sfc}| \ge 0.03 \text{ s}^{-1} \text{ in } t \le 500 \text{ s [~8 min]} \\ \textit{and} \\ \Delta hws_{sfc} \ge 5 \text{ m s}^{-1} \text{ in } t \le 750 \text{ s [~13 min]} \end{array} \right) \\ \textit{or} \end{array} \right. , \tag{3.3}$$

Storms that are slightly weaker than "Positive" storms are categorized into a second group in order to distinguish differences between those potentially producing weak low-level rotation and those producing strong low-level rotation. These are defined as "Intermediate" storms and contain one of the following attributes: (a) only a negative pressure perturbation at the surface of at least -900 Pa occurring in less than 1000 s, or (b) a negative pressure perturbation at the surface between -300 and -

[15] Criteria were established based on preliminary analyses of extracted storm metadata as well as general knowledge of numerically simulated storms and observed storms. These criteria are not universal but instead may vary depending on a numerical model's grid spacing and integration timestep.

900 Pa in less than 1000 s and either of the thresholds set for vertical vorticity magnitude maximum or horizontal wind speed maximum at the surface from the "Positive" storm criteria. These requirements for "Intermediate" storms are summarized in equation 3.4 below.

$$\text{600 s [10 min]} \begin{cases} \Delta p'_{sfc} \leq -900 \text{ Pa in } t \leq 1000 \text{ s [}\sim\text{17 min]} \\ \textit{or} \\ -300 \geq \Delta p'_{sfc} \geq -900 \text{ Pa in } t \leq 1000 \text{ s [}\sim\text{17 min]} \\ \textit{and} \\ \left(\begin{array}{l} |\Delta\zeta_{sfc}| \geq 0.03 \text{ s}^{-1} \text{ in } t \leq 500 \text{ s [}\sim\text{8 min]} \\ \textit{or} \\ \Delta hws_{sfc} \geq 5 \text{ m s}^{-1} \text{ in } t \leq 750 \text{ s [}\sim\text{13 min]} \end{array} \right), \end{cases} \tag{3.4}$$

Storms not categorized as either "Positive" or "Intermediate" are identified as "Negative" storms. The inclusion of "Intermediate and "Negative" storms are needed in the data mining procedure because they provide comparison cases having their own unique meteorological characteristics which differ from "Positive" cases. Distinguishing these differences is the goal of the data mining analysis.

The data mining procedure requires a uniform time sequence across all storms when searching for influential characteristics within the metadata. Because the time period leading up to the development of strong low-level rotation is deemed most important, a 30 minute window is chosen with the end of that window placed at the

initiation of the rotation[16]. The 30 minute time windows for "Intermediate" and "Negative" cases are taken from arbitrary locations within their storm lifetimes because no strong low-level rotation develops.[17] The metadata covering the lifetime of each storm are therefore scaled down to 30 minute segments. These 30 minute storm segments, along with their respective "Positive," "Intermediate" and "Negative" labels, represent the base set of information used in the data mining process.

Because the metadata contain many small fluctuations associated with their 30 s interval time series structure, it is difficult for a computationally automated search to identify and compare the most important variations within each of the meteorological quantities. Therefore, the first step in the data mining process is to discretize the metadata in order to minimize the number of dimensions (or possible values) that any one meteorological quantity can assume while still maintaining its defining characteristics. The term discretize is used, and will be for the remainder of the discussion, with the understanding that the model output is already in a discretized form which means the following procedure further discretizes the metadata.

The discretization method of Lin et al. (2003) is used, which normalizes a data set to fit a Gaussian curve with zero mean and unity standard deviation. The data are then divided into regions corresponding to equal-sized areas under the Gaussian

[16] To identify the initiation of strong low-level rotation, a first derivative Gaussian convolution filter with multiple widths is used to locate the beginning of the negative pressure perturbations associated with the rotation.

[17] Future work will attempt to locate the time period of weak rotation within "Intermediate" storms and use it as the ending point for the defined 30 minute window for these storms.

curve.[18] A data sequence may be discretized into any number of equiprobable Gaussian regions following the standard statistical table shown in Table 3.3. Lin et al. (2003) tested the significance of the number of regions chosen to discretize a wide range of time series data sets and found that between five and eight regions seemed to be an appropriate choice for maintaining the important characteristics of any given data set. For this study, therefore, the meteorological quantity time series metadata sets are each divided into 3, 4, 5, 6, 7 and 8 different Gaussian regions and then tested to see which produce the highest rated data mining results. The use of nine Gaussian regions also was tested but, as will be indicated in the results section, this generated an overload of data which could not be analyzed by the data mining procedure. For simplification purposes a choice of five Gaussian regions will be used as the example for the remainder of the discussion. These five regions are labeled according to the alphabet letters "a", "b", "c", "d" and "e" and therefore the total number of regions within a data set can be interchangeably referred to as its "alphabet size".

[18] A near Gaussian shaped distribution of a time series data set is required for this method to be applied successfully, and while some of the meteorological quantity distributions were found to stray somewhat from this shape, the majority are Gaussian.

Table 3.3: Statistical table giving breakpoints (β_i) used for dividing a Gaussian distribution of mean zero, standard deviation unity into a given number of equiprobable regions (*a*). [From Lin et al. (2003).]

β_i \ *a*	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

After establishing the alphabet size, an appropriate averaging interval must be chosen that will preserve the general structure of the metadata while at the same time minimize any insignificant amplitude variations. The time series data over this interval are averaged together and then assigned the letter of the equiprobable Gaussian region in which the average value falls under. Because the original model output of this study is saved at 30 s intervals, tests were performed using 30 s, 60 s, 90 s, 120 s and 150 s averaging intervals to identify averaging intervals providing the highest quality data mining results. Averaging intervals beyond 150 s were not considered because any further averaging would smooth the data beyond time intervals deemed appropriate for capturing evolving storm features. An averaging interval of 60 s (2 output intervals) will be used throughout the remainder of this discussion for illustrative purposes.

After each of the meteorological quantities within a 30 minute storm segment is properly discretized, the metadata can then be represented by a string of 30 letters. Figure 3.6 illustrates this entire discretization process for a simplified example from Keogh et al. (2005) having only three equiprobable regions (i.e., "a," "b," and "c"), compared to our five regions, and an averaging interval of roughly 18 arbitrary units corresponding with the length of the colored horizontal bars, compared to our choice of 60 s (two output intervals). The generalized data set in Figure 3.6 can therefore be represented by the eight letter discretized sequence, "cbccbaab."

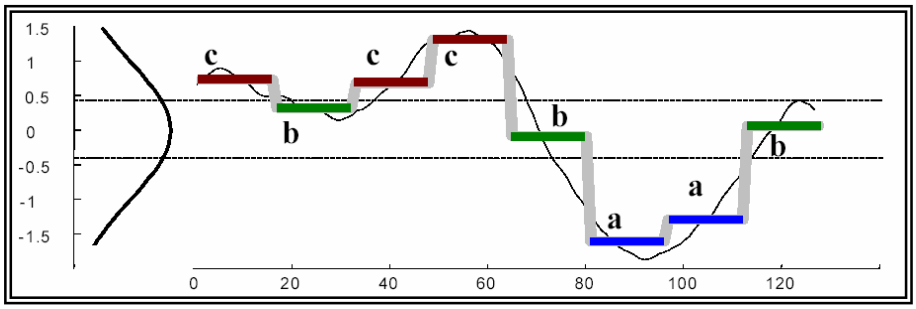


Figure 3.6: Example of how the meteorological quantity metadata are discretized using the Lin et al. (2003) method where the data are normalized to fit a Gaussian curve with mean zero and standard deviation one and then split up into three regions of equal-sized areas under the Gaussian curve. [From Keogh et. al. (2005).]

The final step required prior to data mining is the formation of subsequences from each string of letters, which we refer to as "words." The words can be made up of any number of letters but because the words will be combined later in the data mining procedure it is only necessary to form words with either two or three letters. For this discussion, words composed of three letter segments will be considered.

These words represent not just a simple separation of the strings into three letter segments, but rather contain all three letter combinations that are possible with each single letter increment in time. For example, a meteorological quantity with the letter sequence "aabbcbbbc" contains seven possible word combinations, where the word "aab" is associated with the first letter in the sequence, followed by "abb" associated with the second letter of the sequence, followed by "bbc," "bcb," "cbb," "bbb" and "bbc."

With the 100 meteorological quantities for each 30 minute storm segment now represented by 28 separate 3-minute word combinations, the data are sufficiently scaled down to begin the data mining search process. This process involves searching for repeated words and word combinations occurring in the "Positive" cases but not the "Intermediate" or "Negative" cases, thus making them unique to the development of strong low-level rotation. Two data mining analyses will be performed: one composed of only the "Positive" and "Negative" storms and the other composed of only the "Positive" and "Intermediate" storms. The "Positive" and "Negative" storm analysis will attempt to differentiate storms that presumably have fewer characteristics in common while the "Positive" and "Intermediate" storm analysis will attempt to differentiate storms with similar characteristics.

Figures 3.7 and 3.8 provide a simplified illustration of how the data mining process works. An arbitrary six-storm data set for a single meteorological quantity (generalized as X) is given in Figure 3.7 with words having time labels of 1-7 representing the location, within the original letter sequence, of the first letter of an

individual word. As an example, Storm 1 contains the word combinations of the letter sequence "aabbcbbbc" as discussed previously. Also listed are the "Positive" or "Negative" storm labels, corresponding to the presence or absence, respectively, of strong low-level rotation within the storm. Only the "Positive" and "Negative" storm analysis will be considered for this example. The "Positive" and "Intermediate" storm analysis will be conducted in exactly the same fashion.

Figure 3.8 shows the first step of the search process, which involves building a single contingency table for the presence of a specific word for meteorological quantity X across all storms. These single word contingency tables are built under the assumption that a word existing in a "Positive" case is a "hit" (or true positive) and one existing in a "Negative" case is a "false alarm" (or false positive). Therefore, the presence of a word can be thought of as a prediction that low-level rotation will be present. The example contingency table given in Figure 3.8 is for the 3-letter word "abb" which is highlighted in red in Figure 3.7. This word is present in three "Positive" storm cases which correspond to three "true positives". It also is present in one "Negative" storm case, which corresponds to one "false positive". Note that "abb" occurs twice in Storm 4 but is only counted as one "true positive" because the requirement is simply that the word occurs at any time within a storm. The "Negative" case of Storm 2 does not contain the word and therefore is labeled as a "true negative," while the "Positive" case of Storm 6 does not contain the word either and therefore is labeled as a "false positive".

Storm Label	time / storm	1	2	3	4	5	6	7
pos	1	aab	abb	bbc	bcb	cbb	bbb	bbc
neg	2	aaa	aab	abc	bcb	cba	baa	aab
neg	3	aab	abb	bbb	bbc	bcc	ccc	ccc
pos	4	abb	bbb	bbb	bbb	bba	bab	abb
pos	5	aab	abb	bbc	bcc	ccb	cbb	bbb
pos	6	aab	aba	bab	aba	bab	aba	baa
		Meteorological Quantity X						

Figure 3.7: Simplified illustration of searching for words within discretized time series plots of a meteorological quantity (X). Six storms are listed along with their "Positive" or "Negative" labels associated with the development of strong low-level rotation. The time sequence of three letter words correspond to all possible three letter combinations based on one letter increments from the original letter sequence. For example, the original letter sequence for Storm 1 is "aabbcbbbc". The word "abb" (shown in red) corresponds with the example contingency table and performance measures given in Figure 3.8.

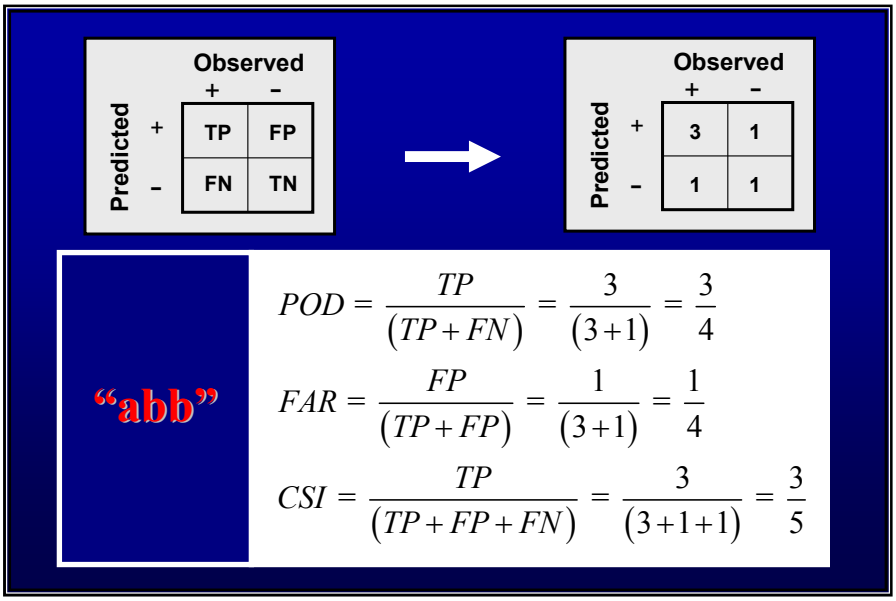


Figure 3.8: Example contingency table and corresponding performance measures of POD, FAR and CSI for the word "abb" from Figure 3.7.

After the contingency table has been built, the performance measures of probability of detection (POD) and false alarm ratio (FAR) (discussed in Chapter 1) are generated (see Figure 3.8) and used for comparing words to identify which are better indicators for the presence of strong low-level rotation (i.e., which words are present only in "Positive" storm cases). If a word is present only in those storms containing strong low-level rotation, then the equations in Figure 3.8 can be used to show that such storms will have a POD of unity and FAR of zero. The critical success index (CSI) also is calculated but is not used until the final sorting step, discussed subsequently.

Contingency tables and performance measures therefore are generated for each 3 letter word within the word sequence of each meteorological quantity across all storms (i.e., every three letter word shown in Figure 3.7). With words representing only a 3 minute time period, multiple word combinations also must be created and their performance measures calculated in order to identify longer sequences potentially influential to the development of strong low-level rotation. These multi-word sequences are built from potentially significant words having POD ≥ 0.8 and FAR ≤ 0.8. These criteria were established based on a trial and error effort to utilize a sufficient number of words without allowing the computational time and data storage to become too large in subsequent data mining steps.[19] Contingency

[19]A more restrictive POD threshold is used to ensure that almost all of the baseline words are occurring in the positive cases. FAR is less restrictive to allow a broader set of initial words for forming the multiple word combinations.

tables and performance measures again are generated for each of the multi-word combinations.

For example, if the words “abb” and “bbc” are identified as influential baseline words in the general example shown in Figure 3.7, then contingency tables and performance measures are generated for each of their word combinations shown in Figure 3.9. These word combinations must be in sequential order which is why, for example, Storm 3 in Figure 3.7 identifies a word combination of “abb,bbc” but not “bbc,abb”. All word combinations thus can be interpreted as a sequence of events. The time separation between combined words in this sequence is not included due to the increased complexities it would produce. For example, in Storm 1 the word combination “abb,bbc” occurs at time “2” and “3” and also occurs at time “2” and “7.” However, in the data mining analysis the word combination “abb,bbc” is identified as simply occurring in Storm 1 beginning at time “2.” Including the time separation between words would complicate subsequent steps in the data mining process and generate sequences that are so specific that they rarely occur with the exact same timing across different storms, therefore making the process of finding repeated patterns nearly impossible. This point will be highlighted and further discussed in the results section.

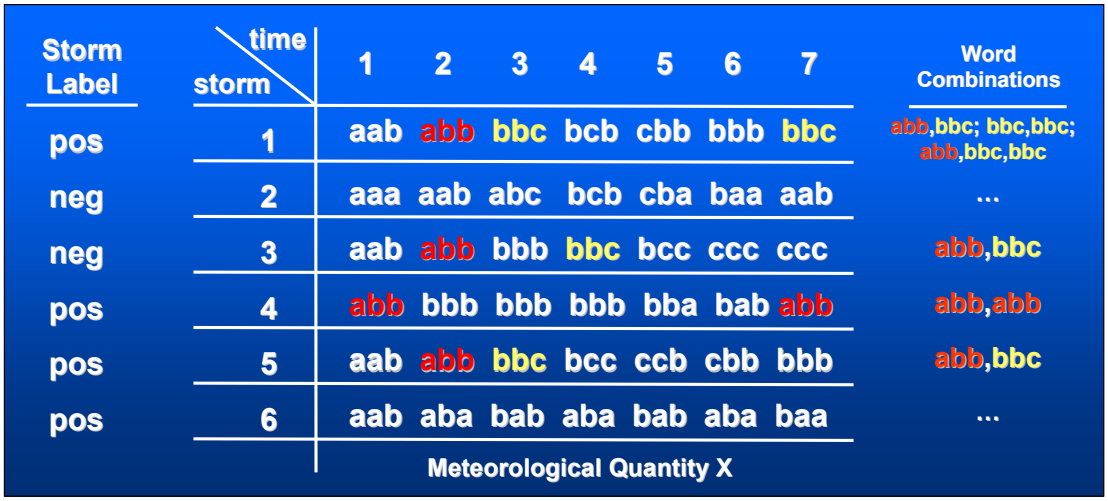


Figure 3.9: Same as Figure 3.7 but includes all baseline word combinations in chronological sequence. "abb" (red) and "bbc" (yellow) are the example baseline words used for the word combinations.

Additionally, contingency tables and performance measures are generated for word combinations from across different meteorological quantities. Thus far, the examples given have involved a single quantity X. With baseline words occurring across many different meteorological quantities, their combined word combinations also must be analyzed to identify any important sequences among them that herald the development of strong low-level rotation. For example, quantity X may drastically increase followed by quantity Y decreasing slightly, followed by quantity Z slowly increasing – all of which may occur only in those instances when strong low-level rotation develops.

All possible word combinations therefore are identified across all meteorological quantities and their performance measures (POD, FAR and CSI) generated. Figure 3.10 provides a simplified, two-quantity example of this process in

which the baseline words for the first three storms of quantity X are combined with those of another quantity defined as Y. The three storms and their "Positive" or "Negative" storm labels are listed, along with the quantity name shown as a storm number subscript. The highlighted words "abb" (red) and "bbc" (yellow) for quantity X and "bbb" (green) for quantity Y correspond with each quantity's identified influential baseline words.

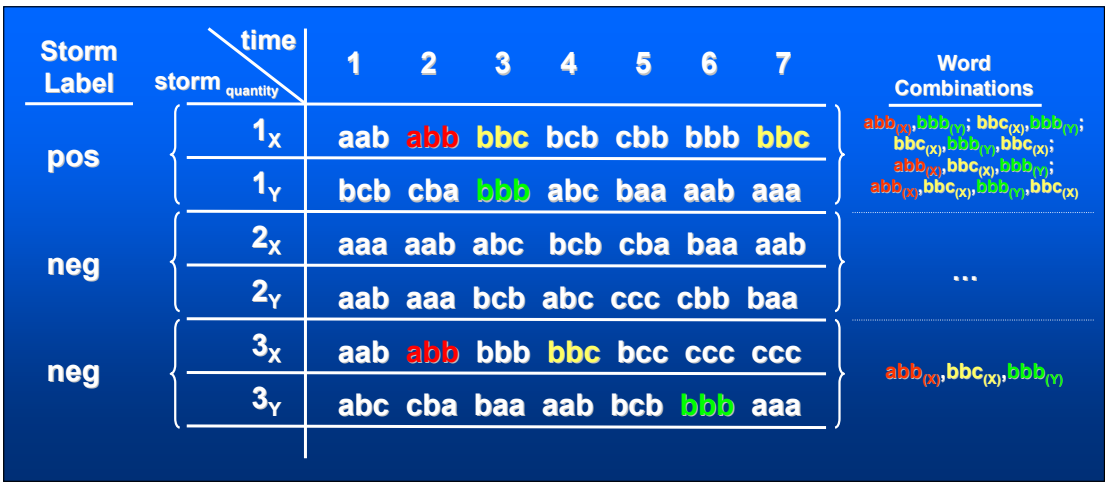


Figure 3.10: Same as Figure 3.7 but lists the baseline word combinations across two meteorological quantities X and Y for a three storm data set. Baseline words for quantity X are "abb" (red) and "bbc" (yellow) and for quantity Y is "bbb" (green).

The word combinations again must occur in chronological order; however, in this instance, the various baseline words may occur within the same time interval (i.e., "bbc" from X and "bbb" from Y both occur at time interval 3 in Storm 1). In such cases, the two words are placed next to one another within a defined word combination. Therefore, successive words in the final word combination listings across quantities can be interpreted as either occurring in chronological order or at the

same time. For example, the cross-quantity word combination "$abb_{(X)}$,$bbc_{(X)}$,$bbb_{(Y)}$" may correspond to the sequence of events "$abb_{(X)}$" followed by "$bbc_{(X)}$" followed by "$bbb_{(Y)}$" (i.e., storm 3 in Figure 3.10) or the sequence of events "$abb_{(X)}$" followed by "$bbc_{(X)}$" and "$bbb_{(Y)}$" at the same timestep (i.e., timesteps 2-3 of Storm 1 in Figure 3.10).

The resulting set of baseline words and multi-word combinations from across all meteorological quantities and their performance measures are collected into a single master list. This list is sorted in a best-to-worst format, reflecting the single word or multi-word combinations most associated with storms containing strong low-level rotation. The performance measure CSI is used to make this assessment because it combines the effects of both POD and FAR, as shown in Figure 3.11. A CSI of unity is best and corresponds with a POD of unity and FAR of zero, thus meaning a word or word combination was present only in storms containing strong low-level rotation. Improving CSI requires POD to increase and FAR to decrease, and therefore CSI provides an effective performance measure for comparing various single- and multi- word combinations. The final sorted list of words and multi-word combinations provide a collection of events within the meteorological quantities for various simulated storms that are rated according to how well they predict the occurrence of strong low-level rotation.

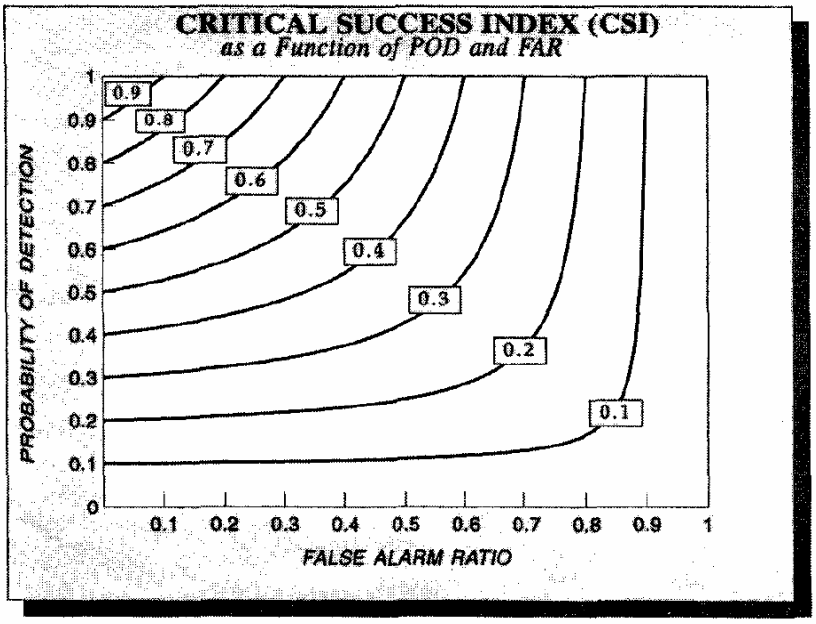


Figure 3.11: CSI as a function of both POD and FAR. [From Polger et al. (1994).]

Finally, to provide a more robust set of results, 10-fold cross-validation is applied to test the validity of the final sorted list of words and multi-word combinations. Ten-fold cross-validation separates a full data set into 10 sub-groups and performs an analysis on nine of the sub-groups for training purposes and then performs the analysis on the remaining sub-group as a test to validate the results. This process is repeated 10 times with each sub-group used exactly once as a test case. The results of the 10 test cases can then be averaged together to determine whether the sub-groups generate results comparable to the combined analysis. For the example used in the previous discussion, the "Positive" and "Negative" storm cases are distributed evenly between 10 sub-groups with an equal ratio of "Positive" to "Negative" storms in each sub-group to insure the results are not biased by an

unequal “Positive” or “Negative” storm distribution from one sub-group to another. The same is done for the analysis of the “Positive” and “Intermediate” storms before the 10-fold cross-validation procedure is conducted. The top results of the full analysis and 10-fold cross-validation analysis for a substantial number of discretization scheme variations (50 different scheme variations) are presented and discussed in Chapter 4.

CHAPTER 4

RESULTS

The results of the numerical simulation experiments and data mining analyses are presented in this chapter. In the first section, an example simulation experiment is presented along with corresponding extracted meteorological quantities to illustrate the extraction process. An overview of storm intensities and storms producing strong low-level rotation is also provided to quantify the full range of storm intensities generated within the chosen parameter space. In the second section, the interrelated, repeated quantity patterns occurring prior to the development of strong low-level rotation are presented for the data mining analyses performed.

4.1 Extracted Meteorological Quantities

The numerical simulation experiments (256 simulations) each generated, on average, between three and six official storms (SER_{5km}) meeting the storm definition requirements presented in sections 3.3.1 and 3.3.2. A total of 1168 storms were generated over all simulations, each containing a metadata set of 100 extracted maximum and minimum meteorological quantities. Figure 4.1 provides output from an example numerical simulation to illustrate the evolution of identified storms. The images on the right display updraft $\geq$ 15 m s^{-1} at 4 km altitude, corresponding to the storm indicator region (SIR) described in sections 3.3.1 and 3.3.2. A single continuously tracked SIR remains the same color as long as it maintains SIR status.

The left images show derived radar reflectivity[20] at 4 km altitude as well as a 5 km extension of the SIR (solid dark line) which corresponds to the storm extraction region (SER_{5km}) defined in Section 3.3.1. The fields are provided at 15 minute intervals over the entire three-hour simulation.

The initial convective development at the center of the domain, forced by a thermal bubble perturbation, is apparent in Figure 4.1(a). At this time the updraft has already surpassed the criteria for a SIR and is highlighted blue in the right column. A SER_{5km} corresponding to the SIR is located in the derived radar reflectivity plot (left column) but does not imply the existence of an officially tracked storm at this time because the criteria also require a 30-minute lifetime. Meteorological quantities are being extracted but will be discarded if the 30-minute criterion is not met. The initial convection maintains its SIR (blue) and associated SER_{5km} in Figure 4.1(b), 30 minutes into the simulation, but is not yet considered an official storm because the SIR (blue) developed 690 s (11.5 min) into the simulation (not shown) and therefore has been present for only 1110 s (18.5 min). After 45 minutes into the simulation (Figure 4.1(c)), the SIR (blue) surpasses the 30 minute threshold and the corresponding SER_{5km} is officially designated a storm. The extracted metadata from the SER_{5km} is now allowed in the data mining procedure, beginning with those quantities extracted when the SIR (blue) was first established (690 s into the simulation). Also at 45 minutes (Figure 4.1(c)) a new SIR (green) is identified,

[20] See Appendix C for calculation of model-derived radar reflectivity.

corresponding with a region of developing convection just to the north[21] of the original SIR (blue). To simplify the following discussion, a storm will be labeled according to the order in which it became an official storm and the color of its corresponding SIR (i.e., Storm 1 (blue)). Also, SIR's will be labeled according to their sequential order and color (i.e., 1st SIR (blue) and 2nd SIR (green)).

After an hour into the simulation (Figure 4.1(d)), Storm 1 (blue) continues to be the dominant storm in the center of the domain. This central location is the result of its intense updraft and the automated grid translation's preferential adjustment of the domain towards the directional movement of the most intense updraft regions. The 2nd SIR (green) has translated farther to the north and is now adjacent to a 3rd SIR (red) and 4th SIR (light blue). Analyzing the full simulation (not shown) reveals that the 3rd SIR (red) split off of the 2nd SIR (green) and does not acquire an official SER_{5km} until it moves far enough away from the SER_{5km} of the 2nd SIR (green) for their overlap to meet the conditions set by the storm tracking scheme (Section 3.3.2). For that reason, no SER_{5km} accompanies the 3rd SIR (red) in Figure 4.1(d). The 4th SIR (light blue) develops far enough away from any adjacent updrafts and therefore has a corresponding SER_{5km}.

[21] The upper (lower) region of the simulated domain will be referred to as north (south) and the right (left) region will be referred to as east (west).

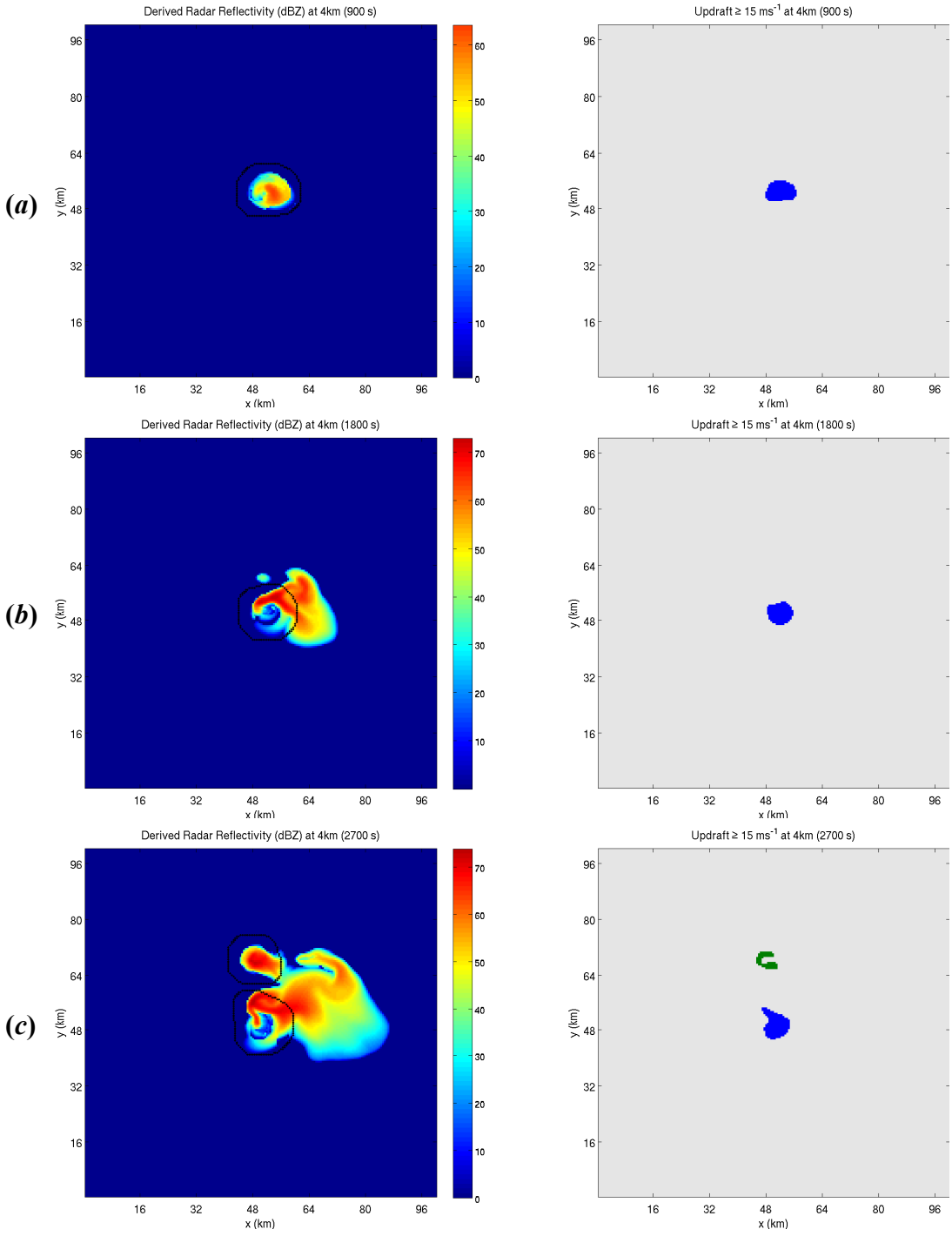


Figure 4.1: Example numerical simulation for initial environment WK14_half_r15_4km. Right column contains storm indicator regions (SIR), distinguished by color, indicating updraft ≥ 15 m s^{-1} at 4 km altitude and left column contains derived radar reflectivity (dBZ) and the storm extraction region (SER_{5km}) corresponding with a 5 km extension (black solid line) of a SIR for (*a*) 15 minutes, (*b*) 30 minutes and (*c*) 45 minutes into the simulation.

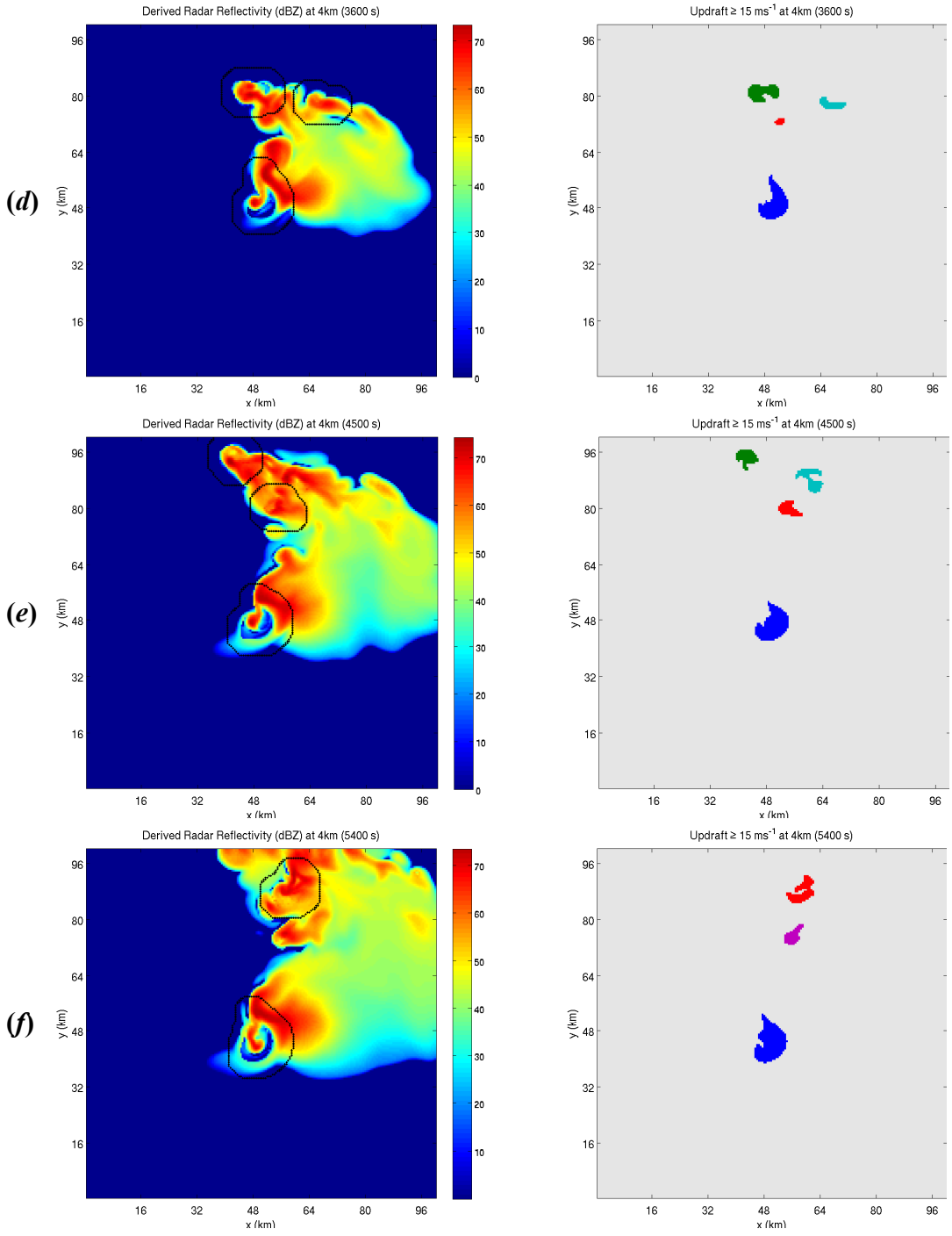


Figure 4.1: continued, for (*d*) 60 minutes, (*e*) 75 minutes and (*f*) 90 minutes.

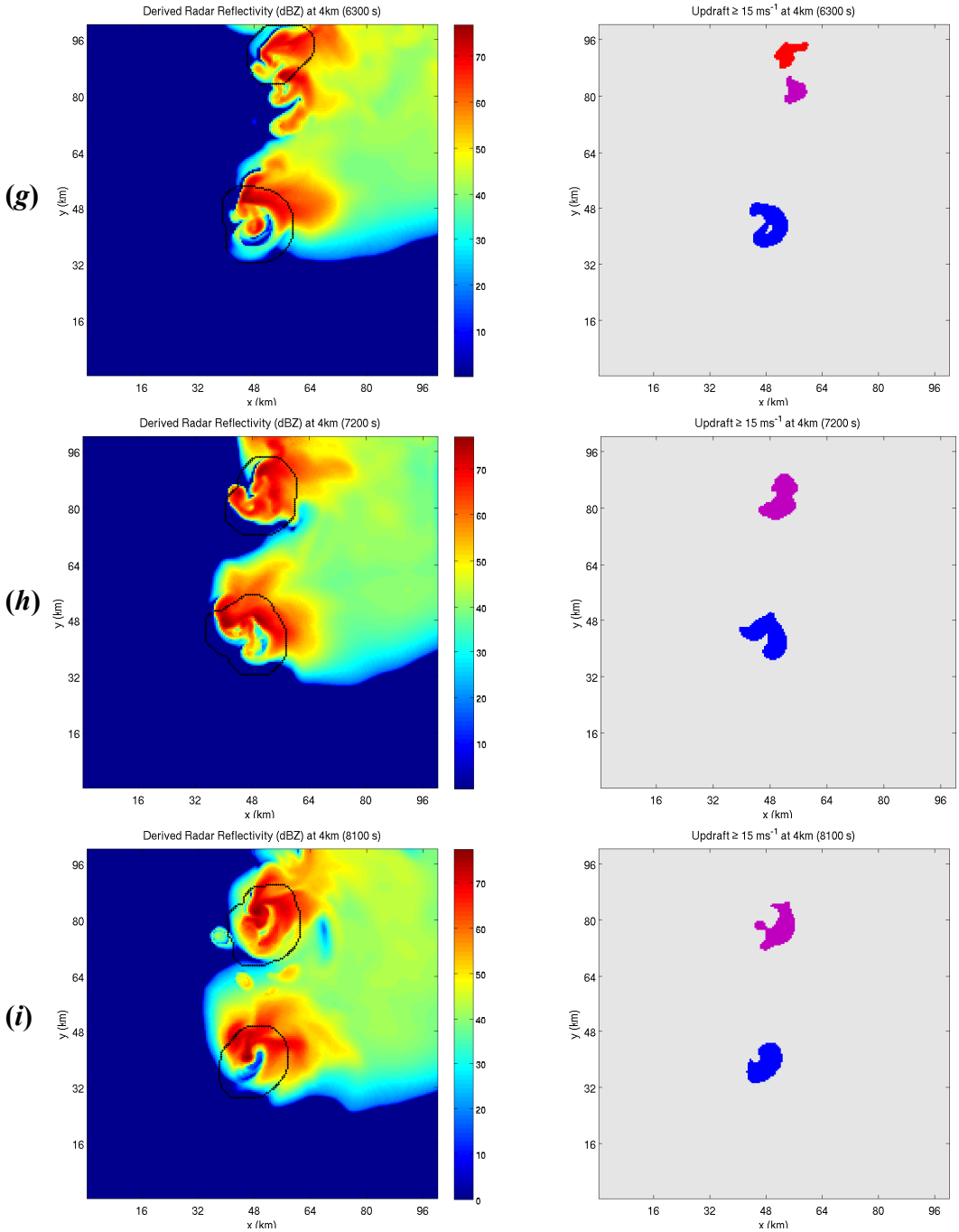


Figure 4.1: continued, for (*g*) 105 minutes, (*h*) 120 minutes and (*i*) 135 minutes.

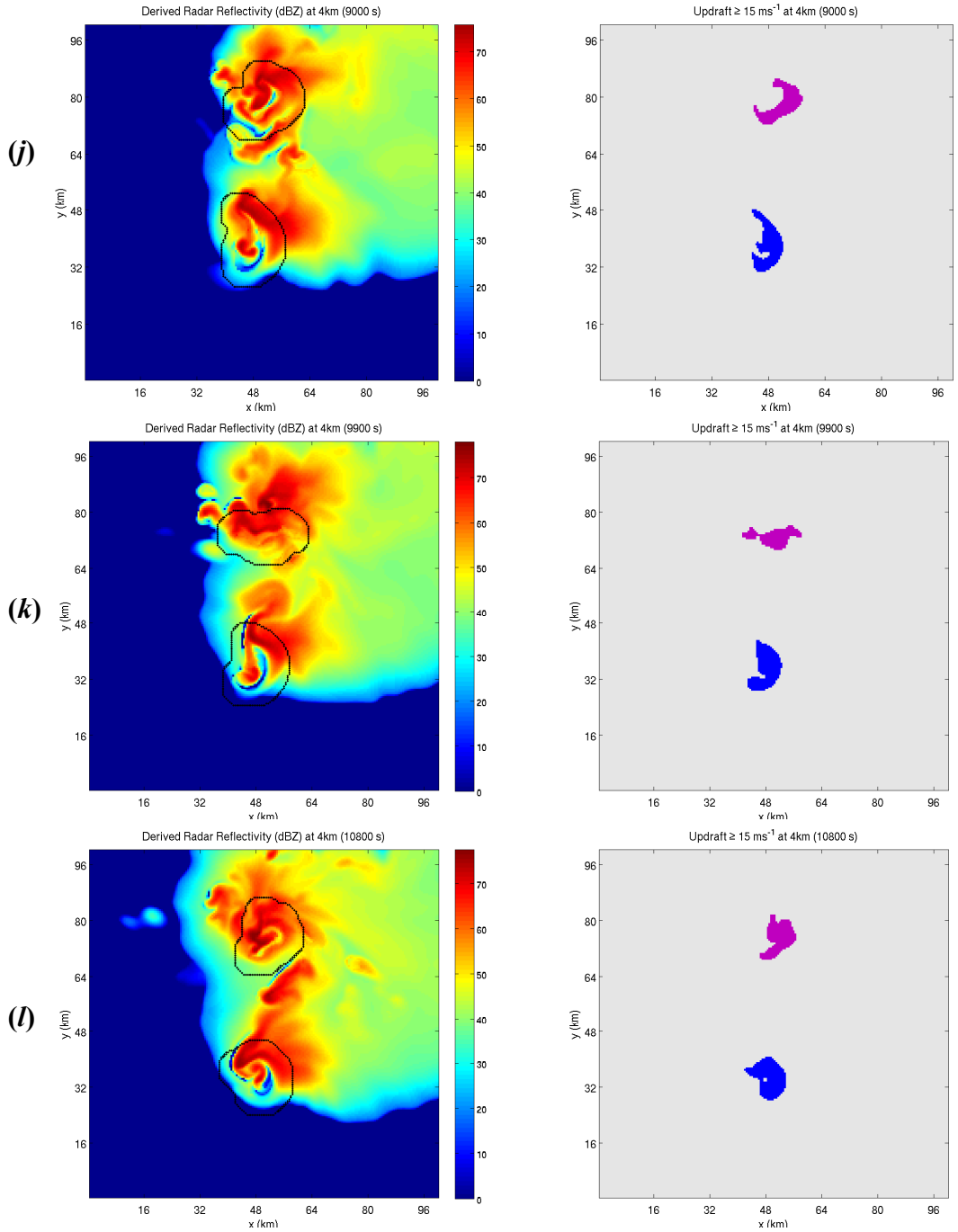


Figure 4.1: continued, for (*j*) 150 minutes, (*k*) 165 minutes and (*l*) 180 minutes.

In Figure 4.1(e) (75 minutes), Storm 1 (blue) maintains its central position while the other convective regions move farther to the north, approaching the northern boundary. The 2nd SIR (green), and its SER_{5km}, has been present for more than 30 minutes and therefore is labeled an official storm (Storm 2 (green)). The 3rd SIR (red) has moved a sufficient distance away from the 2nd SIR (green) to attain its own SER_{5km}. The 4th SIR (light blue) is still present but no longer has a corresponding SER_{5km} because between 60 minutes (Figure 4.1(d)) and 75 minutes (Figure 4.1(e)) it moved too close to the 2nd SIR (green), which took precedence due to its longer lifetime, and then once it moved far enough away from the 2nd SIR (green) it was too close to the recently created SER_{5km} of the 3rd SIR (red). Therefore, based off the storm tracking requirements (Section 3.3.2), it could not retain its own SER_{5km}.

Between 75 min (Figure 4.1(e)) and 90 min (Figure 4.1(f)), Storm 2 (green) exits the northern boundary of the domain. Also during this time, the 3rd SIR (red) and 4th SIR (light blue) merge into a single SIR and retain the label 3rd SIR (red). A 5th SIR (purple) forms just to the south of the 3rd SIR (red) but does not acquire a corresponding SER_{5km} because of its proximity to the 3rd SIR (red) (Figure 4.1(f)). Again, Storm 1 (blue) maintains its central position. At 105 min (Figure 4.1(g)), the 3rd SIR (red), and its SER_{5km}, has been present for longer than 30 minutes and becomes Storm 3 (red). The 5th SIR (purple) still is too close to Storm 3 (red) and therefore has no SER_{5km}.

Between 105 min (Figure 4.1(g)) and 120 min (Figure 4.1(h)), Storm 3 (red) exits the northern boundary which allows the 5th SIR (purple) to acquire a SER_{5km} and eventually become Storm 4 (purple) by 150 min (Figure 4.1(j)). Both Storm 4 (purple) and Storm 1 (blue) remain in the domain and maintain a similar distance from one another throughout the rest of the simulation (Figure 4.1(h) – Figure 4.1(l)). In the reference frame of the translating domain, there is a tendency for the two storms to move slightly south during the later portions of the simulation with the midpoint between their centers moving close to the center of the domain. This implies that both storms contain updrafts of comparable size and strength with similar directional motion and speed. The continued presence of Storm 1 (blue) throughout the entire example shown in Figure 4.1 also highlights the fact that storms containing the most intense updrafts can be kept within a simulation's horizontal domain throughout an entire simulation.

Each of the four storms identified in Figure 4.1 have maximum and minimum meteorological quantities extracted from their SER_{5km} (Section 3.3.1). Figure 4.2 provides a subset of quantities extracted from Storm 1 (blue) with Appendix D containing the full set of 100 extracted quantities for this storm. Each plot contains extracted metadata for a single maximum or minimum quantity over all vertical regions in which the particular quantity is extracted (see Table 3.2 for quantities and their vertical extraction regions). The quantities are plotted with respect to simulation time and therefore the metadata do not begin at time zero but rather at the time when the storm is established. Metadata time series extend throughout the remainder of the

3-hour simulation for this storm but not all storms will have metadata lasting throughout the entirety of a simulation because the storms may either leave the domain, merge with another storm or dissipate to below defined storm thresholds. Therefore, storms may subsist for different periods of time, producing a wide range of metadata intervals necessitating a defined, uniform 30 minute data mining analysis period (Section 3.3.3).

The notable feature of the time series metadata shown in Figure 4.2 are the larger perturbations within some of the quantities between 7000 – 8000 s (i.e., vertical velocity maximum below 2 km, vertical vorticity maximum at each level and minimum at the surface, pressure perturbation minimum at each level, horizontal wind speed at each level, perturbation potential temperature below 2 km and horizontal divergence min below 2 km). These perturbations are associated with the development of strong low-level rotation as defined in Section 3.3.3. The 30 minute period immediately prior to this development is the window used to search for quantity characteristics unique to the development of strong low-level rotation. The necessity of implementing an automated data mining procedure is apparent when examining this small set of quantities. Of the 1168 generated storms, 58 were "Positive," 373 were "Intermediate" and 737 were "Negative." It would be impractical to manually search through all 100 quantity time series from each of the simulated storms in an attempt to identify characteristics unique only to those cases where strong low-level rotation developed.

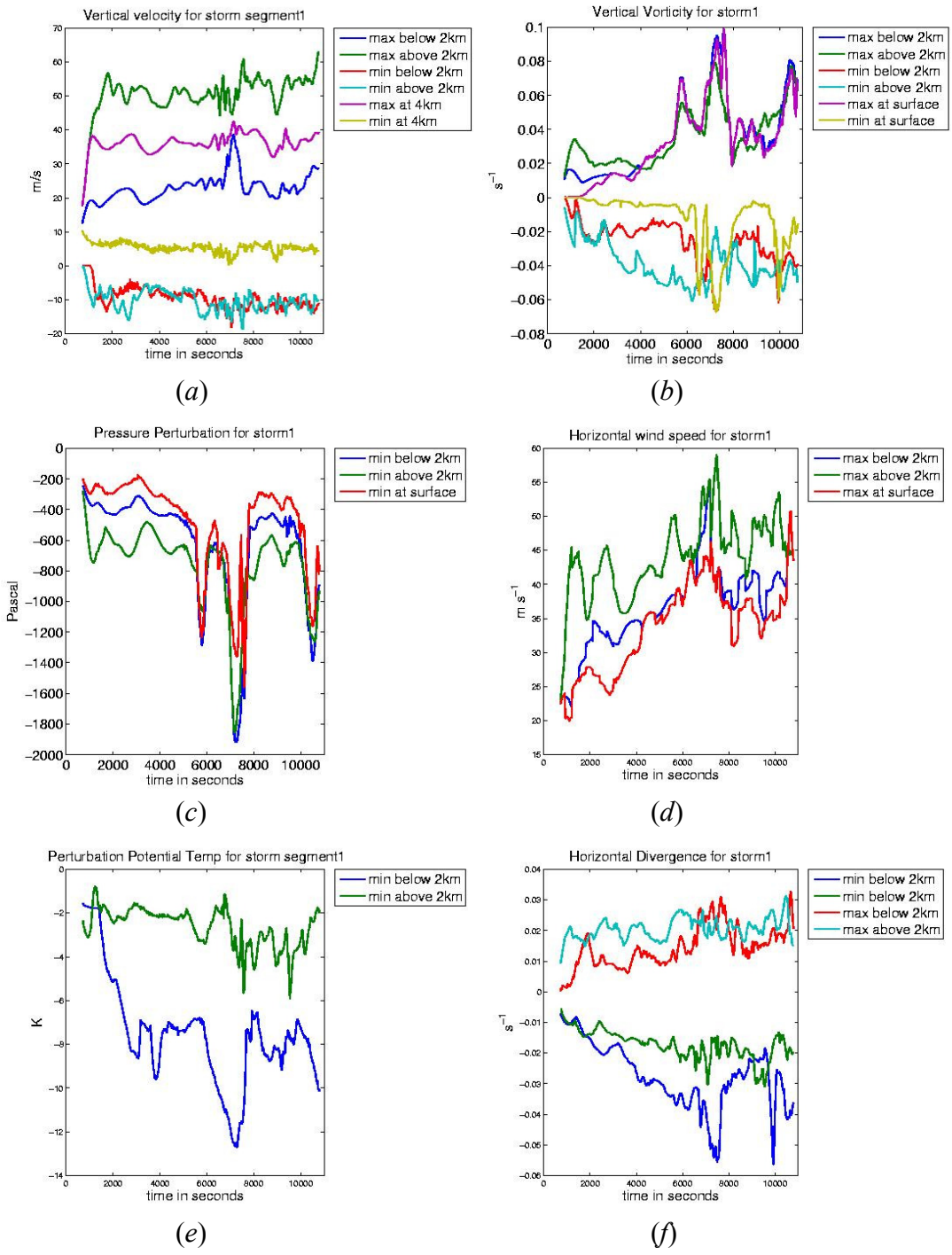


Figure 4.2: Meteorological quantities extracted from Storm 1 (blue) in the simulation WK14_half_r15_4km plotted in Figure 4.1. Maximum and minimum quantities for various vertical regions are plotted for (*a*) vertical velocity, (*b)* vertical vorticity, (*c*) pressure perturbation, (*d*) horizontal wind speed, (*e*) perturbation potential temperature and (*f*) horizontal divergence. The full set of 100 extracted quantities for this storm can be found in Appendix D.

Using the extracted meteorological quantities from each storm it is possible to quantify a relative range of storm intensities generated by the numerical simulation experiments. Figure 4.3 provides distributions of the most intense value a select set of quantities attain over each storm's lifetime. The quantities shown include vertical velocity maximum above 2 km, vertical vorticity maximum at the surface, horizontal wind speed maximum at the surface, pressure perturbation minimum at the surface, perturbation potential temperature minimum below 2 km and rainwater mixing ratio maximum above 2 km. As an example, the extracted storm quantity of vertical vorticity maximum at the surface shown in Figure 4.2(b) reaches a maximum value for the entire storm of approximately 0.1 s^{-1} near 7600 s. This maximum value is represented in the distribution plot in Figure 4.3(b) as a single storm entry within the 0.1 s^{-1} column.

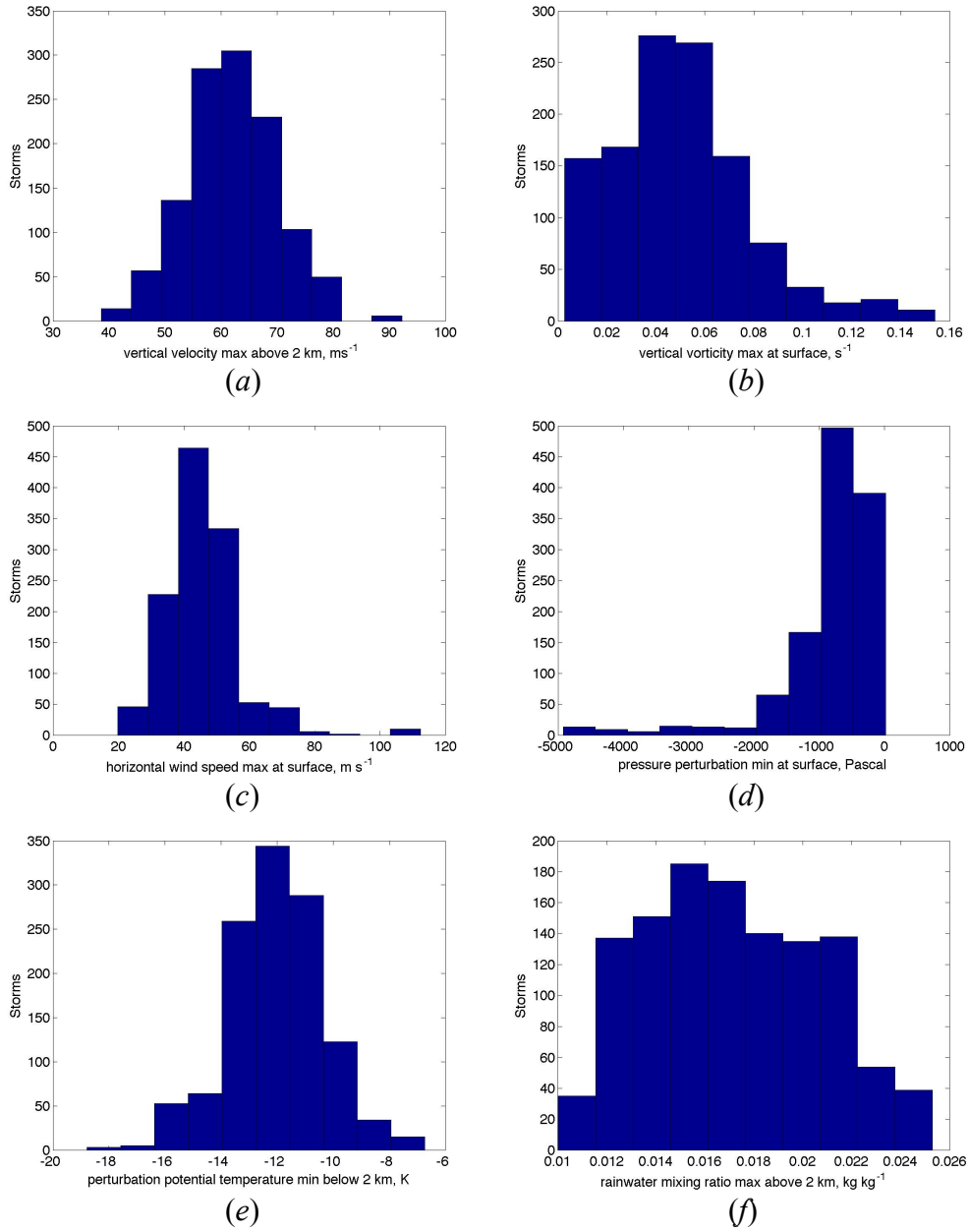


Figure 4.3: Distributions for the most intense value reached by each storm for the quantities (*a*) vertical velocity maximum above 2 km, (*b*) vertical vorticity maximum at the surface, (*c*) horizontal wind speed maximum at the surface, (*d*) pressure perturbation minimum at the surface, (*e*) perturbation potential temperature minimum below 2 km and (*f*) rainwater mixing ratio maximum above 2 km.

All 1168 storms are represented in the distribution plots of Figure 4.3 and it is obvious that a wide range of storm intensities was generated. The most intense storms contain vertical velocities above 2 km reaching 90 m s^{-1},[22] vertical vorticity maximum at the surface near 0.15 s^{-1}, horizontal wind speed maximum at the surface around 80 m s^{-1},[23] pressure perturbation minimum at the surface near -50 mb, perturbation potential temperature minimum below 2 km near -18 K and rainwater mixing ratio maximum above 2 km near 25 g kg^{-1}. The average values for each of the quantities were roughly 60 m s^{-1} for vertical velocity maximum above 2 km, 0.05 s^{-1} for vertical vorticity maximum at the surface, 45 m s^{-1} for horizontal wind speed maximum at the surface, -10 mb for pressure perturbation minimum at the surface, -12 K for perturbation potential temperature minimum below 2 km and 17 g kg^{-1} for rainwater mixing ratio maximum above 2 km. The presence of weaker storm intensities is expected given the number of smaller, less organized storms that are likely to be generated alongside the dominant storms in each simulation. These weaker storms are less important in terms of the presence of strong low-level rotation but they provide valuable "Negative" comparison cases when performing the data mining analysis.

Meteorological quantity extreme values also can be quantified for each simulation. Once extreme values are identified for each storm they can be compared

[22] It is generally accepted that the maximum updraft can be roughly equated to the expression $(2*CAPE)^{1/2}$. The maximum CAPE used in the model initial environments is 3407 J kg^{-1} which results in an estimated maximum updraft of 83 m s^{-1}. This value is close to the largest maximums shown in Figure 4.3(a) which are generally at or below 80 m s^{-1}.

[23] There are a small number of values attaining horizontal wind speeds greater than 100 m s^{-1}. This may be due to non-physical instabilities from within the model.

within a given simulation to identify the most extreme values generated within an individual simulation experiment. Figure 4.4 displays the initial environments (plotted by hodograph type as a function of BRN and 0-3 km SRH) containing the strongest vertical velocity maximum values above 2 km (left column) and strongest horizontal wind speed maximum values at the surface (right column). Initial environments producing the 10, 50 and 100 strongest values (out of 256 environments) are shown separately to highlight regions of the parameter space (Figure 3.3) generating the strongest storms. Figure 4.5 shows the same information but for values of vertical vorticity absolute maximums at the surface (left column) and pressure perturbation minimums at the surface (right column). The four quantities used in Figure 4.4 and Figure 4.5 provide a general representation of storm intensities attained within the initial environments.

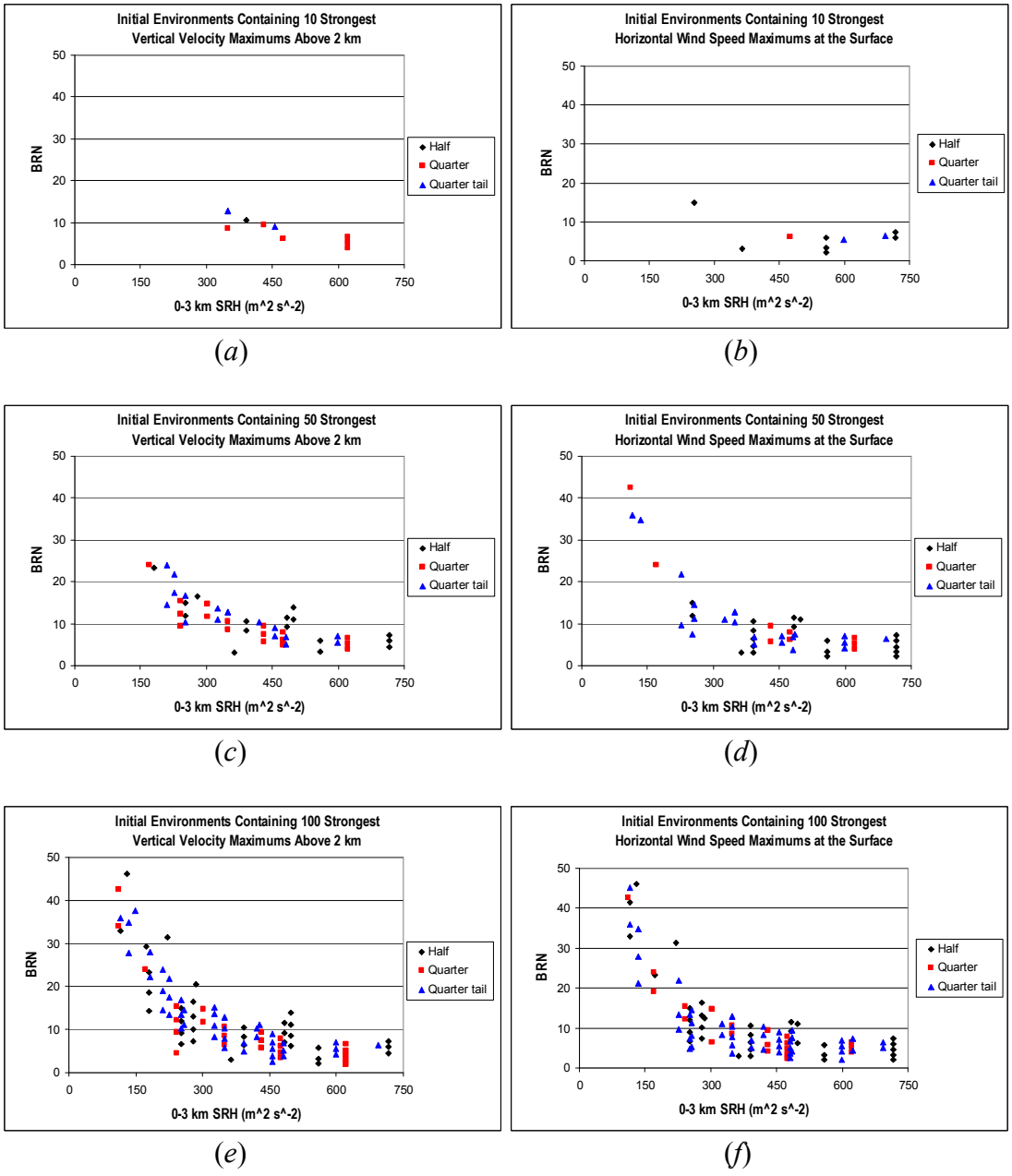


Figure 4.4: Initial environments associated with storms having the strongest vertical velocity maximums above 2 km for (*a*) 10 strongest (75 - 92 m s^{-1}), (*c*) 50 strongest (68 - 92 m s^{-1}) and (*e*) 100 strongest (63 - 92 m s^{-1}) as well as strongest horizontal wind speed maximums at the surface for (*b*) 10 strongest (73 - 112 m s^{-1}), (*d*) 50 strongest (53 - 112 m s^{-1}) and (*f*) 100 strongest (47 - 112 m s^{-1}). Initial environments are identified by their hodograph type and are plotted in terms of their BRN and 0-3 km SRH. For comparison to initial parameter space, see Figure 3.3.

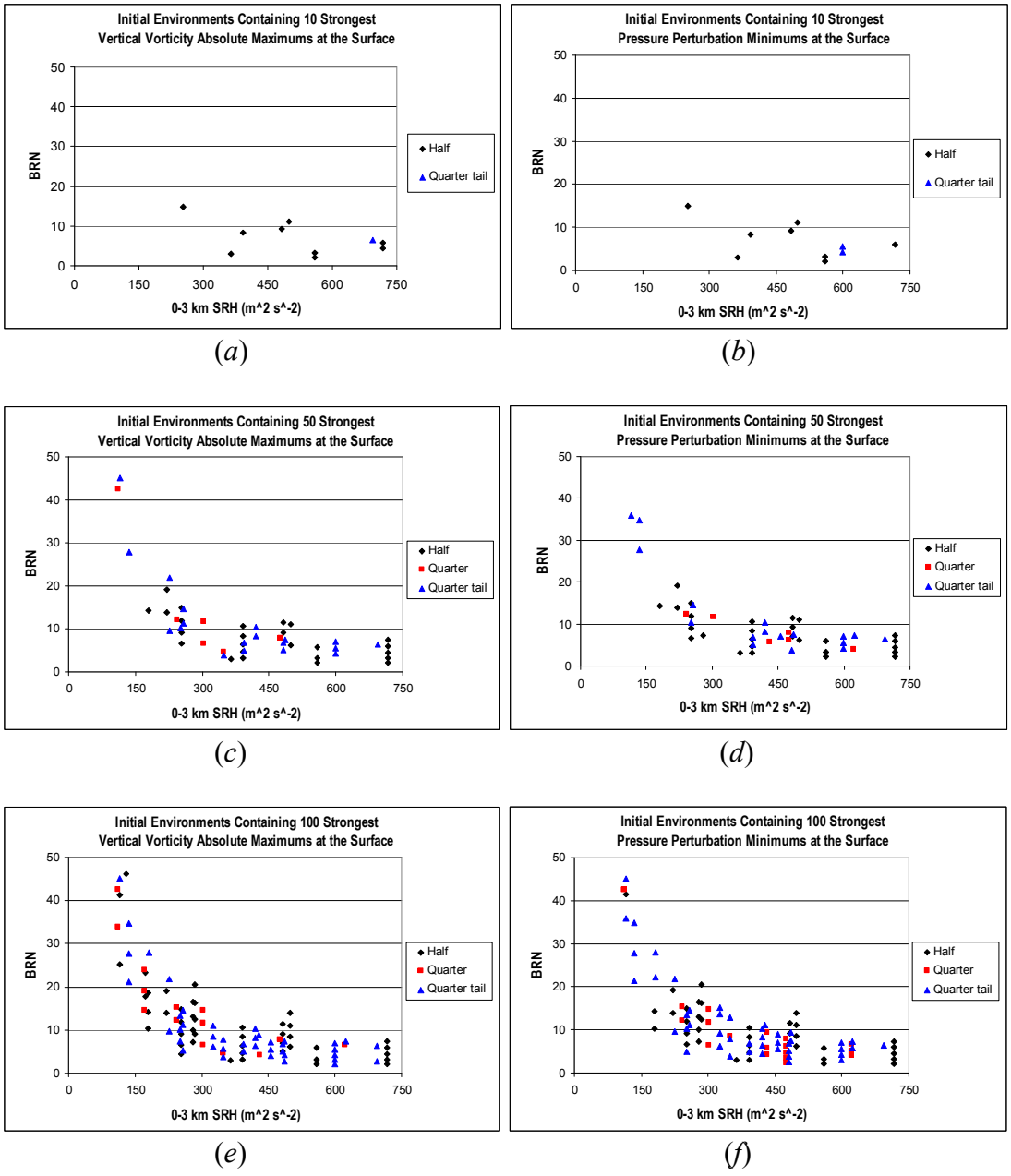


Figure 4.5: Same as Figure 4.4 but for initial environments associated with storms having the strongest vertical vorticity absolute maximums at the surface for (*a*) 10 strongest (0.12 to 0.15 s^{-1}), (*c*) 50 strongest (0.073 to 0.15 s^{-1}) and (*e*) 100 strongest (0.057 to 0.15 s^{-1}) as well as pressure perturbation minimum at the surface for (*b*) 10 strongest (-3182 to -4922 Pa), (*d*) 50 strongest (-1137 to -4922 Pa) and (*f*) 100 strongest (-810 to -4922 Pa).

It is evident that the environments in Figure 4.4 and Figure 4.5 containing the 10 strongest values generally occur in a similar region of the parameter space for all four quantities (e.g., BRN < 20, 0-3 km SRH > 200 m^2 s^{-2}). The 50 strongest tend to occur in similar environments as the 10 strongest but they also extend to environments of higher BRN and lower 0-3 km SRH (e.g., BRN < 45, 0-3 km SRH > 100 m^2 s^{-2}). The 100 strongest are in similar locations as the 50 strongest, filling in the surrounding region. Thus, relative to the defined parameter space (Figure 3.3), the strongest storms occur in initial environments of lower BRN and higher 0-3 km SRH. This is only a general statement for locating the strongest cases because simulations containing weaker storms are present within all regions of the parameter space (i.e., initial environments containing the 10 strongest quantity cases are directly adjacent to weaker cases shown in the 100 strongest quantity cases as well as weaker cases beyond the top 100 shown in Figure 3.3).

For comparison purposes, the example simulation experiment WK14_half_r15_4km used to illustrate the storm extraction process (Figures 4.1 and 4.2) used an initial environment with BRN = 6.17 and 0-3km SRH = 498 m^2 s^{-2} (see Appendix A). This simulation was identified as producing storms in the top 100 strongest experiments for updraft max above 2 km (Figure 4.4(e)) (max = 65 m s^{-1}) and horizontal wind speed at the surface (Figure 4.4(f)) (max = 51 m s^{-1}). It was also in the top 50 strongest experiments for vertical vorticity absolute maximum at the surface (Figure 4.5(c)) (max = 0.1 s^{-1}) and pressure perturbation minimum at the surface (Figure 4.5(d)) (min = -15.6 mb).

The initial environment differentiations can be extended to the generation of strong low-level rotation. Figure 4.6 presents the environments producing "Positive," "Intermediate" and "Negative" storms as well as environments producing no defined storms. The environments are labeled according to the highest rated storm generated within a simulation experiment in terms of the development of strong low-level rotation (i.e., highest to lowest ranking is in the following order: "Positive," "Intermediate," "Negative," "No Storm"). For example, if a simulation experiment had one "Intermediate" storm and one "Negative" storm it would be labeled "Intermediate." Further, if there were one "Positive" storm and three "Negative" storms the simulation experiment would be labeled "Positive."

Appendix A lists each simulation experiment (along with their initial environmental indices) and provides a count of the number of storms in each experiment that were identified as either "Positive," "Intermediate" or "Negative." Those with no defined storms (only two cases) obviously contain no storm counts. Referring to the example simulation used in Figure 4.1 (experiment WK14_half_r15_4km), the four identified storms are accounted for in Appendix A with one categorized as "Positive," one "Intermediate" and two "Negative." Therefore the simulation itself is labeled "Positive."

A majority of the simulation experiments in Figure 4.6 are identified as "Intermediate" cases (140) with relatively fewer "Negative" cases (69) and "Positive" cases (45). The remaining 2 simulation experiments did not contain a defined storm and therefore no maximum or minimum quantities were extracted from their model

output. They are shown here for comparison purposes only. Comparing the "Positive," "Intermediate" and "Negative simulation experiment totals (45, 140 and 69) to the storm totals (58, 373 and 737), respectively, indicates that only a limited number of simulations contained more than one "Positive" storm while multiple "Intermediate" and especially "Negative" storms were present in a large number of the experiments (see storm counts in Appendix A).

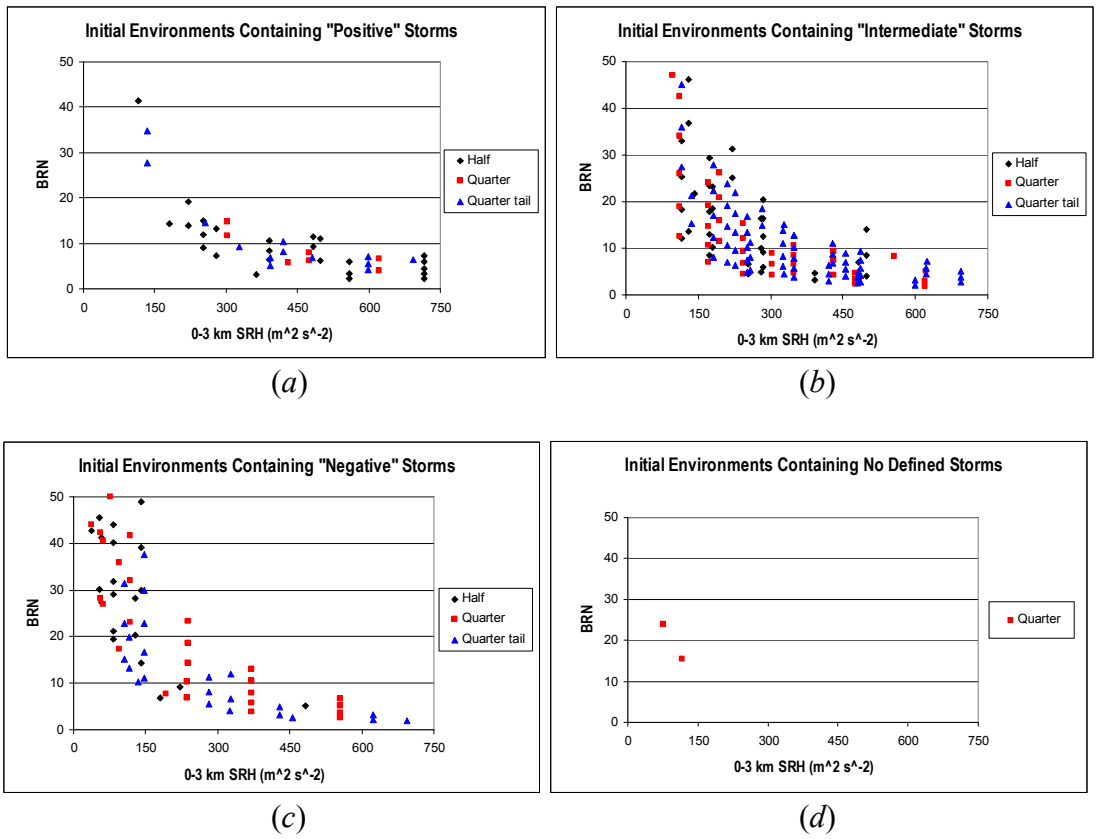


Figure 4.6: Initial environments in which simulation experiments generated storms identified as (*a*) "Positive" (45 cases) (*b*) "Intermediate" (140 cases) and (*c*) "Negative" (69 cases) with regards to the development of strong low-level rotation as well as (*d*) those environments producing no defined storms. Initial environments are given in terms of BRN and 0-3 km SRH and are labeled according to their hodograph type.

There exists general agreement between the initial environments generating "Positive" storms in Figure 4.6(a) and those environments generating the top 10 and top 50 strongest storms in Figure 4.4 and Figure 4.5. A majority of the "Positive" storms occur in initial environments of BRN < 20 and 0-3 km SRH >150 m^2 s^{-2}, which also is true of the environments containing the top 10 strongest storms in Figure 4.4(a),(b) and Figure 4.5(a),(b). The parameter space region encompassing the "Positive" cases, however, is not exclusively "Positive" because "Intermediate" cases (Figure 4.6(b)) and "Negative" cases (Figure 4.6(c)) also are present in this region. There are also instances of "Positive" storm cases occurring outside of this region (e.g., three cases with BRN > 25). The "Intermediate" and "Negative" cases each span most of the parameter space with "Intermediate" cases slightly more concentrated at lower BRN's (BRN < 30) and between 150 m^2 s^{-2} < 0-3 km SRH < 500 m^2 s^{-2} compared to "Negative" cases slightly more concentrated at relatively lower 0-3 km SRH (0-3 km SRH < 200 m^2 s^{-2}). The two initial environments not producing a storm strong enough to be tracked were located in a similar region of the parameter space with low BRN and low 0-3 km SRH.

A general summary of the initial environments generating "Positive," "Intermediate" and "Negative" storms is provided in Figure 4.7.[24] The significant overlap between the three regions supports the chosen parameter space because most of the simulation experiments generally were capable of producing strong low-level rotation. Inclusion of environments generating "Intermediate" and "Negative" cases

[24] The "No Storm" initial environments are excluded from this general summary because they are not included in the data mining analysis.

but not "Positive" cases are still acceptable because non-"Positive" storms are needed in the data mining analysis to distinguish feature differences between storms that do and do not produce strong low-level rotation. Future work may consider further constraining the parameter space to test for any variability in the data mining results, but the wider range applied here provides a thorough initial analysis.

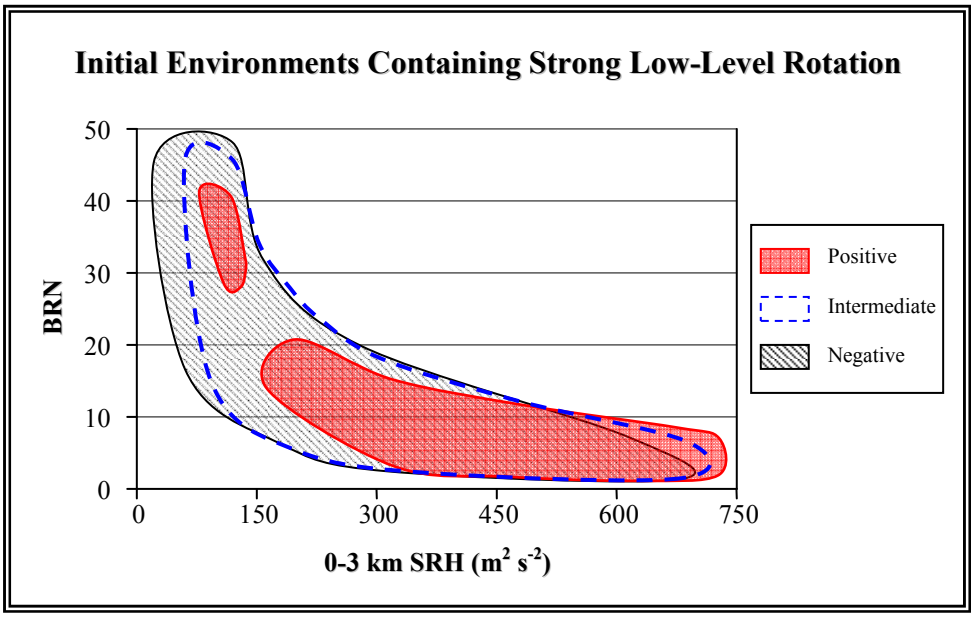


Figure 4.7: General representation of initial model environments associated with strong low-level rotation identifiers of "Positive" (red shaded), "Intermediate" (blue dashed) and "Negative" (diagonal lined) as displayed in Figure 4.6. Environments are given in terms of their BRN and 0-3 km SRH corresponding to the initial parameter space given in Figure 3.3.

4.2 Data Mining Results

The data mining procedure discussed in Section 3.3.3 is used to analyze the extracted meteorological quantities in an attempt to identify interrelated, repeated patterns occurring prior to the development of strong low-level rotation. Section

4.2.1 highlights the top rules generated when applying the data mining analysis to the 58 "Positive" and 737 "Negative" storms only, while Section 4.2.2 provides results from the 58 "Positive" and 373 "Intermediate" storms only.

4.2.1 Rules Generated Comparing "Positive" and "Negative" Storms

The results from the data mining discretization scheme parameter variation analysis for "Positive" and "Negative" storms are presented in Table 4.1. The schemes are listed according to alphabet size, word size and averaging interval and include the performance measures of POD, FAR, and CSI for the top rule identified when analyzing all 58 "Positive" and 737 "Negative" storms as well as for the rule with the top average value over 10 folds when using 10-fold cross-validation. The schemes are ranked according to their 10-fold cross-validation scores because it provides a more robust and representative result. The top 10 discretization scheme parameter variation results are highlighted in light blue. These top 10 schemes contain a wide variation of discretization setups. Five out of the six alphabet size variations are represented in the top 10 schemes as well as both word sizes (two and three) and all five averaging intervals. There exists a higher concentration of schemes in the top 10 containing an alphabet size of five and six letters but the number two and three rated schemes are not in this grouping.

Table 4.1: Discretization scheme parameter variations applied to the data mining analysis for "Positive" and "Negative" storm cases. Schemes are listed according to variations in alphabet size, word size and averaging interval and include the performance measures of POD, FAR, and CSI for the top rule identified when analyzing all "Positive" and "Negative" storms and for the average value (across all 10 folds) when using 10-fold cross-validation. Discretization schemes are rated according to 10-fold cross-validation CSI with the top 10 schemes highlighted in light blue.

Alpha Size	Word Size	Avg Interval	All "Pos" and "Neg" Storms			10-Fold Cross-Validation			Rank
			POD	FAR	CSI	POD	FAR	CSI	
3	2	1	0.808	0.401	0.523	0.800	0.420	0.514	44
3	2	2	0.823	0.381	0.543	0.760	0.433	0.460	50
3	2	3	0.886	0.366	0.586	0.810	0.438	0.497	46
3	2	4	0.831	0.249	0.651	0.813	0.251	0.622	17
3	2	5	0.825	0.164	0.710	0.790	0.220	0.652	10
3	3	1	0.814	0.390	0.533	0.743	0.411	0.483	49
3	3	2	0.856	0.324	0.606	0.836	0.293	0.590	25
3	3	3	0.819	0.188	0.688	0.746	0.236	0.575	29
3	3	4	0.821	0.183	0.693	0.690	0.320	0.521	42
3	3	5	0.812	0.143	0.714	0.746	0.240	0.603	20
4	2	1	0.885	0.300	0.641	0.850	0.351	0.583	28
4	2	2	0.848	0.244	0.665	0.793	0.338	0.559	31
4	2	3	0.821	0.146	0.720	0.776	0.147	0.688	3
4	2	4	0.810	0.163	0.699	0.720	0.257	0.553	34
4	2	5	0.812	0.108	0.738	0.746	0.141	0.650	11
4	3	1	0.860	0.240	0.676	0.836	0.268	0.634	13
4	3	2	0.817	0.159	0.707	0.746	0.284	0.587	27
4	3	3	0.823	0.169	0.704	0.726	0.175	0.601	22
4	3	4	0.810	0.142	0.713	0.640	0.238	0.485	48
4	3	5	0.812	0.165	0.698	0.686	0.235	0.530	40
5	2	1	0.865	0.153	0.748	0.823	0.190	0.684	4
5	2	2	0.827	0.119	0.743	0.773	0.181	0.665	8
5	2	3	0.818	0.118	0.737	0.726	0.154	0.667	7
5	2	4	0.812	0.151	0.709	0.673	0.234	0.551	36
5	2	5	0.817	0.170	0.700	0.720	0.217	0.599	23
5	3	1	0.827	0.123	0.740	0.823	0.128	0.704	1
5	3	2	0.810	0.147	0.710	0.656	0.275	0.505	45
5	3	3	0.817	0.126	0.731	0.703	0.224	0.602	21
5	3	4	0.812	0.147	0.711	0.680	0.258	0.528	41
5	3	5	0.808	0.178	0.686	0.743	0.195	0.598	24

Table 4.1: Continued

Alpha Size	Word Size	Avg Interval	All "Pos" and "Neg" Storms POD	FAR	CSI	10-Fold Cross-Validation POD	FAR	CSI	Rank
6	2	1	0.827	0.099	0.758	0.796	0.184	0.684	5
6	2	2	0.821	0.110	0.745	0.693	0.163	0.609	19
6	2	3	0.817	0.110	0.741	0.670	0.254	0.552	35
6	2	4	0.814	0.159	0.704	0.636	0.282	0.516	43
6	2	5	0.814	0.108	0.740	0.630	0.131	0.553	33
6	3	1	0.829	0.103	0.757	0.753	0.126	0.683	6
6	3	2	0.812	0.129	0.724	0.723	0.268	0.587	26
6	3	3	0.821	0.132	0.729	0.783	0.214	0.654	9
6	3	4	0.817	0.123	0.733	0.700	0.142	0.611	18
6	3	5	0.814	0.196	0.678	0.693	0.224	0.544	38
7	2	1	0.819	0.098	0.752	0.700	0.206	0.573	30
7	2	2	0.827	0.102	0.755	0.670	0.260	0.547	37
7	2	3	0.817	0.109	0.742	0.643	0.196	0.543	39
7	2	4	0.819	0.124	0.734	0.626	0.262	0.492	47
7	2	5	0.819	0.090	0.757	0.693	0.258	0.558	32
8	2	1	0.827	0.122	0.741	0.743	0.167	0.627	16
8	2	2	0.818	0.074	0.767	0.713	0.153	0.630	15
8	2	3	0.835	0.103	0.761	0.763	0.147	0.643	12
8	2	4	0.817	0.056	0.779	0.756	0.117	0.693	2
8	2	5	0.819	0.103	0.748	0.726	0.150	0.633	14

The top rated discretization scheme shown in Table 4.1 has an alphabet size of five, word size of three and averaging interval of one (no averaging) which can be identified as scheme 531. Discretization scheme identifiers will be labeled using this same three number method for the remainder of the discussion. Table 4.2 provides a list of the top 10 rules from scheme 531 (according to CSI) when analyzing all available "Positive" and "Negative" storms. The quantities constituting each rule are listed in sequential order in the right-hand column of Table 4.2. Recall from Section 3.3.3 that words originating from different quantities can occur at the same time and

therefore the term sequential is interpreted to mean that no quantity can occur prior to the quantity listed before it.

Table 4.2: Top 10 data mining rules (according to CSI) identified for top rated discretization scheme (531) when analyzing all available "Positive" and "Negative" storms (Table 4.1). Meteorological quantities within a rule are listed in the order in which they occur.

Rule #	***CSI / POD / FAR***	***Rule***
1	CSI = 0.74 POD = 0.83 FAR = 0.13	- baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km - pressure perturbation vertical gradient force max above 2 km - vertical stretching term min below 2 km
2	CSI = 0.74 POD = 0.83 FAR = 0.13	- baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km - vertical stretching term min below 2 km - pressure perturbation vertical gradient force min above 2 km
3	CSI = 0.72 POD = 0.81 FAR = 0.13	- baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km - vertical stretching term min below 2 km - pressure perturbation vertical gradient force max above 2 km
4	CSI = 0.72 POD = 0.81 FAR = 0.13	- baroclinic generation term (vertical) min below 2 km - vertical stretching term min below 2 km - pressure perturbation vertical gradient force min above 2 km - vertical stretching term max below 2 km
5	CSI = 0.72 POD = 0.83 FAR = 0.16	- baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km - tilting term max below 2 km - pressure perturbation vertical gradient force min above 2 km
6	CSI = 0.71 POD = 0.84 FAR = 0.18	- baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km - pressure perturbation vertical gradient force max below 2 km - pressure perturbation vertical gradient force min above 2 km
7	CSI = 0.71 POD = 0.83 FAR = 0.17	- baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km - pressure perturbation vertical gradient force max above 2 km - pressure perturbation vertical gradient force max below 2 km
8	CSI = 0.70 POD = 0.86 FAR = 0.21	- baroclinic generation term (vertical) min below 2 km - pressure perturbation vertical gradient force max above 2 km - pressure perturbation vertical gradient force max below 2 km - vertical stretching term min below 2 km
9	CSI = 0.70 POD = 0.81 FAR = 0.16	- pressure perturbation vertical gradient force max above 2 km - baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km - vertical stretching term min below 2 km
10	CSI = 0.70 POD = 0.81 FAR = 0.16	- vertical stretching term min below 2 km - baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km - tilting term max below 2 km

To illustrate an individual rule sequence, Figure 4.8 provides an expanded view of Rule # 1 from Table 4.2. The left column identifies the discretized, 3-letter meteorological quantity words forming the rule in sequential order from top to bottom (4-word rule in this example). The three letters making up each word are plotted in their respective equiprobable Gaussian regions with each region demarcated by light gray horizontal lines. Each colored horizontal line segment represents the 30 s time period corresponding with each letter. There is no actual time separation between letters in an individual word (x-axis units are arbitrary): the separation is included simply to better distinguish each three-letter sequence. The right column of Figure 4.8 provides an example storm in which this rule is found. This is useful because the left column word sequence does not provide the time separation between each word so it is unclear when the words appear in the actual metadata.

Plotted in the right column is the time series for Storm 1 of WK14_half_r15_4km (from previous discussion associated with Figure 4.1 and Figure 4.2) which is a "Positive" storm containing the highlighted rule. Time series are given (in 30 s output intervals) for each quantity of the rule over the entire storm lifetime. The 30 minute period immediately prior to the development of strong low-level rotation used for the data mining analysis is marked with two vertical gray lines. The exact location of each word in the rule is indicated with a red line segment superimposed on the black time series. It should be noted that the example storm shown is simply one example out of the many in which the rule appears. Other storms may contain the same rule but have slightly different time separation between

the words composing the rule. Recall that the time separation between words is not included in the data mining process and therefore the exact timing of the words shown in this example storm, and future examples, should not be interpreted as being part of the official rule. Words simply need to be present at the same time or following the word listed before it in the left column.

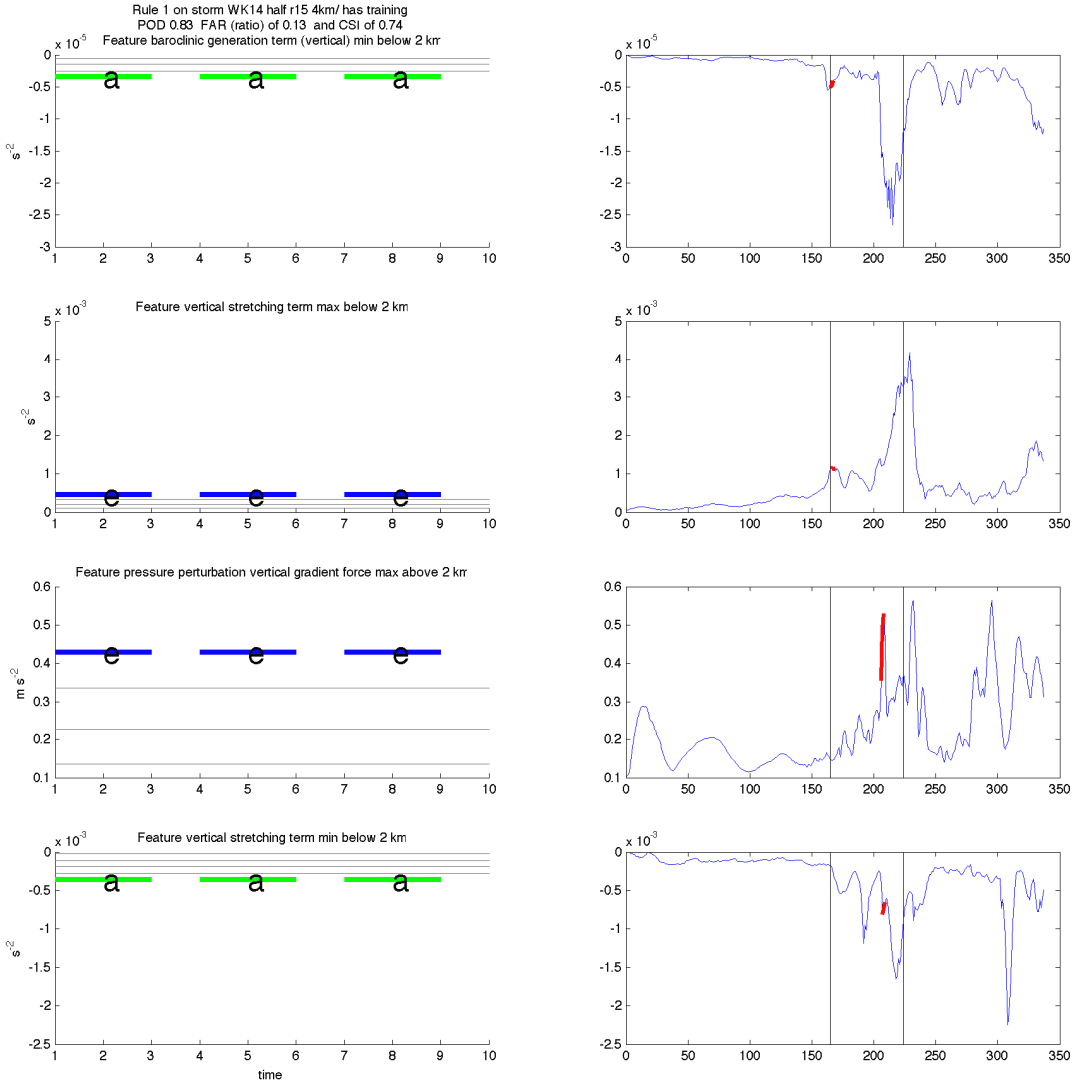


Figure 4.8: Rule #1 from Table 4.2. Left column identifies 3-letter words comprising the rule in sequential order (baroclinic generation term (vertical) min below 2 km, vertical stretching term max below 2 km, pressure perturbation vertical gradient force max above 2 km and vertical stretching term min below 2 km). Time step within each word plot in left column is arbitrary but colored line letter segments correspond to a 30 s time period (one output interval). The five equiprobable Gaussian regions associated with each word are demarcated by light gray horizontal lines. Right column provides meteorological quantity metadata (30 s output intervals) from an example storm (Storm 1 (blue) from WK14_half_r15_4km) that contains the rule. The metadata are given in number of 30 s output intervals, beginning when the storm was initially identified (e.g., not the same as simulation time). Each word from the left column is identified by a red line segment in the metadata. The 30 minute window prior to the development of strong low-level rotation is contained within the two vertical gray lines. Corresponding performance measures are listed at the top of the figure.

It is apparent from the rule given in Figure 4.8 that all letters within each word are located in one of two *extreme* equiprobable Gaussian regions (e.g., regions "a" and "e"). This characteristic was discovered in all generated rules over the entire study and was a result not originally anticipated. When developing the methodology, it was assumed that a sequence of quantity (letter) variations would be identified in each individual quantity (each word) and then multiple sets of quantity variations over different quantities (different words) would be linked together into one larger sequence (sequence of words) that would be associated with the development of strong low-level rotation. The results show that a rule sequence was generated, but the individual quantities (individual words) within each rule are made up of extreme values rather than a sequence of quantity variations. This implies that attaining an extreme value is more significant than the sequence of events leading to that extreme value for an individual quantity. Each extreme quantity value is still linked together with other quantities to form a sequence of events, but these events can now be interpreted as a set of individual quantities reaching extremes. In other words, the rules consist of a set of meteorological quantities passing an extreme threshold in a particular sequence. This makes sense qualitatively because strong low-level rotation is an extreme event and therefore any features associated with it can be expected to have extreme attributes.

The first two words of the top rule shown in Figure 4.8 (baroclinic generation term (vertical) min below 2 km and vertical stretching term max below 2 km) reach relatively extreme values and then roughly 20 minutes later, the third and fourth

words in the rule (pressure perturbation vertical gradient force max above 2 km and vertical stretching term min below 2 km) reach relatively extreme values. The combination of these four quantities reaching extreme values is the interrelated repeated pattern identified as unique to those instances when strong low-level rotation develops in the simulations. As the performance measures show, this sequence of events occurred in 83% of all storms developing strong low-level rotation (POD), and when this sequence was observed within a storm, only 13% of the time did strong low-level rotation not occur (FAR).

Considering the top rules for scheme 531 in Table 4.2, it is apparent that the rules are made up of similar meteorological quantities (similar words) simply arranged in a different sequential order. This is a characteristic common to all of the top discretization schemes and seems logical given that each word represents a quantity reaching an extreme value. Quantities simply need to reach extreme values in slightly different orders to generate multiple rules containing the same extreme quantities.

The top rule shown in Figure 4.8, and the other rules shown in Table 4.2, result only from discretization scheme 531. The other schemes listed in Table 4.1 also have their own set of rules which can potentially vary from those found using scheme 531. To illustrate the difference in rules generated from different discretizations schemes, Table 4.3 provides the top 10 rules for the second rated scheme from Table 4.1 (scheme 824). All of the top 10 rules for scheme 824 differ slightly from the top 10 rules of scheme 531 (compare Table 4.3 and Table 4.2).

Many of the same extreme quantities are found in both sets of rules, but there also exist additional quantities unique to particular schemes.

For example, the baroclinic generation term (vertical) min below 2 km is present in every one of the top 10 rules of scheme 531 but is not present in any of the top 10 rules of scheme 824. Also, vertical velocity max at 4 km and horizontal divergence min below 2 km are in many of the top 10 rules for scheme 824 but are not present in any of the top 10 rules for scheme 531. This apparent discrepancy between schemes is attributed to how the discretization methods alter the original model output. An important difference between the two schemes discussed thus far is that scheme 824 averages four 30 s output intervals together (2 min intervals) while scheme 531 averages one output interval (e.g., no averaging). This changes the metadata from which the data mining analysis searches and can alter the quantity variations that best indicate the development of strong low-level rotation. As will be shown later in the discussion, however, when looking over the results of a number of data mining schemes, certain meteorological quantities repeatedly stand out as important to the development of strong low-level rotation.

Table 4.3: Top 10 data mining rules (according to CSI) identified for the number two rated discretization scheme (824) when analyzing all available "Positive" and "Negative" storms (Table 4.1). Meteorological quantities within a rule are listed in the order in which they occur.

Rule #	*CSI / POD / FAR*	*Rule*
1	CSI = 0.78 POD = 0.81 FAR = 0.04	- vertical velocity max at 4 km - pressure perturbation vertical gradient force max below 2 km - vertical stretching term max below 2 km - horizontal divergence min below 2 km
2	CSI = 0.77 POD = 0.81 FAR = 0.06	- horizontal divergence min below 2 km - pressure perturbation vertical gradient force max below 2 km - vertical stretching term max below 2 km - horizontal divergence min below 2 km
3	CSI = 0.76 POD = 0.83 FAR = 0.09	- vertical velocity max at 4 km - pressure perturbation vertical gradient force max below 2 km - vertical stretching term max below 2 km - vertical velocity horizontal gradient max below 2 km
4	CSI = 0.76 POD = 0.81 FAR = 0.08	- vertical velocity max at 4 km - pressure perturbation vertical gradient force max below 2 km - vertical stretching term max below 2 km - pressure perturbation vertical gradient force max below 2 km
5	CSI = 0.75 POD = 0.83 FAR = 0.11	- horizontal divergence min below 2 km - pressure perturbation vertical gradient force min below 2 km - pressure perturbation vertical gradient force max below 2 km - vertical velocity horizontal gradient max below 2 km
6	CSI = 0.74 POD = 0.83 FAR = 0.13	- vertical velocity max at 4 km - pressure perturbation vertical gradient force max below 2 km - vertical stretching term max below 2 km
7	CSI = 0.72 POD = 0.83 FAR = 0.16	- pressure perturbation vertical gradient force max below 2 km - vertical stretching term max below 2 km - vertical velocity horizontal gradient max below 2 km
8	CSI = 0.71 POD = 0.81 FAR = 0.15	- vertical velocity max at 4 km - pressure perturbation vertical gradient force min below 2 km - pressure perturbation vertical gradient force max below 2 km
9	CSI = 0.71 POD = 0.81 FAR = 0.15	- vertical velocity max at 4 km - vertical stretching term max below 2 km - pressure perturbation vertical gradient force max below 2 km
10	CSI = 0.70 POD = 0.86 FAR = 0.21	- pressure perturbation vertical gradient force min below 2 km - horizontal divergence min below 2 km - vertical velocity horizontal gradient max below 2 km

It also should be noted that the top rules for the second rated scheme (842) in Table 4.3 appear to have higher performance measure values than the top rated discretization scheme (531) shown in Table 4.2. This is because discretization schemes in Table 4.1 were rated based on the 10-fold cross validation procedure while the top 10 listed rules in Table 4.2 and Table 4.3 are the results of the data mining analysis applied to all "Positive" and "Negative" storms. Therefore, when the storms were evenly distributed into 10 separate subgroups and analyzed using scheme 842, the averaged CSI score was below the average score of scheme 531 even though scheme 842 scored higher when the analysis was performed on all "Positive" and "Negative" storms. This implies that the top rated rules from scheme 531 are more consistent across the "Positive" and "Negative" storms than scheme 842.

The top rule from scheme 824 (Table 4.3) is provided in Figure 4.9. Because scheme 824 splits the metadata into eight equiprobable Gaussian regions, the additional region separators (gray lines) are shown in the left column (compare with five regions from scheme 531 in Figure 4.8. Also, scheme 824 contains only two letters in each word compared to three letters per word in scheme 531. According to the example storm in the right hand column, the rule begins with an extreme value of vertical velocity max at 4 km and then approximately 15 minutes prior to the development of strong low-level rotation, a sequence of three quantities – all acquiring relatively extreme levels – (pressure perturbation vertical gradient force max below 2 km, vertical stretching term max below 2 km, horizontal divergence min below 2 km) develops. Again, this simply is an example of the rule so the exact

timing of the quantities reaching extreme levels prior to the development of strong low-level rotation may not be the same in other storms.

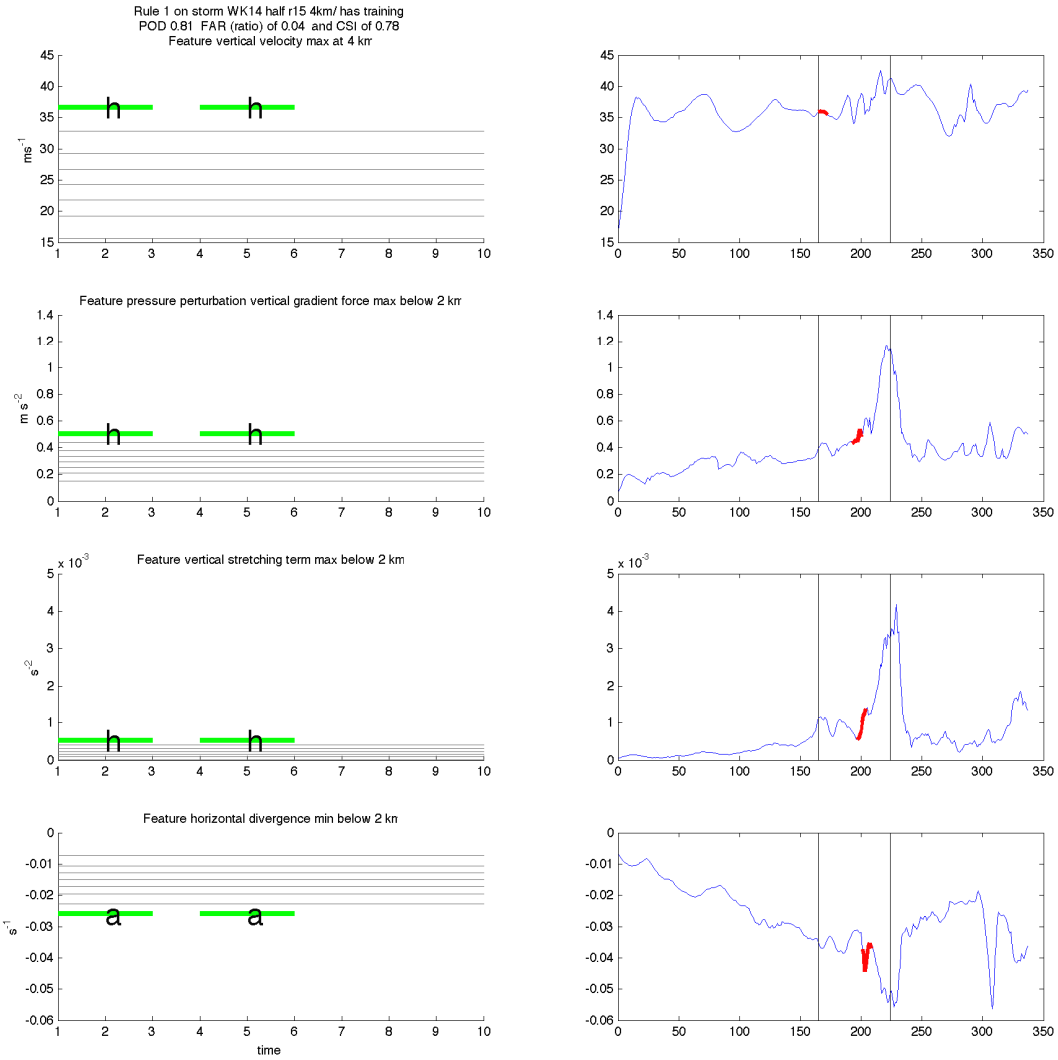


Figure 4.9: Same as Figure 4.8 but for top rule found in discretization scheme 824 (Table 4.3).

Another rule from scheme 824 contains a specific characteristic worth noting. Figure 4.10 provides the fourth rated rule from Table 4.3. The second and fourth quantities within this rule are actually the same quantity (pressure perturbation vertical gradient force max below 2 km). Having a word repeated in the same rule is allowed because the data mining procedure searches all potential word combinations, which includes words that are repeated in the same quantity as long as they are sequentially separated (see discussion associated with Figure 3.10 in Section 3.3.3). The presence of the repeated word in this rule provides additional insight on an evolving storm containing the rule. The first word indicates the presence of an extreme value of vertical velocity maximum at 4 km. An extreme value of pressure perturbation vertical gradient force maximum below 2 km (associated with a decrease in pressure aloft which strengthens the updraft) develops, which increases vertical stretching of vertical vorticity in the updraft (vertical stretching term maximum below 2 km). The increased vertical vorticity and rotation further decrease pressure aloft, which increases the vertical perturbation pressure gradient force max below 2 km.

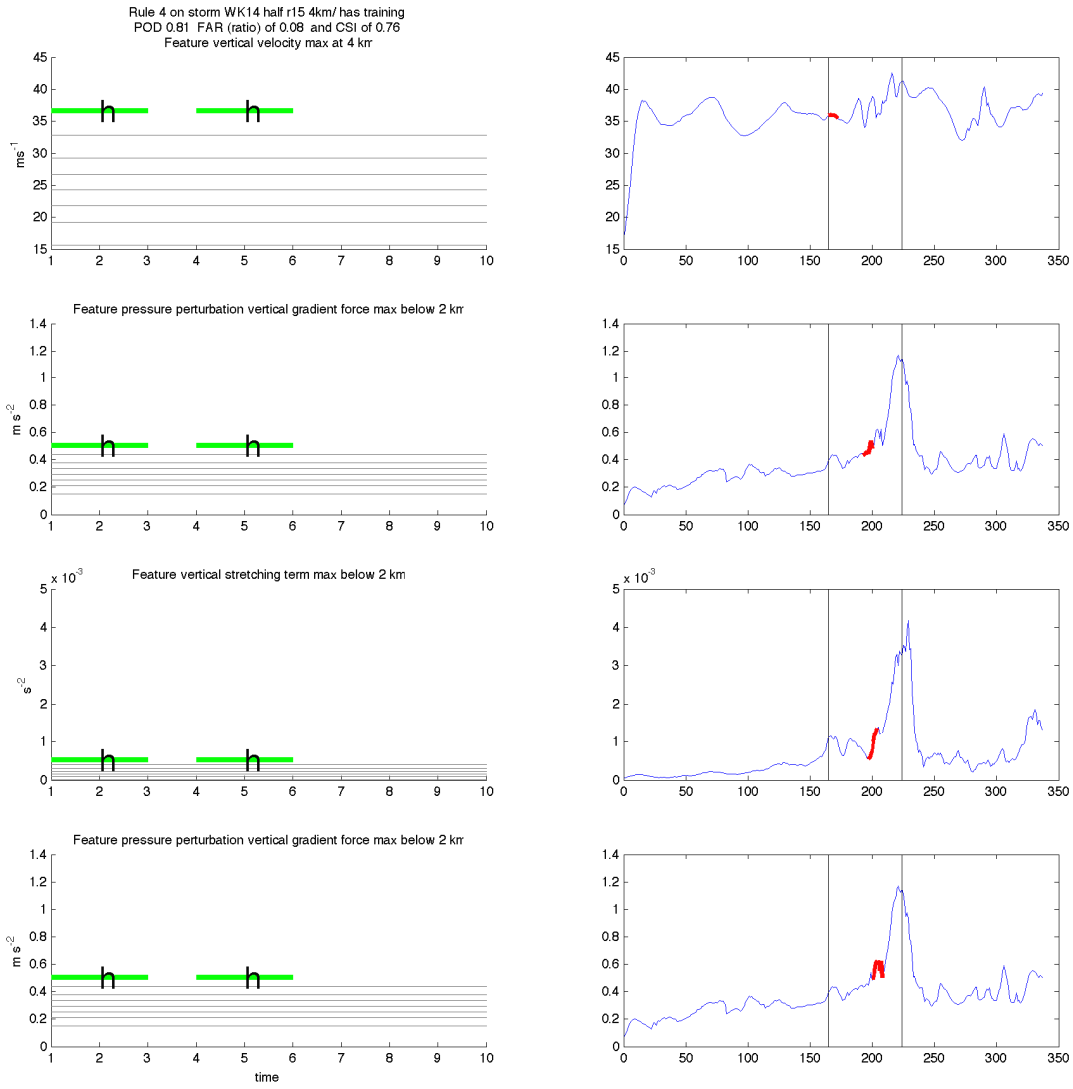


Figure 4.10: Same as Figure 4.8 but for the fourth rated rule of discretization scheme 824 (Table 4.3).

The example rules shown thus far are just a few of the hundreds of rules generated over the various discretization schemes. Because the top rules within particular discretization schemes contain similar extreme quantities in different orders, it is useful to identify those extremes occurring most often in the top rules to establish which are the most significant precursors to the development of strong low-level rotation. Figure 4.11 and Table 4.4 provide frequency totals for the number of times an extreme quantity appeared in the top 10 rules from each of the top three data mining discretization parameter variations for all "Positive" and "Negative" storms shown in Table 4.1 (schemes 531, 824 and 423). A total of 11 different extreme quantities were identified in these top rules with four or five standing out as occurring appreciably more often than the others. For a description of how these quantities are related to an evolving supercell storm, refer to Table 4.8 later in the discussion.

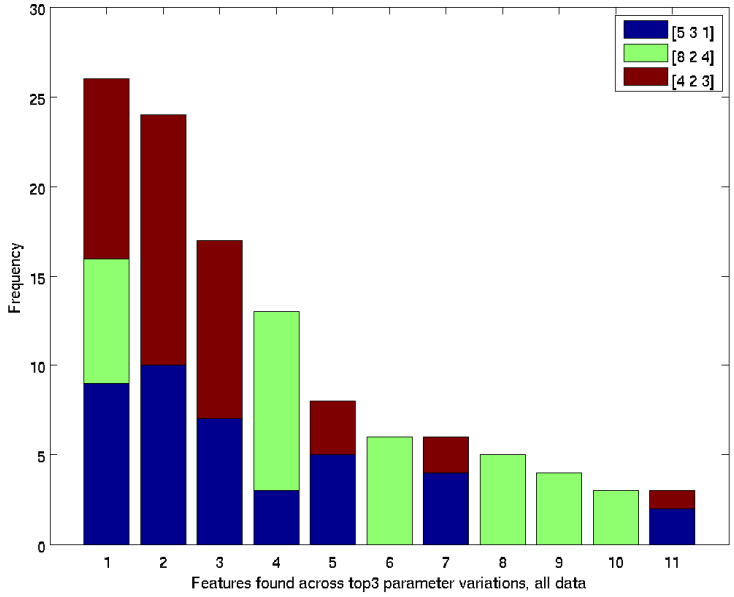


Figure 4.11: Frequency totals for top 10 rules within the top three data mining discretization parameter variations schemes (531, 824 and 423) for all "Positive" and "Negative" storms from Table 4.1. Specific meteorological quantities associated with numbers 1 – 11 are given in Table 4.4.

Table 4.4: Meteorological quantities and frequency totals corresponding to Figure 4.11. Frequency totals are given for each of the top three discretization schemes (531, 824 and 423) as well as their combined totals.

Rank	Meteorological Quantity	Frequency			
		Total	531	824	423
1	vertical stretching term max below 2 km	26	9	7	10
2	baroclinic generation term (vertical) min below 2 km	24	10	0	14
3	vertical stretching term min below 2 km	17	7	0	10
4	pressure perturbation vertical gradient force max below 2 km	13	3	10	0
5	pressure perturbation vertical gradient force max above 2 km	8	5	0	3
6	vertical velocity max at 4 km	6	0	6	0
7	pressure perturbation vertical gradient force min above 2 km	6	4	0	2
8	horizontal divergence min below 2 km	5	0	5	0
9	vertical velocity horizontal gradient max below 2 km	4	0	4	0
10	pressure perturbation vertical gradient force min below 2 km	3	0	3	0
11	tilting term max below 2 km	3	2	0	1

Figure 4.12 and Table 4.5 provide the same frequency total information as Figure 4.11 and Table 4.4 only for the top 10 rules within the top 10 data mining discretization parameter variation schemes (531, 824, 423, 521, 621, 631, 523, 522, 633, 325). Frequencies within the top 10 schemes are significantly greater than those for the top three schemes because they include all words used to generate the top 10 rules of 10 different schemes. Even with the greater number of words and wide variation of discretization schemes, the quantities identified across the top 10 schemes are quite similar to those found in the top three schemes (Table 4.4). To better illustrate this point, alongside the quantity rankings for the top 10 schemes in Table 4.5 is the quantity's ranking from the top three schemes from Table 4.4 shown in subscript brackets. The most significant quantities from the top three discretization schemes remain the most significant quantities for the top 10 schemes with only slight variations in ordering. A few additional quantities appear in the top 10 schemes but these have relatively small frequency totals. Therefore it can be concluded that the top rules from the top discretization schemes for all "Positive" and "Negative" storms generally were composed of similar extreme quantities.

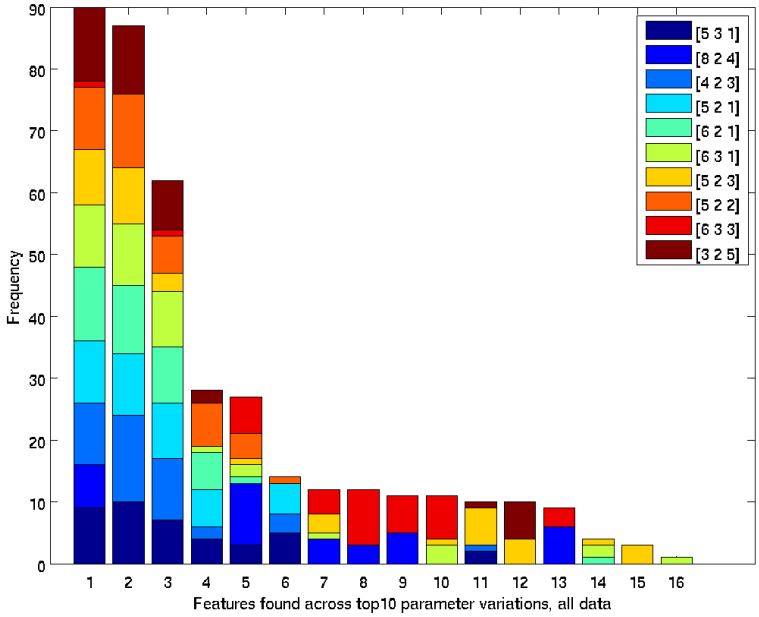


Figure 4.12: Frequency totals for top 10 rules within the top ten data mining discretization parameter variations schemes (531, 824, 423, 521, 621, 631, 523, 522, 633, 325) from Table 4.1 for all "Positive" and "Negative" storms. Specific meteorological quantities associated with numbers 1 – 16 are given in Table 4.5.

Table 4.5: Meteorological quantities and frequency totals corresponding to Figure 4.12. Frequency totals are given for each of the top ten discretization schemes (531, 824, 423, 521, 621, 631, 523, 522, 633, 325) as well as their combined totals. Meteorological quantity rankings over the top three schemes (from Table 4.4) are given in subscript brackets next to the rankings over the top 10 schemes.

Rank	Meteorological Quantity	Frequency	
		Total	Top 10 Variations
1 [1]	vertical stretching term max below 2 km	90	9,7,10,10,12,10,9,10,1,12
2 [2]	baroclinic generation term (vertical) min below 2 km	87	0,14,10,11,10,9,12,0,11
3 [3]	vertical stretching term min below 2 km	62	7, 0, 10, 9, 9, 9, 3, 6, 1, 8
4 [7]	pressure perturbation vertical gradient force min above 2 km	28	4, 0, 2, 6, 6, 1, 0, 7, 0, 2
5 [4]	pressure perturbation vertical gradient force max below 2 km	27	3, 10, 0, 0, 1, 2, 1, 4, 6, 0
6 [5]	pressure perturbation vertical gradient force max above 2 km	14	5, 0, 3, 5, 0, 0, 0, 1, 0, 0
7 [9]	vertical velocity horizontal gradient max below 2 km	12	0, 4, 0, 0, 0, 1, 3, 0, 4, 0
8 [10]	pressure perturbation vertical gradient force min below 2 km	12	0, 3, 0, 0, 0, 0, 0, 0, 9, 0
9 [8]	horizontal divergence min below 2 km	11	0, 5, 0, 0, 0, 0, 0, 0, 6, 0
10	vertical stretching term max above 2 km	11	0, 0, 0, 0, 0, 3, 1, 0, 7, 0
11 [11]	tilting term max below 2 km	10	2, 0, 1, 0, 0, 0, 6, 0, 0, 1
12	baroclinic generation term (vertical) max below 2 km	10	0, 0, 0, 0, 0, 0, 4, 0, 0, 6
13 [6]	vertical velocity max at 4 km	9	0, 6, 0, 0, 0, 0, 0, 0, 3, 0
14	vertical velocity max below 2 km	4	0, 0, 0, 0, 1, 2, 1, 0, 0, 0
15	vertical stretching term min above 2 km	3	0, 0, 0, 0, 0, 0, 3, 0, 0, 0
16	tilting term min below 2 km	1	0, 0, 0, 0, 0, 1, 0, 0, 0, 0

To test the robustness of the results, the data mining analysis is rerun on all "Positive" and "Negative" storms using 10-fold cross-validation. Figure 4.13 and Table 4.6 provide the same frequency results for the top three schemes over all "Positive" and "Negative" storms as shown in Figure 4.11 and Table 4.4 only for the 10-fold cross-validation procedure. Frequencies are tabulated by adding up the number of times a specific extreme quantity occurred in the top 10 rules of the 10-fold subgroups from each of the top 3 schemes, which is why there are significantly greater frequencies. The rank of each quantity is given in Table 4.13 as well as the original rank (given in subscript brackets) of the same quantity from mining across all "Positive" and "Negative" storms (from Table 4.4). There is good agreement between the analysis using all "Positive" and "Negative" storms and the 10-fold cross-validation analysis. The top 11 quantities remain ranked in the top 11 with only a single place movement of some of the quantities.

There are a number of additional quantities present in the 10-fold cross-validation analysis but they contain relatively small frequencies. These additional quantities are to be expected given that separate analyses are performed on each of the 10 subgroups within each discretization scheme. The most significant results of this 10-fold cross-validation analysis are the reappearance of the top quantities containing the highest frequencies. This strengthens the confidence for the top quantities to be regarded as precursors to the development of strong low-level rotation in the "Positive" and "Negative" storms because, even when the storms are split into separate subgroups, a common set of quantities still stands out.

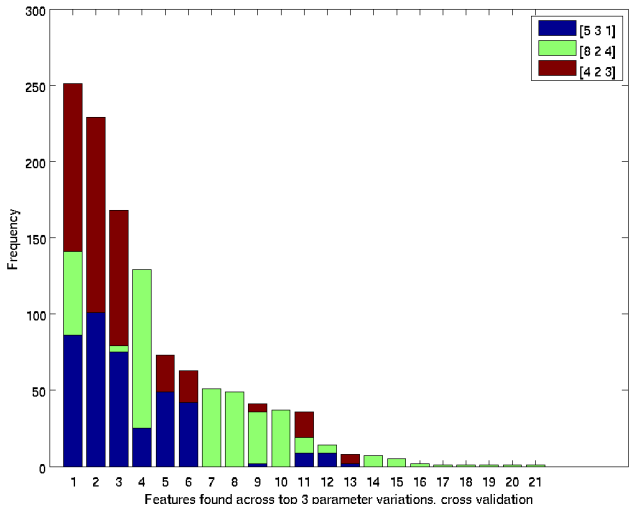


Figure 4.13: Frequency totals for top 10 rules within the top three data mining discretization parameter variations schemes (531, 824 and 423) from Table 4.1 for the 10-fold cross-validation procedure for all "Positive" and "Negative" storms. Specific meteorological quantities associated with numbers 1 – 21 are given in Table 4.6.

Table 4.6: Meteorological quantities and frequency totals corresponding to Figure 4.13. Frequency totals are given for each of the top three discretization schemes (531, 824 and 423) as well as their combined totals. Meteorological quantity rankings for analysis using all "Positive" and "Negative" storms (from Figure 4.11 and Table 4.4) are given in subscript brackets next to the rankings.

Rank	Meteorological Quantity	Frequency			
		Total	531	824	423
1 [1]	vertical stretching term max below 2 km	251	86	55	110
2 [2]	baroclinic generation term (vertical) min below 2 km	229	101	0	128
3 [3]	vertical stretching term min below 2 km	168	75	4	89
4 [4]	pressure perturbation vertical gradient force max below 2 km	129	25	104	0
5 [5]	pressure perturbation vertical gradient force max above 2 km	73	49	0	24
6 [7]	pressure perturbation vertical gradient force min above 2 km	63	42	0	21
7 [6]	vertical velocity max at 4 km	51	0	51	0
8 [8]	horizontal divergence min below 2 km	49	0	49	0
9 [10]	pressure perturbation vertical gradient force min below 2 km	41	2	34	5
10 [9]	vertical velocity horizontal gradient max below 2 km	37	0	37	0
11 [11]	tilting term max below 2 km	36	9	10	17
12	vertical velocity max below 2 km	14	9	5	0
13	baroclinic generation term (vertical) max below 2 km	8	2	0	6
14	tilting term min below 2 km	7	0	7	0
15	tilting term max above 2 km	5	0	5	0
16	baroclinic generation term (horizontal x direction) min below 2 km	2	0	2	0
17	vertical velocity min below 2 km	1	0	1	0
18	vertical velocity min above 2 km	1	0	1	0
19	hail mixing ratio vertical gradient max above 2 km	1	0	1	0
20	vertical stretching term min above 2 km	1	0	1	0
21	Updraft and horizontal Laplacian of radar reflectivity correlation max above 2 km	1	0	1	0

The 10-fold cross-validation analysis also is performed on the top 10 schemes (Figure 4.14 and Table 4.7). Again the top quantities correlate well with the results of the analysis over all "Positive" and "Negative" storms with only a few variations in the ordering of the top 16 quantities (refer to ranking comparisons in Table 4.7). The remaining quantities identified in the 10-fold cross-validation analysis have relatively small frequencies and are not as significant. Therefore, confidence in the identified important quantities also is strengthened for the top 10 rules over the top 10 schemes for all "Positive" and "Negative" storms.

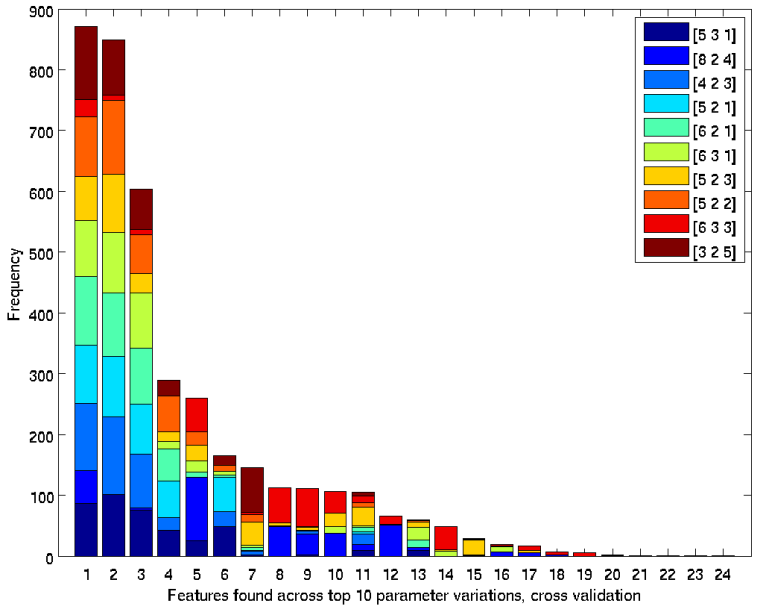


Figure 4.14: Frequency totals for top 10 rules within the top ten data mining discretization parameter variations schemes (531, 824, 423, 521, 621, 631, 523, 522, 633, 325) from Table 4.1 for the 10-fold cross-validation procedure for all "Positive" and "Negative" storms. Specific meteorological quantities associated with numbers 1 – 24 are given in Table 4.7.

Table 4.7: Meteorological quantities and frequency totals corresponding to Figure 4.14. Frequency totals are given for each of the top three discretization schemes (531, 824, 423, 521, 621, 631, 523, 522, 633, 325) as well as their combined totals. Meteorological quantity rankings for analysis using all "Positive" and "Negative" storms (from Figure 4.12 and Table 4.5) are given in subscript brackets next to the rankings.

Rank	Meteorological Quantity	Frequency	
		Total	Top 10 Variations
1 [1]	vertical stretching term max below 2 km	871	86, 55, 110, 96, 113, 92, 72, 98, 29, 120
2 [2]	baroclinic generation term (vertical) min below 2 km	849	101, 0, 128, 100, 104, 99, 96, 122, 8, 91
3 [3]	vertical stretching term min below 2 km	604	75, 4, 89, 82, 92, 91, 32, 64, 8, 67
4 [4]	pressure perturbation vertical gradient force min above 2 km	289	42, 0, 21, 60, 53, 13, 16, 58, 0, 26
5 [5]	pressure perturbation vertical gradient force max below 2 km	260	25, 104, 0, 0, 9, 19, 25, 22, 56, 0
6 [6]	pressure perturbation vertical gradient force max above 2 km	165	49, 0, 24, 57, 3, 6, 0, 10, 0, 16
7 [12]	baroclinic generation term (vertical) max below 2 km	146	2, 0, 6, 1, 5, 4, 38, 12, 4, 74
8 [9]	horizontal divergence min below 2 km	112	0, 49, 0, 0, 0, 1, 5, 0, 57, 0
9 [8]	pressure perturbation vertical gradient force min below 2 km	111	2, 34, 5, 0, 1, 0, 5, 2, 62, 0
10 [7]	vertical velocity horizontal gradient max below 2 km	106	0, 37, 0, 0, 0, 11, 22, 0, 36, 0
11 [11]	tilting term max below 2 km	105	9, 10, 17, 4, 7, 3, 30, 8, 11, 6
12 [13]	vertical velocity max at 4 km	66	0, 51, 0, 0, 0, 0, 1, 0, 14, 0

Table 4.7: Continued

Rank	Meteorological Quantity	Frequency	
		Total	Top 10 Variations
$13_{[14]}$	vertical velocity max below 2 km	60	9, 5, 0, 0, 12, 21, 9, 2, 2, 0
$14_{[10]}$	vertical stretching term max above 2 km	48	0, 0, 0, 0, 0, 8, 3, 0, 37, 0
$15_{[15]}$	vertical stretching term min above 2 km	29	0, 1, 0, 0, 0, 1, 24, 2, 1, 0
$16_{[16]}$	tilting term min below 2 km	19	0, 7, 0, 0, 0, 8, 2, 0, 2, 0
17	tilting term max above 2 km	16	0, 5, 0, 0, 0, 0, 4, 0, 7, 0
18	baroclinic generation term (horizontal x direction) min below 2 km	7	0, 2, 0, 0, 0, 0, 0, 0, 5, 0
19	horizontal divergence min above 2 km	6	0, 0, 0, 0, 0, 0, 0, 0, 6, 0
20	vertical velocity min above 2 km	2	0, 1, 0, 0, 0, 0, 0, 0, 1, 0
21	vertical velocity min below 2 km	1	0, 1, 0, 0, 0, 0, 0, 0, 0, 0
22	hail mixing ratio vertical gradient max above 2 km	1	0, 1, 0, 0, 0, 0, 0, 0, 0, 0
23	Updraft and horizontal Laplacian of radar reflectivity correlation max above 2 km	1	0, 1, 0, 0, 0, 0, 0, 0, 0, 0
24	Updraft and horizontal Laplacian of radar reflectivity correlation min from 0-12km	1	0, 0, 0, 0, 0, 0, 0, 0, 1, 0

The quantities identified in this discussion are all simply extracted maximum or minimum meteorological quantities reaching extreme levels in a certain order within the 30 minute time period prior to the development of strong low-level rotation. The quantities do not provide the spatial evolution of storm features, but using our understanding of supercell storms, the quantities can still be used to deduce changing storm characteristics.

Table 4.8 provides brief descriptions of each of the top quantities in terms of how they may be interpreted for an evolving supercell storm. While most of the quantities are commonly understood, involving increases in updraft, downdraft, convergence or the tilting and stretching of vorticity, a few need further explanation. The third rated quantity is a minimum vertical stretching below 2 km. Most commonly, vertical stretching is associated positive vertical vorticity in a storm's updraft, which happens to be the top rated term shown in Table 4.8. This term,

however, is associated with the vertical stretching of negative vertical vorticity on the right-hand-side of the RFD outflow in the upward flowing air along the RFD gust front. A region of negative vertical vorticity of this nature has been identified in previous studies (see review in Markowski (2002)) and may be linked to either a strengthening of the RFD or a quasi-steady flow pattern best able to maintain counter-rotating vortices on either side of the RFD outflow (e.g., Straka et al. 2007; Markowski et al. 2008).

Table 4.8: Brief description of how the top meteorological quantities identified in the data mining analysis for the top 10 schemes from all "Positive" and "Negative" storms may be interpreted in terms of an evolving supercell storm. Quantities are listed according to their ranking in Table 4.5.

Rank	Meteorological Extreme Quantity and Relation to Supercell Storm
1	vertical stretching term max below 2 km - convergence of positive vertical vorticity into the base of the updraft
2	baroclinic generation term (vertical) min below 2 km - forms on the west side of a large negative pressure perturbation (associated with increasing low level rotation) aligned with a large density gradient (on left-hand-side of RFD in the interior of the hook echo appendage) at low altitudes (see discussion)
3	vertical stretching term min below 2 km - convergence of negative vertical vorticity along the RFD gust front on the right-hand-side of the RFD outflow (e.g., region of enhanced negative vertical vorticity)
4	pressure perturbation vertical gradient force min above 2 km - strengthening downdraft above 2 km
5	pressure perturbation vertical gradient force max below 2 km - strengthening updraft below 2 km
6	pressure perturbation vertical gradient force max above 2 km - strengthening updraft above 2 km
7	baroclinic generation term (vertical) max below 2 km - forms on the east side of a large negative pressure perturbation (associated with increasing low level rotation) aligned with a large density gradient (on left-hand-side of RFD in the interior of the hook echo appendage) at low altitudes (see discussion)
8	horizontal divergence min below 2 km - strong horizontal convergence into the base of the updraft
9	pressure perturbation vertical gradient force min below 2 km - strengthening downdraft below 2 km
10	vertical velocity horizontal gradient max below 2 km - strong updraft adjacent to strong RFD at low altitudes
11	tilting term max below 2 km - tilting of horizontal vorticity to a positive vertical orientation at low altitudes
12	vertical velocity max at 4 km - strong updraft at 4 km
13	vertical velocity max below 2 km - strong updraft at low altitudes
14	vertical stretching term max above 2 km - convergence of positive vertical vorticity within the updraft at or near 2 km
15	vertical stretching term min above 2 km - convergence of negative vertical vorticity on right-hand-side of RFD at or near 2 km - divergence of positive vertical vorticity aloft
16	tilting term min below 2 km - downward tilting of horizontal vortex lines within the RFD at low altitudes.

The other term that has not been thoroughly investigated in previous studies is the vertical component of the baroclinic generation of vorticity, which shows up as a minimum below 2 km (quantity number 2 in Table 4.8) and as a maximum below 2 km (quantity number 7 in Table 4.8). This term is associated with the close proximity and orientation of both a large density gradient located on the interior of the hook echo appendage and a large negative pressure perturbation region associated with increasing low-level rotation.

Figure 4.15 shows an example time series of the maximum (dark blue) and minimum (red) vertical term of the baroclinic generation of vorticity from Storm 1 of experiment WK14_half_r15_4km (Figures 4.1, 4.2 and Appendix D). The large perturbation in the minimum term (red), and moderate perturbation in the maximum term (dark blue) at 7140 s, mirror the perturbations found in the perturbation pressure time series for the same storm (see Figure 4.2(c)). This term is actually a by-product of the strong negative perturbations in pressure associated with increased rotation at low altitudes. Recall that the perturbation pressure (along with vertical vorticity and horizontal wind speed) was removed from the data mining analysis because it was used to define strong low-level rotation and already anticipated to intensify prior to strong rotational development. This, however, does not mean that the vertical component of the baroclinic generation of vorticity term needs to be excluded due to the fact that it includes attributes of the density field along with the pressure perturbation field as specified by:

$$\frac{1}{\rho^2}\left(\frac{\partial\rho}{\partial x}\frac{\partial p}{\partial y}-\frac{\partial\rho}{\partial y}\frac{\partial p}{\partial x}\right), \tag{4.1}$$

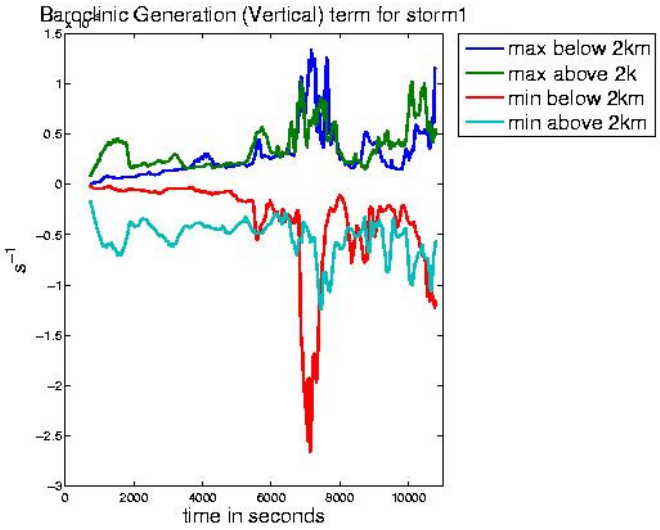


Figure 4.15: Extracted time series metadata from Storm 1 of experiment WK14_half_r15_4km (Figures 4.1, 4.2 and Appendix D) for the vertical component of the baroclinic generation of vorticity equation. Magnitudes along the y-axis are (x 10^{-5}).

It is interesting, however, that the largest perturbations identified in this quantity, which is associated with the generation of vertical vorticity, is one and sometimes two orders of magnitude smaller than other vertical vorticity generation terms (i.e., vertical stretching and tilting). This implies that the quantity itself is not generating significant vertical vorticity but instead is providing information about other features within a storm that are important to the development of strong low-level rotation.

Figures 4.16 and 4.17 illustrate the vertical component of baroclinic generation in a simulated supercell storm. To simplify the discussion, this term will be labeled the baroclinic (vertical) term. The figures show Storm 1 from experiment WK14_half_r15_4km at a time when the maximum and minimum of the baroclinic (vertical) term below 2 km reach their most extreme (7140 s). Figure 4.16(a) provides the vertical velocity (color) and rainwater mixing ratio (white lines starting at 2 g kg^{-1} with 2 g kg^{-1} intervals) at the surface to show where the main supercell features are located. The white outline of the rainwater mixing ratio field shows the main region of the storm, along with the hook echo appendage located near the strongest updraft region. The RFD outflow is visible on the western and southern portions of the storm (darker blue), with the RFD gust front evident by the transition from negative to positive vertical velocities. The most intense updraft and downdraft are located adjacent to one another near the hook echo appendage. Figure 4.16(b) provides the same rainwater mixing ratio field outlined in white from Figure 4.16(a) but now with the color field of the baroclinic (vertical) term. There exists an obvious negative-positive couplet of more intense baroclinically generated vorticity along the updraft-downdraft interface near the hook echo appendage.

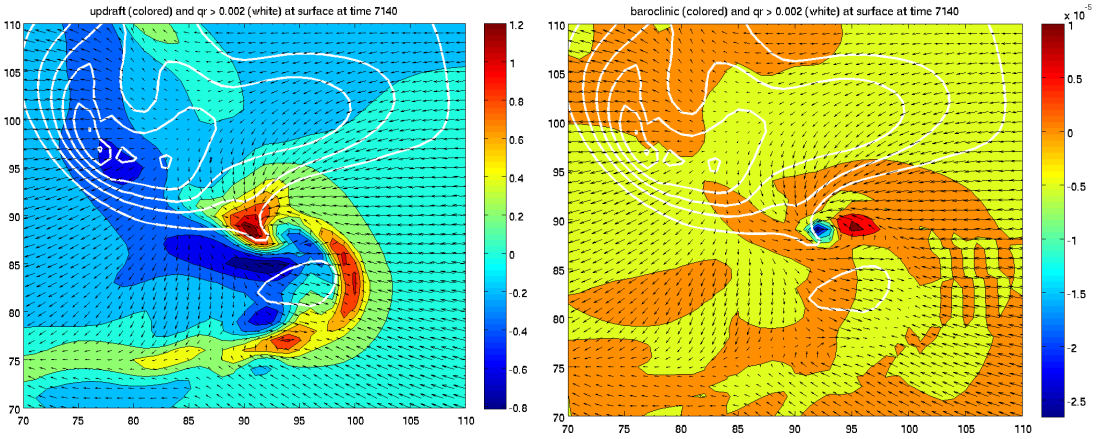


Figure 4.16: Simulated fields at time 7140s from Storm 1 of experiment WK14_half_r15_4km. (a) Vertical velocity in m s^{-1} (color) and rainwater mixing ratio in g kg^{-1} (white lines with 2 g kg^{-1} intervals starting at 2 g kg^{-1}) and (b) vertical component of baroclinic generation of vorticity equation in s^{-1} (color) and the same rainwater mixing ratio field from (a).

Figure 4.17 provides the perturbation pressure and density fields for the same time as Figure 4.16. Figure 4.17(a),(c) display the perturbation pressure and density fields alone while Figure 4.17(b),(d) include the baroclinic (vertical) field outlined in white to locate the extreme maximum and minimum couplet shown in Figure 4.16(b). The maximum and minimum couplet straddles the negative pressure perturbation circular region along the density gradient that separates the cool RFD outflow and warm inflow. Therefore, although this term is influenced by the large decrease in pressure associated with increasing low-level rotation, it also is influenced by the orientation of the density gradient that characterizes a supercell's lower altitudes near the hook appendage. For further discussion on the significance of the identified quantities in Table 4.15 refer to the discussion at the end of the next section as well as the concluding chapter (Chapter 5).

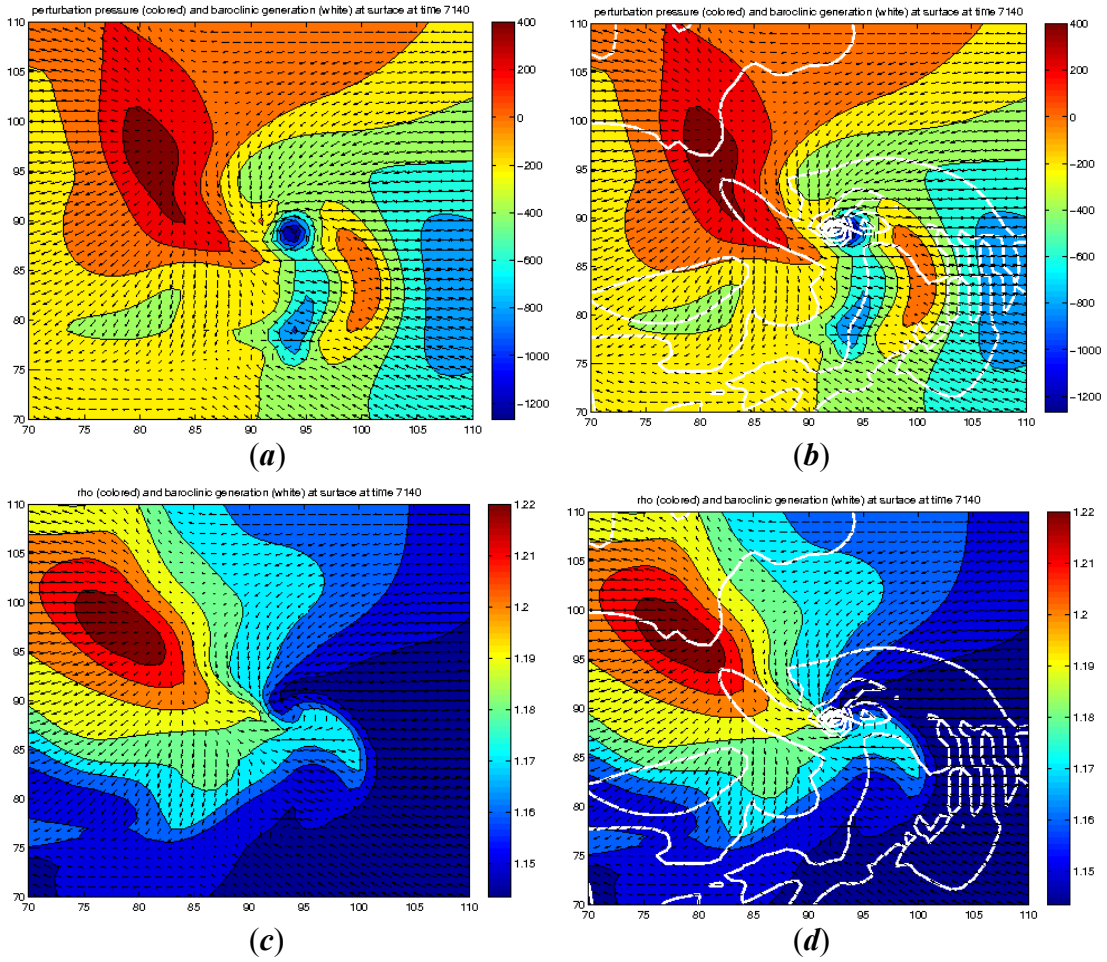


Figure 4.17: Same as Figure 4.16 but for (a) pressure perturbation field in Pa (color), (b) pressure perturbation field in Pa (color) and white outline of the vertical component of baroclinic generation of vorticity equation in s^{-1}, (c) density field in kg m^{-3} (color) and (d) density field in kg m^{-3} (color) and white outline of the vertical component of baroclinic generation of vorticity equation in s^{-1}.

4.2.2 Rules Generated Comparing "Positive" and "Intermediate" Storms

The rules discussed in the previous section were generated from data mining analyses performed on the "Positive" and "Negative" simulated storms. Those analyses compared storms with characteristics presumably farthest away from one

another in our simulated storm spectrum. Now the "Positive" and "Intermediate" storms will be analyzed. Recall that "Intermediate" storms contain some attributes of low-level rotation but are not sufficiently strong to be considered "Positive" cases (see Section 3.3.3). Therefore, by comparing these two fairly similar storm types, additional information may be gained regarding the specific characteristics that differentiate storms containing weak low-level rotation and those containing strong low-level rotation.

Table 4.9 summarizes the results of the data mining discretization scheme parameter variation analysis over the 58 "Positive" and 373 "Intermediate" storms. As with the results of the "Positive" and "Negative" analysis in Table 4.1, the schemes are listed according to alphabet size, word size and averaging interval. Performance measures of POD, FAR, and CSI are provided for the top rule identified when analyzing all "Positive" and "Intermediate" storms as well as for the rule with the top average value over 10 folds when using 10-fold cross-validation. The discretization schemes are ranked according to their 10-fold cross-validation CSI scores with the top 10 schemes highlighted in light blue. A comparison of the top 10 discretization schemes from Table 4.1 and Table 4.9 indicates that the data mining analyses of the "Positive" and "Intermediate" storms produce rules with appreciably lower 10-fold cross-validation CSI values (0.48 – 0.40) than the analyses on the "Positive" and "Negative" storms (0.70 – 0.65). This is due to the greater similarities between storms generating strong low-level rotation and weak low-level rotation.

These two storm types are anticipated to contain features following similar evolutionary paths up to the time period when strong low-level rotation develops.

Table 4.9: Discretization scheme parameter variations applied to the data mining analysis for "Positive" and "Intermediate" storm cases. Schemes are listed according to variations in alphabet size, word size and averaging interval and include the performance measures of POD, FAR, and CSI for the top rule identified when analyzing all "Positive" and "Intermediate" storms and for the average value (across all 10 folds) when using 10-fold cross-validation. Discretization schemes are rated according to 10-fold cross-validation CSI and the top 10 schemes highlighted in light blue.

Alpha size	Word Size	Avg Interval	All "Pos" and "Neg" Storms			10-Fold Cross-Validation			Rank
			POD	FAR	CSI	POD	FAR	CSI	
3	2	1	0.808	0.448	0.487	0.730	0.489	0.434	3
3	2	2	0.749	0.424	0.482	0.640	0.549	0.353	30
3	2	3	0.768	0.415	0.496	0.686	0.470	0.426	5
3	2	4	0.725	0.389	0.495	0.596	0.474	0.386	12
3	2	5	0.731	0.360	0.517	0.556	0.467	0.372	19
3	3	1	0.749	0.398	0.500	0.740	0.424	0.482	1
3	3	2	0.760	0.410	0.496	0.566	0.536	0.338	34
3	3	3	0.722	0.389	0.493	0.463	0.625	0.271	52
3	3	4	0.727	0.431	0.468	0.443	0.571	0.291	48
3	3	5	0.718	0.433	0.462	0.550	0.480	0.355	28
4	2	1	0.783	0.413	0.503	0.683	0.499	0.406	8
4	2	2	0.733	0.365	0.515	0.523	0.554	0.315	41
4	2	3	0.729	0.361	0.516	0.523	0.510	0.344	32
4	2	4	0.724	0.404	0.485	0.580	0.519	0.363	24
4	2	5	0.726	0.445	0.457	0.483	0.554	0.289	49
4	3	1	0.785	0.401	0.513	0.623	0.516	0.389	11
4	3	2	0.726	0.342	0.526	0.533	0.449	0.376	14
4	3	3	0.716	0.410	0.477	0.510	0.532	0.311	42
4	3	4	0.726	0.490	0.427	0.640	0.518	0.373	17
4	3	5	0.752	0.575	0.371	0.720	0.596	0.346	31
5	2	1	0.749	0.344	0.536	0.550	0.472	0.375	15
5	2	2	0.726	0.360	0.514	0.483	0.551	0.310	43
5	2	3	0.731	0.365	0.513	0.560	0.475	0.362	25
5	2	4	0.729	0.418	0.477	0.680	0.464	0.424	6
5	2	5	0.727	0.473	0.440	0.496	0.522	0.316	40
5	3	1	0.727	0.360	0.516	0.560	0.524	0.360	26
5	3	2	0.726	0.324	0.538	0.603	0.401	0.450	2
5	3	3	0.726	0.493	0.424	0.626	0.568	0.340	33
5	3	4	0.772	0.578	0.372	0.700	0.619	0.325	38
5	3	5	0.733	0.554	0.381	0.676	0.584	0.355	29

Table 4.9: Continued

Alpha size	Word Size	Avg Interval	All "Pos" and "Neg" Storms POD	FAR	CSI	10-Fold Cross-Validation POD	FAR	CSI	Rank
6	2	1	0.714	0.354	0.513	0.540	0.464	0.364	23
6	2	2	0.724	0.382	0.499	0.543	0.480	0.373	16
6	2	3	0.729	0.474	0.438	0.686	0.542	0.370	22
6	2	4	0.735	0.489	0.428	0.656	0.479	0.372	18
6	2	5	0.745	0.564	0.377	0.593	0.648	0.278	51
6	3	1	0.718	0.377	0.500	0.536	0.465	0.371	20
6	3	2	0.752	0.454	0.461	0.660	0.512	0.397	10
6	3	3	0.769	0.588	0.365	0.683	0.601	0.338	35
6	3	4	0.752	0.555	0.387	0.673	0.620	0.327	37
6	3	5	0.717	0.559	0.375	0.533	0.658	0.265	53
7	2	1	0.720	0.348	0.519	0.536	0.399	0.377	13
7	2	2	0.724	0.440	0.461	0.500	0.547	0.306	44
7	2	3	0.729	0.503	0.417	0.573	0.547	0.334	36
7	2	4	0.772	0.576	0.375	0.660	0.643	0.299	45
7	2	5	0.745	0.561	0.381	0.743	0.575	0.371	21
7	3	1	0.718	0.396	0.487	0.480	0.491	0.298	46
7	3	2	0.779	0.581	0.369	0.600	0.661	0.261	54
7	3	3	0.758	0.583	0.367	0.620	0.644	0.296	47
7	3	4	0.724	0.533	0.396	0.723	0.511	0.404	9
7	3	5	0.711	0.792	0.191	0.333	0.818	0.133	55
8	2	1	0.718	0.343	0.521	0.506	0.460	0.359	27
8	2	2	0.727	0.461	0.447	0.500	0.538	0.284	50
8	2	3	0.722	0.550	0.382	0.590	0.567	0.323	39
8	2	4	0.793	0.557	0.396	0.800	0.524	0.429	4
8	2	5	0.741	0.537	0.398	0.750	0.515	0.411	7

Similar to the top schemes identified for the "Positive" and "Negative" storms (Table 4.1), the top 10 schemes for the "Positive" and "Intermediate" storms in Table 4.9 (shaded light blue) contain a wide variation of discretization setups. All six alphabet sizes are represented, as well as both word sizes and all 5 averaging intervals.

The top rated discretization scheme shown in Table 4.9 is scheme 331. Table 4.10 provides a list of the top 10 rules from scheme 331 (according to CSI) for all

available “Positive” and “Intermediate” storms, and Figure 4.18 provides the expanded view of rule number one from Table 4.10. The quantities composing these top rules resemble those identified in the previous section (e.g., see Table 4.8). Also, the rules themselves (Figure 4.18) contain the same characteristics as those found in the previous section, the most important of which is that each rule consists of a sequence of meteorological quantities reaching extreme values. It is also apparent from the CSI values for the top 10 rules (Table 4.10) that rules over all “Positive” and “Intermediate” storms ($0.47 < CSI < 0.50$) are significantly lower than those over all “Positive” and “Negative” storms ($0.70 < CSI < 0.74$) as was discussed previously (compare with Table 4.2).

Table 4.10: Top 10 data mining rules (according to CSI) identified for top rated discretization scheme (331) when analyzing all available "Positive" and "Intermediate" storms (Table 4.9). Meteorological quantities within a rule are listed in the order in which they occur.

Rule #	*CSI / POD / FAR*	*Rule*
1	CSI = 0.50 POD = 0.74 FAR = 0.39	- pressure perturbation vertical gradient force max above 2 km - vertical stretching term min below 2 km - vertical stretching term max below 2 km - baroclinic generation term (vertical) min below 2 km
2	CSI = 0.49 POD = 0.76 FAR = 0.41	- pressure perturbation vertical gradient force max above 2 km - pressure perturbation vertical gradient force min below 2 km - vertical stretching term min below 2 km - vertical stretching term max below 2 km
3	CSI = 0.48 POD = 0.79 FAR = 0.45	- pressure perturbation vertical gradient force max above 2 km - vertical stretching term min below 2 km - vertical velocity max below 2 km - vertical stretching term max below 2 km
4	CSI = 0.48 POD = 0.78 FAR = 0.44	- pressure perturbation vertical gradient force max above 2 km - vertical stretching term max below 2 km - vertical stretching term min below 2 km - vertical stretching term max below 2 km
5	CSI = 0.48 POD = 0.78 FAR = 0.44	- pressure perturbation vertical gradient force max above 2 km - vertical stretching term min below 2 km - vertical stretching term max below 2 km - vertical velocity max below 2 km
6	CSI = 0.48 POD = 0.76 FAR = 0.44	- pressure perturbation vertical gradient force max above 2 km - pressure perturbation vertical gradient force min below 2 km - tilting term max below 2 km - vertical stretching term max below 2 km
7	CSI = 0.47 POD = 0.79 FAR = 0.46	- pressure perturbation vertical gradient force max above 2 km - pressure perturbation vertical gradient force min below 2 km - vertical stretching term min below 2 km - vertical velocity max below 2 km
8	CSI = 0.47 POD = 0.76 FAR = 0.44	- pressure perturbation vertical gradient force max above 2 km - vertical stretching term max below 2 km - vertical stretching term max below 2 km - vertical stretching term min below 2 km
9	CSI = 0.47 POD = 0.76 FAR = 0.44	- pressure perturbation vertical gradient force max above 2 km - vertical stretching term min below 2 km - vertical stretching term max below 2 km - vertical stretching term min below 2 km
10	CSI = 0.47 POD = 0.72 FAR = 0.42	- pressure perturbation vertical gradient force max above 2 km - vertical stretching term min below 2 km - baroclinic generation term (vertical) min below 2 km - vertical stretching term max below 2 km

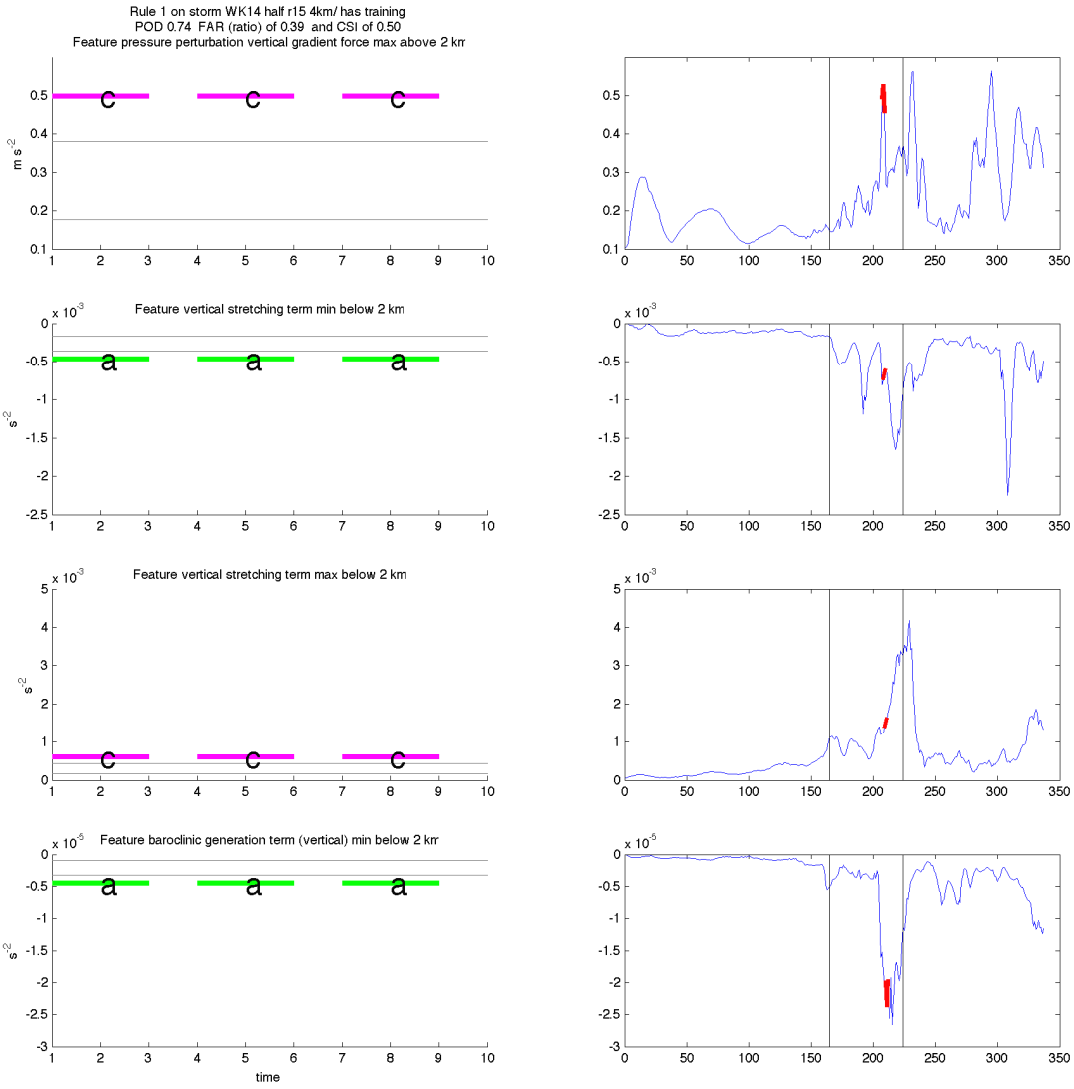


Figure 4.18: Rule #1 from Table 4.10. Left column identifies 3-letter words comprising the rule in sequential order (pressure perturbation vertical gradient force max above 2 km, vertical stretching term min below 2 km, vertical stretching term max below 2 km and baroclinic generation term (vertical) min below 2 km). Time step within each word plot in left column is arbitrary but colored line letter segments correspond to a 30 s time period (one output interval). The three equiprobable Gaussian regions associated with each word are demarcated by light gray horizontal lines. Right column provides meteorological quantity metadata (30 s output intervals) from an example storm (Storm 1 from WK14_half_r15_4km) that contains the rule. The metadata are given in minutes, beginning when the storm was initially identified. Each word from the left column is identified by a red line segment in the metadata. The 30 minute window prior to the development of strong low-level rotation is contained within the two vertical gray lines. Corresponding performance measures are listed at the top of the figure.

Because the top rules, again, are composed of a set of quantities reaching extreme values in varying sequential orders, identifying the number of times a specific quantity appears in the top rules across the top schemes will provide a good measure of the importance of that quantity to the development of strong low-level rotation. Figure 4.19 and Table 4.11 provide quantity frequencies for the top three discretization schemes across all "Positive" and "Intermediate" storms and Figure 4.20 and Table 4.12 show the same information for the top 10 schemes. The quantities identified as the most significant precursors to strong low-level rotation in the top three schemes are the same as those in the top 10 schemes, with variations in the ordering (see top 10 scheme ratings and subscript bracket ratings for top three schemes in Table 4.12). Some of the quantities from the top three schemes are present at the bottom of the list of the top 10 schemes (e.g., vertical velocity max at 4 km and vertical velocity min above 2 km) due to their near-zero frequencies. The additional quantities identified in the top 10 schemes contain relatively smaller frequencies and therefore are not as important.

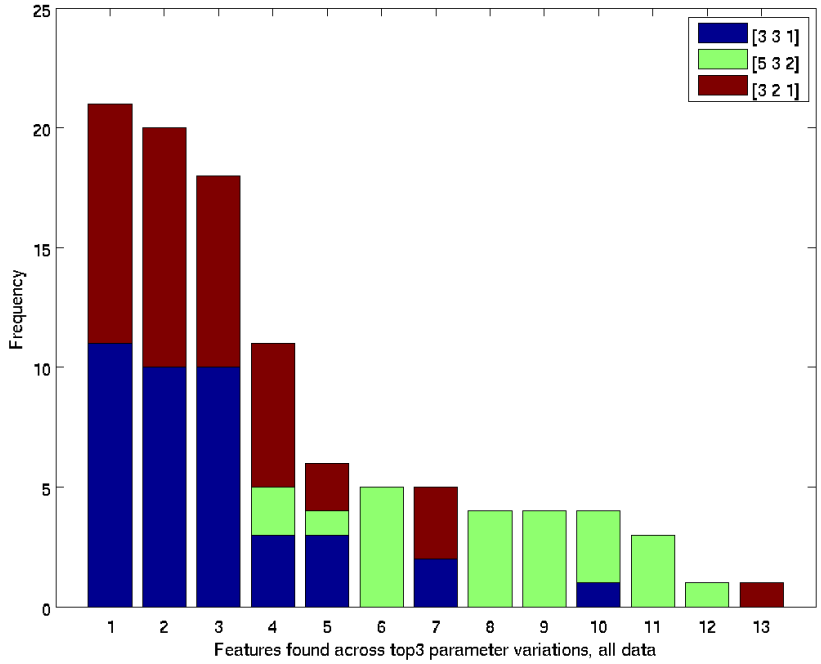


Figure 4.19: Frequency totals for top 10 rules within the top three data mining discretization parameter variation schemes (331, 532 and 321) for all "Positive" and "Intermediate" storms from Table 4.9. Specific meteorological quantities associated with numbers 1 – 13 are given in Table 4.11.

Table 4.11: Meteorological quantities and frequency totals corresponding to Figure 4.19. Frequency totals are given for each of the top three discretization schemes (331, 532 and 321) as well as the combined totals.

Rank	Meteorological Quantity	Frequency			
		Total	331	532	321
1	vertical stretching term max below 2 km	21	11	0	10
2	pressure perturbation vertical gradient force max above 2 km	20	10	0	10
3	vertical stretching term min below 2 km	18	10	0	8
4	vertical velocity max below 2 km	11	3	2	6
5	pressure perturbation vertical gradient force min below 2 km	6	3	1	2
6	tilting term max above 2 km	5	0	5	0
7	baroclinic generation term (vertical) min below 2 km	5	2	0	3
8	vertical velocity horizontal gradient max below 2 km	4	0	4	0
9	pressure perturbation vertical gradient force max below 2 km	4	0	4	0
10	tilting term max below 2 km	4	1	3	0
11	Updraft and horizontal Laplacian of radar reflectivity correlation max above 2 km	3	0	3	0
12	vertical velocity min above 2 km	1	0	1	0
13	vertical velocity max at 4 km	1	0	0	1

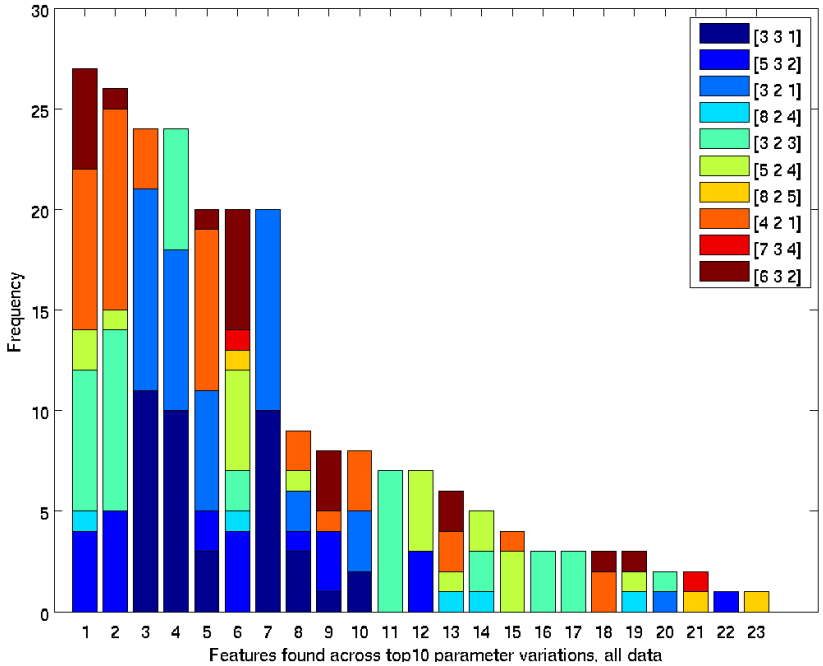


Figure 4.20: Frequency totals for top 10 rules within the top ten data mining discretization parameter variations schemes (331, 532, 321, 824, 323, 524, 825, 421, 734, 632) from Table 4.9 for all "Positive" and "Intermediate" storms. Specific meteorological quantities associated with numbers 1 – 23 are given in Table 4.12.

Table 4.12: Meteorological quantities and frequency totals corresponding to Figure 4.20. Frequency totals are given for each of the top ten discretization schemes (331, 532, 321, 824, 323, 524, 825, 421, 734, 632) as well as the combined totals. Meteorological quantity rankings over the top three schemes (from Table 4.11) are given in subscript brackets next to the rankings over the top 10 schemes.

Rank	Meteorological Quantity	Frequency	
		Total	Top 10 Variations
1 $_{[9]}$	pressure perturbation vertical gradient force max below 2 km	27	0, 4, 0, 1, 7, 2, 0, 8, 0, 5
2 $_{[6]}$	tilting term max above 2 km	26	0, 5, 0, 0, 9, 1, 0, 10, 0, 1
3 $_{[1]}$	vertical stretching term max below 2 km	24	11, 0, 10, 0, 0, 0, 0, 3, 0, 0
4 $_{[3]}$	vertical stretching term min below 2 km	24	10, 0, 8, 0, 6, 0, 0, 0, 0, 0
5 $_{[4]}$	vertical velocity max below 2 km	20	3, 2, 6, 0, 0, 0, 0, 8, 0, 1
6 $_{[8]}$	vertical velocity horizontal gradient max below 2 km	20	0, 4, 0, 1, 2, 5, 1, 0, 1, 6
7 $_{[2]}$	pressure perturbation vertical gradient force max above 2 km	20	10, 0, 10, 0, 0, 0, 0, 0, 0, 0
8 $_{[5]}$	pressure perturbation vertical gradient force min below 2 km	9	3, 1, 2, 0, 0, 1, 0, 2, 0, 0
9 $_{[10]}$	tilting term max below 2 km	8	1, 3, 0, 0, 0, 0, 0, 1, 0, 3
10 $_{[7]}$	baroclinic generation term (vertical) min below 2 km	8	2, 0, 3, 0, 0, 0, 0, 3, 0, 0
11	tilting term min above 2 km	7	0, 0, 0, 0, 7, 0, 0, 0, 0, 0
12 $_{[11]}$	Updraft and horizontal Laplacian of radar reflectivity correlation max above 2 km	7	0, 3, 0, 0, 0, 4, 0, 0, 0, 0

Table 4.12: Continued

Rank	Meteorological Quantity	Frequency	
		Total	Top 10 Variations
13	vertical stretching term min above 2 km	6	0, 0, 0, 1, 0, 1, 0, 2, 0, 2
14	horizontal divergence min below 2 km	5	0, 0, 0, 1, 2, 2, 0, 0, 0, 0
15	vertical stretching term max above 2 km	4	0, 0, 0, 0, 0, 3, 0, 1, 0, 0
16	pressure perturbation vertical gradient force min above 2 km	3	0, 0, 0, 0, 3, 0, 0, 0, 0, 0
17	tilting term min below 2 km	3	0, 0, 0, 0, 3, 0, 0, 0, 0, 0
18	baroclinic generation term (vertical) max below 2 km	3	0, 0, 0, 0, 0, 0, 0, 2, 0, 1
19	baroclinic generation term (horizontal x direction) min below 2 km	3	0, 0, 0, 1, 0, 1, 0, 0, 0, 1
$20_{[13]}$	vertical velocity max at 4 km	2	0, 0, 1, 0, 1, 0, 0, 0, 0, 0
21	radar reflectivity horizontal gradient max below 2 km	2	0, 0, 0, 0, 0, 0, 1, 0, 1, 0
$22_{[12]}$	vertical velocity min above 2 km	1	0, 1, 0, 0, 0, 0, 0, 0, 0, 0
23	radar reflectivity vertical gradient min below 2 km	1	0, 0, 0, 0, 0, 0, 1, 0, 0, 0

To test the robustness of the results, 10-fold cross-validation again is performed. Figure 4.21 and Table 4.13 show the 10-fold cross-validation analysis for the top three discretization schemes, while Figure 4.22 and Table 4.14 provide the results of the analysis on the top 10 schemes. Both analyses indicate general agreement between the 10-fold cross-validation analyses and the analyses performed on all "Positive" and "Intermediate" storms (refer to rankings in Table 4.13 and Table 4.14). Any additional quantities identified have very low frequencies. This strengthens our confidence in the results of the full analyses performed on all "Positive" and "Intermediate" storms.

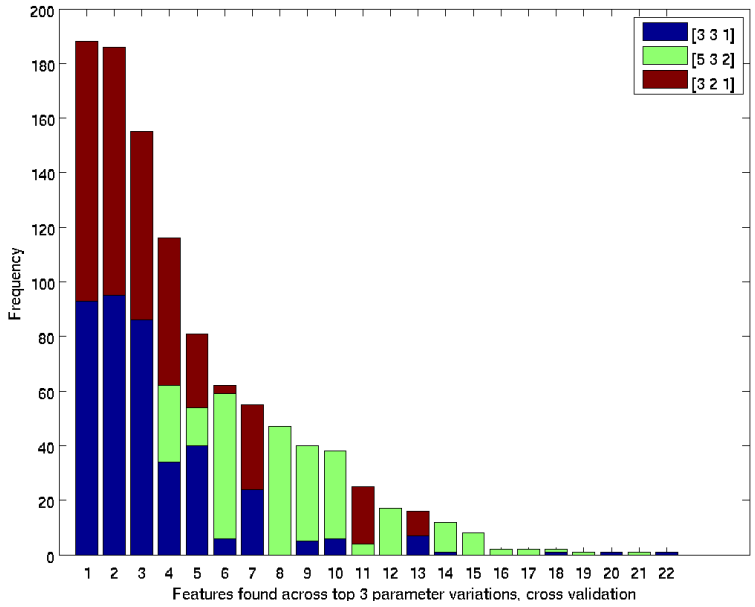


Figure 4.21: Frequency totals for top 10 rules within the top three data mining discretization parameter variations schemes (331, 532 and 321) from Table 4.9 for the 10-fold cross-validation procedure for all "Positive" and "Intermediate" storms. Specific meteorological quantities associated with numbers 1 – 22 are given in Table 4.13.

Table 4.13: Meteorological quantities and frequency totals corresponding to Figure 4.21. Frequency totals are given for each of the top three discretization schemes (331, 532 and 321) as well as the combined totals. Meteorological quantity rankings for analysis using all "Positive" and "Intermediate" storms (from Table 4.11) are given in subscript brackets next to the rankings.

Rank	Meteorological Quantity	Frequency			
		Total	331	532	321
1 $_{[2]}$	pressure perturbation vertical gradient force max above 2 km	188	93	0	95
2 $_{[1]}$	vertical stretching term max below 2 km	186	95	0	91
3 $_{[3]}$	vertical stretching term min below 2 km	155	86	0	69
4 $_{[4]}$	vertical velocity max below 2 km	116	34	28	54
5 $_{[5]}$	pressure perturbation vertical gradient force min below 2 km	81	40	14	27
6 $_{[9]}$	pressure perturbation vertical gradient force max below 2 km	62	6	53	3
7 $_{[7]}$	baroclinic generation term (vertical) min below 2 km	55	24	0	31
8 $_{[6]}$	tilting term max above 2 km	47	0	47	0
9 $_{[8]}$	vertical velocity horizontal gradient max below 2 km	40	5	35	0
10 $_{[10]}$	tilting term max below 2 km	38	6	32	0
11 $_{[13]}$	vertical velocity max at 4 km	25	0	4	21
12 $_{[11]}$	Updraft and horizontal Laplacian of radar reflectivity correlation max above 2 km	17	0	17	0
13	pressure perturbation vertical gradient force min above 2 km	16	7	0	9
14	horizontal divergence min below 2 km	12	1	11	0

Table 4.13: Continued

Rank	Meteorological Quantity	Frequency			
		Total	331	532	321
$15_{[12]}$	vertical velocity min above 2 km	8	0	8	0
16	horizontal divergence min above 2 km	2	0	2	0
17	vertical stretching term max above 2 km	2	0	2	0
18	tilting term min below 2 km	2	1	1	0
19	vertical velocity min below 2 km	1	0	1	0
20	vertical stretching term min above 2 km	1	1	0	0
21	baroclinic generation term (vertical) max below 2 km	1	0	1	0
22	baroclinic generation term (horizontal x direction) min below 2 km	1	1	0	0

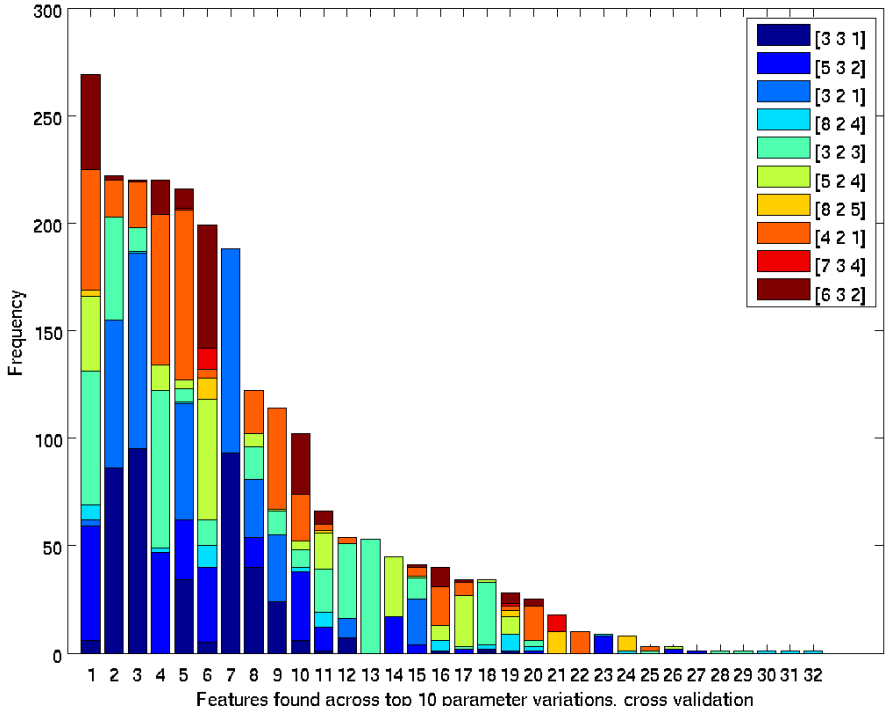


Figure 4.22: Frequency totals for top 10 rules within the top ten data mining discretization parameter variations schemes (331, 532, 321, 824, 323, 524, 825, 421, 734, 632) from Table 4.9 for the 10-fold cross-validation procedure for all "Positive" and "Intermediate" storms. Specific meteorological quantities associated with numbers 1 – 32 are given in Table 4.14.

Table 4.14: Meteorological quantities and frequency totals corresponding to Figure 4.22. Frequency totals are given for each of the top three discretization schemes (331, 532, 321, 824, 323, 524, 825, 421, 734, 632) as well as the combined totals. Meteorological quantity rankings for analysis using all "Positive" and "Intermediate" storms (from Table 4.12) are given in subscript brackets next to the rankings.

Rank	Meteorological Quantity	Frequency	
		Total	Top 10 Variations
1 [1]	pressure perturbation vertical gradient force max below 2 km	269	6, 53, 3, 7, 62, 35, 3, 56, 0,44
2 [4]	vertical stretching term min below 2 km	222	86, 0, 69, 0, 48, 0, 0, 17, 0, 2
3 [3]	vertical stretching term max below 2 km	220	95, 0, 91, 1, 11, 0, 0, 21, 0, 1
4 [2]	tilting term max above 2 km	220	0, 47, 0, 2, 73, 12, 0, 70, 0,16
5 [5]	vertical velocity max below 2 km	216	34, 28, 54, 1, 6, 4, 0, 79, 1, 9
6 [6]	vertical velocity horizontal gradient max below 2 km	199	5, 35, 0,10, 12,56,10,4, 10,57
7 [7]	pressure perturbation vertical gradient force max above 2 km	188	93, 0, 95, 0, 0, 0, 0, 0, 0, 0
8 [8]	pressure perturbation vertical gradient force min below 2 km	122	40, 14, 27, 0, 15, 6, 0, 20, 0,0
9 [10]	baroclinic generation term (vertical) min below 2 km	114	24, 0, 31, 0, 11, 1, 0, 47, 0, 0
10 [9]	tilting term max below 2 km	102	6, 32, 0, 2, 8, 4, 0, 22, 0, 28
11 [14]	horizontal divergence min below 2 km	66	1, 11, 0, 7, 20, 17, 1, 3, 0, 6
12 [16]	pressure perturbation vertical gradient force min above 2 km	54	7, 0, 9, 0, 35, 0, 0, 3, 0, 0
13 [11]	tilting term min above 2 km	53	0, 0, 0, 0, 53, 0, 0, 0, 0, 0
14 [12]	Updraft and horizontal Laplacian of radar reflectivity correlation max above 2 km	45	0, 17, 0, 0, 0, 28, 0, 0, 0, 0
15 [20]	vertical velocity max at 4 km	41	0, 4, 21, 0, 10, 1, 0, 4, 0, 1
16 [13]	vertical stretching term min above 2 km	40	1, 0, 0, 5, 0, 7, 0, 18, 0, 9
17 [15]	vertical stretching term max above 2 km	34	0, 2, 0, 0, 1, 24, 0, 6, 0, 1
18 [17]	tilting term min below 2 km	34	1, 1, 0, 2, 29, 1, 0, 0, 0, 0
19 [19]	baroclinic generation term (horizontal x direction) min below 2 km	28	1, 0, 0, 8, 0, 8, 3, 2, 1, 5
20 [18]	baroclinic generation term (vertical) max below 2 km	25	0, 1, 0, 2, 3, 0, 0, 16, 0, 3
21 [21]	radar reflectivity horizontal gradient max below 2 km	18	0, 0, 0, 0, 0, 0, 10, 0, 8, 0
22	hail mixing ratio vertical gradient max above 2 km	10	0, 0, 0, 0, 0, 0, 0, 10, 0, 0
23 [22]	vertical velocity min above 2 km	9	0, 8, 0, 1, 0, 0, 0, 0, 0, 0
24 [23]	radar reflectivity vertical gradient min below 2 km	8	0, 0, 0, 1, 0, 0, 7, 0, 0, 0
25	vertical velocity max above 2 km	3	0, 0, 0, 0, 1, 0, 0, 2, 0, 0
26	horizontal divergence min above 2 km	3	0, 2, 0, 0, 0, 1, 0, 0, 0, 0
27	vertical velocity min below 2 km	1	0, 1, 0, 0, 0, 0, 0, 0, 0, 0
28	rainwater mixing ratio horizontal gradient max above 2 km	1	0, 0, 0, 0, 1, 0, 0, 0, 0, 0
29	rainwater mixing ratio vertical gradient min above 2 km	1	0, 0, 0, 0, 1, 0, 0, 0, 0, 0
30	radar reflectivity max above 2 km	1	0, 0, 0, 1, 0, 0, 0, 0, 0, 0
31	horizontal Laplacian of radar reflectivity max below 2 km	1	0, 0, 0, 1, 0, 0, 0, 0, 0, 0
32	Updraft and horizontal Laplacian of radar reflectivity correlation max below 2 km	1	0, 0, 0, 1, 0, 0, 0, 0, 0, 0

The top quantities identified by the analysis on "Positive' and "Intermediate storms are listed in Table 4.15, along with a brief description of how they may be interpreted for an evolving supercell storm. The quantities are listed in the order of their ranking in the top 10 schemes (Table 4.12). Also provided is the ranking each quantity acquired in the data mining analysis using only "Positive" and "Negative" storms from Table 4.8 in the previous section. The two different analyses generate a remarkably similar list of quantities identified as generally unique to storms producing strong low-level rotation. The only major difference in the top quantities is the addition of two tilting term quantities (tilting term max above 2 km and tilting term min above 2 km). It may be possible to explain these additions after first observing that the quantities "tilting term max *below* 2 km" and "tilting term min *below* 2 km" are already present in both lists. Therefore, distinguishing between "Positive" and "Negative" storms may potentially be accomplished simply by identifying the extreme values reached by these two tilting terms below 2 km because the "Negative" storms will likely have relatively minimal values of tilting compared to the "Positive" storms. However, distinguishing between "Positive" and "Intermediate" storms may be more difficult because both types of storms will probably contain similar values of both tilting terms below 2 km. Therefore the distinguishing factor between the "Positive" and "Intermediate" storms may be the identification of tilting term values which have grown so intense at regions below 2 km that their presence now extends into regions above 2 km.

There is also the addition of the 12th rated quantity (updraft and horizontal Laplacian of radar reflectivity correlation max above 2 km) which corresponds with a strong updraft and presence of a BWER. The other new quantities contain relatively small frequencies and are not as significant (see Table 4.12).

Table 4.15: Brief description of how the top meteorological quantities identified in the data mining analysis for the top 10 schemes for all "Positive" and "Intermediate" storms may be interpreted in terms of an evolving supercell storm. Quantities are listed according to their ranking in Table 4.12 and the quantity's ranking from the "Positive" and "Negative" data mining analysis is given in subscript brackets (from Table 4.8).

Rank	Meteorological Extreme Quantity and Supercell Storm Significance
1 [5]	pressure perturbation vertical gradient force max below 2 km - strengthening updraft
2	tilting term max above 2 km - upward tilting of horizontal vorticity by the updraft
3 [1]	vertical stretching term max below 2 km - convergence of positive vertical vorticity into the base of the updraft
4 [3]	vertical stretching term min below 2 km - convergence of negative vertical vorticity along the RFD gust front on the right-hand-side of the RFD outflow (e.g., region of enhanced negative vertical vorticity)
5 [13]	vertical velocity max below 2 km - strong updraft at low altitudes
6 [10]	vertical velocity horizontal gradient max below 2 km - strong updraft adjacent to strong RFD at lower altitudes
7 [6]	pressure perturbation vertical gradient force max above 2 km - strengthening updraft
8 [9]	pressure perturbation vertical gradient force min below 2 km - strengthening downdraft
9 [11]	tilting term max below 2 km - tilting of horizontal vorticity to a positive vertical orientation
10[2]	baroclinic generation term (vertical) min below 2 km - on the west side of an alignment of negative pressure perturbation associated with increasing low level rotation and negative density gradient in y direction on left-hand-side of RFD and hook appendage at low altitudes (see discussion)
11	tilting term min above 2 km - downward tilting of horizontal vorticity by the RFD
12	updraft and horizontal Laplacian of radar reflectivity correlation max above 2 km - presence of a bounded weak echo region
13[15]	vertical stretching term min above 2 km - convergence of negative vertical vorticity on right-hand-side of RFD at or near 2 km - divergence of positive vertical vorticity aloft
14[8]	horizontal divergence min below 2 km - strong horizontal convergence into the base of the updraft
15[14]	vertical stretching term max above 2 km - convergence of positive vertical vorticity within the updraft at or near 2 km
16[4]	pressure perturbation vertical gradient force min above 2 km - strengthening downdraft
17[16]	tilting term min below 2 km - downward tilting of horizontal vortex lines within the RFD.

Table 4.15: Continued

Rank	Meteorological Extreme Quantity and Supercell Storm Significance
$18_{[7]}$	baroclinic generation term (vertical) max below 2 km - on the east side of an alignment of negative pressure perturbation associated with increasing low level rotation and negative density gradient in y direction on left-hand-side of RFD and hook appendage at low altitudes (see discussion)
19	baroclinic generation term (horizontal x direction) min below 2 km - baroclinic generation of horizontal vorticity along the FFD gust front where vorticity is oriented in the negative x direction towards the updraft (FFD gust fronts usually have east-west orientation in simulations)
$20_{[12]}$	vertical velocity max at 4 km - strong updraft at 4 km
21	radar reflectivity horizontal gradient max below 2 km - strong horizontal gradient in radar reflectivity near the hook echo appendage
22	vertical velocity min above 2 km - strong downdraft above 2 km
23	radar reflectivity vertical gradient min below 2 km - low-level extension of the hook echo appendage out from directly underneath the storm produces a negative radar reflectivity gradient in its vicinity

The principal result from both the "Positive" – "Negative" and "Positive" – "Intermediate" analyses is that rules were generated with extracted extreme quantities that can be used to explain the physical characteristics of an evolving supercell storm. Therefore, the generated rules can be trusted as physically relevant. The incorporation of a data mining procedure replaced what would have been hundreds of hours of required analysis time (if possible at all) to identify particular sequences of events unique to "Positive" storms only within the extracted metadata of hundreds of simulated storms. The successful implementation of this data mining procedure will foster further expansions of this work as well as the development of more advanced search techniques. These future developments will be addressed in the concluding chapter, in which a summary of the important implications of this study also will be provided.

CHAPTER 5

CONCLUSION

5.1 Summary and Discussion

Despite considerable progress made in recent decades in the observation, modeling, and theoretical understanding of tornadoes, warning and forecasting their occurrence remains a considerable challenge. Quantitative statistics clearly show that warning probability of detection (POD) and lead time have plateaued in recent years, with false alarm ratio (FAR) remaining relatively constant (Figure 1.1), principally because existing surveillance radars and hazardous weather detection methodologies suffer from fundamental limitations that allow key meteorological quantities and associated features to go undetected. New advances will be required if substantial improvements in warning and forecasting accuracy are to take place.

One promising avenue is the use of a data assimilation procedure, in *real-time*, which is capable of increasing the number of meteorological quantities available for use in diagnostic detection algorithms and full dynamic prediction models. Data assimilation incorporates information from all available observing systems to produce a single three-dimensional (3-D) gridded analysis of the atmosphere that includes retrieved quantities, at high spatial resolution, that are dynamically and thermodynamically consistent with observations. The current application of data assimilation, at the scale of convective storms, is limited principally to the initialization of numerical weather prediction models due to time requirements

associated with creating the analysis, but with computational advances, these assimilated analyses soon will be available to forecasters in real-time. Applying detection algorithms to these real-time analyses should provide a considerable advantage over existing techniques, which, in the case of tornadoes, mostly utilize radar radial velocity and reflectivity data in their native spherical-polar coordinates and are strongly linked to the particular type of radar under consideration (i.e., the algorithms must be changed to accommodate new radars). When hazardous weather detection algorithms are applied to gridded analyses produced via data assimilation, new observing systems can be added to the assimilation process, thereby improving the analyses, without requiring a change to detection and anticipation algorithms that utilize them.

The methodology of detection and anticipation using real-time assimilated data sets rather than direct observations necessitate the development of a new suite of algorithms able to operate on regular grids and accommodate retrieved fields. The performance of these algorithms will depend upon their ability to identify storm features and feature interrelationships prior to the development of a severe weather event (e.g., tornado) as observed within a 3-D gridded domain containing all important meteorological fields.

This study provided an initial framework for identifying important features and feature interrelationships, intended for future hazardous weather detection algorithms, which herald the development of strong low-level rotation within deep convective storms. Numerical simulations were used instead of observational data

because simulations provide all meteorological fields within a 3-D gridded structure, analogous to future assimilated analyses, and could generate a large number of storms under controlled conditions that utilize varying initial background environments. Assimilated data sets based upon observations could not be used because too few of these data sets exist, and it would be difficult to verify the retrieved fields of those available. The low-level rotation qualifier was used because tornadic vortices can not be resolved with the model grid spacing employed here, though as computational advances foster higher resolution simulations, the methodology may be extended to include smaller scale vortices.

A total of 1168 numerically simulated storms were generated from within initial environments characteristic of supercell storms and were categorized by whether they produced strong low-level rotation ("Positive" storms), weak low-level rotation ("Intermediate" storms) or no low-level rotation ("Negative" storms). A computational data mining procedure was developed to search the gridded fields for meteorological precursors, occurring in repeatable patterns, that lead to the development of strong low-level rotation ("Positive" storm). This procedure was performed on two separate storm subgroups ("Positive" vs. "Negative" and "Positive" vs. "Intermediate") to isolate the differences between storms that differ most significantly in their distinguishing characteristics ("Positive" vs. "Negative") as well as those more closely related ("Positive" vs. "Intermediate").

Our results identified sets of precursors, in the form of meteorological quantities reaching extreme values in a particular temporal sequence, unique to

storms producing strong low-level rotation ("Positive" storms) in both subgroups. These sequences (termed rules) were assigned probability of detection (POD), false alarm ration (FAR) and critical success index (CSI) values according to their performance in predicting strong low-level rotation ("Positive" storms). The highest rated rules generated similar performance measures and consisted of the same meteorological quantities with small variations in temporal ordering. This implied that the *order* in which quantities reached extreme values was less significant than the requirement that the quantities simply *reach* extreme values.

With this in mind, frequency distributions of quantity occurrence in the top rated rules were generated, identifying the most important quantities reaching extreme values prior to the development of strong low-level rotation. The top five quantities identified when analyzing all "Positive" and "Negative" storms were: maximum in vertical vorticity stretching below 2 km, minimum in baroclinic vertical vorticity generation below 2 km, minimum in vertical vorticity stretching below 2 km, minimum vertical perturbation pressure gradient force above 2 km and maximum vertical perturbation pressure gradient force below 2 km. These, and the other quantities identified, each could be explained physically in terms of an evolving supercell storm (Table 4.8). For example, the top five listed quantities correlate to the following storm characteristics respectively: convergence of positive vertical vorticity into the base of the updraft on the left-hand-side of the RFD, alignment of strong horizontal gradients of perturbation pressure and density on the interior side of the hook appendage, convergence of negative vertical vorticity on the right-hand-side

of the RFD near the RFD gust front, strengthening downdraft above 2 km and strengthening updraft below 2 km.

The second analysis group ("Positive" vs. "Intermediate" storms) generated rules containing similar quantities as the "Positive" and "Negative" analysis (Table 4.15) but with less statistical significance. This is invariably due to the greater similarities in storm characteristics between "Positive" and "Intermediate" storms

The large number of rules identified by this study should prove useful in the development of algorithms for anticipating strong low-level rotation in real-time 3-D gridded assimilated analyses. The goal would be for the algorithms to target and continuously track meteorological quantities within significant rules and alert forecasters when a specific rule (sequence of events) has occurred in a storm. As additional rules are identified, their performance measures from the data mining analysis can be displayed to the forecaster, indicating a storm's potential for developing strong low-level rotation. With computational advances, finer simulation grids can be used to extend the concept to the tornado scale.

5.2 Limitations and Future Work

The present study is obviously limited by its use of numerical simulations. The output from these simulations are dependent on the model set up and initialization parameters (e.g., horizontal and vertical grid spacing, integration time step, cloud microphysics, etc.) which may hinder the accurate reproduction of conditions within a storm (i.e., model cannot resolve tornadic vortices). Also, any

variation in the model's initial parameters is certain to change the extracted storm metadata which, in turn, may alter the data mining results (e.g., changing the horizontal grid spacing will alter those quantities containing horizontal partial derivatives in their denominator, such as divergence and vorticity). Future studies should focus on how sensitive the data mining results are when model parameters are systematically altered.

Another important limitation of this study is that the extracted maximum and minimum quantities used to track a storm's evolutionary characteristics provide limited to no information regarding the spatial variations of storm features. The only spatial attributes included are the designation of which vertical region a quantity is being extracted, which does not guarantee that successive output intervals within a quantity's time series are located within the same general horizontal location. For example, if a region of significant positive vertical vorticity in one part of a SER_{5km} begins to weaken while at the same time another region of positive vertical vorticity intensifies at a different location then it may be possible for the extracted metadata for positive vertical vorticity to show almost no change during this time period because only the most extreme quantity within the search space is extracted. The spatial evolution of storm features is an important aspect not incorporated into the current study that should be addressed in future research.

The current study also does not explicitly include a lead time component. As discussed previously, the time separation of words contained within a rule is not included in the data mining analysis. For an initial study, this was deemed acceptable

because it minimized the complexity in the data mining analysis and allowed a larger number of significant rules to be identified. The time separation between words should be included in future studies and the results compared to those found in this work.

Another limiting factor is the parameter space chosen for generating the simulated storms. The initial environments used in the study were chosen because they fell within a particular range of the storm spectrum capable of producing supercell storms. This parameter space will need to be expanded to include the wider range of environments observed in the real atmosphere if results from this method are to be applied to real-time assimilated analyses in the future. Temperature and moisture profiles differing from the Weisman and Klemp (1983) profiles also should be applied to generate alternative simulated storms for the data mining analysis.

The data mining analysis results discussed in this study focused on those generated by the top discretization schemes (Table 4.1 and Table 4.9). A wider and more detailed investigation should be conducted across many of the schemes in search of variations that may occur from one type of scheme to another.

Besides the improvements needed to address the specific limitations of this study, there also will need to be improvements made in the spatial and temporal resolution of observing systems that will be incorporated into the future real-time assimilation analyses. Even if suitable search algorithms can be constructed, they still will depend on the accuracy of the analyses in which they are applied. Based on the results of this study, it is evident that the meteorological quantities identified as

containing the most important precursors to the development of strong low-level rotation tended to be located in the lowest 2 km of the atmosphere. Therefore, it will be essential that additional observing systems be developed which are capable of better sampling this region.

Additionally, if the results of the methodology employed in this study are to be incorporated into detection and anticipation algorithms in future assimilated analyses then a method will need to be developed to verify that identified precursors are indeed real life precursors within actual storms. In the present study, numerical simulations are used that provide data sets in which the results are known and can be controlled. In the real life atmosphere, the results (e.g., actual observed meteorological quantities) are most often retrieved and therefore unknown which necessitates the development of a new verification procedure able to establish the accuracy of the algorithms. Finer scale observational data will be required if such a verification procedure is to be effective.

This thesis represents the first study in what is expected to be an extensive line of research utilizing data mining techniques to computationally search through large meteorological data sets (numerical and/or observational) to uncover precursors to significant meteorological events that would otherwise be too difficult to search through manually. The current study has already stimulated tangential projects utilizing knowledge gained from the generated data mining technique (e.g., McGovern et al. 2008) and future work is expected to further increase our ability to

identify universal spatial and temporal patterns that will ultimately be incorporated into severe weather recognition algorithms for real-time weather forecasters.

REFERENCES

Adlerman, E.J., K.K. Droegemeier, and R. Davies-Jones, 1999: A numerical simulation of cyclic mesocyclogenesis. *J. Atmos. Sci.*, **56**, 2045-2069.

Adlerman, E.J., and K.K. Droegemeier, 2002: The sensitivity of numerically simulated cyclic mesocyclogenesis to variations in model physical and computational parameters. *Mon. Wea. Rev.*, **130**, 2671-2691.

Adlerman, E.J., and K.K. Droegemeier, 2005: The dependence of numerically simulated cyclic mesocyclogenesis upon environmental vertical wind shear. *Mon. Wea. Rev.*, **133**, 3595-3623.

Barnes, S.L., 1978a: Oklahoma thunderstorms on 29-30 April 1970. Part I: Morphology of a tornadic storm. *Mon. Wea. Rev.*, **106**, 673-684.

Barnes, S.L., 1978b: Oklahoma thunderstorms on 29-30 April 1970. Part II: Radar-observed merger of twin hook echoes. *Mon. Wea. Rev.*, **106**, 685-696.

Barnes, L.R., E.C. Gruntfest, M.H. Hayden, D.M. Schultz, and C. Benight, 2007: False alarms and close calls: A conceptual model of warning accuracy. *Wea. Forecasting*, **22**, 1140-1147.

Beck, J.R., J.L. Schroeder, and J.M. Wurman, 2006: High-resolution dual-Doppler analyses of the 29 May 2001 Kress, Texas, cyclic supercell. *Mon. Wea. Rev.*, **134**, 3125-3172.

Bieringer , P., and P.S. Ray, 1996: A comparison of tornado warning lead times with and without NEXRAD Doppler radar. *Wea. Forecasting*, **11**, 47-52.

Brandes, E.A., 1977a: Flow in a severe thunderstorm observed by dual-Doppler radar. *Mon. Wea. Rev.*, **105**, 113-120.

Brandes, E.A., 1977b: Gust front evolution and tornadogenesis as viewed by Doppler radar. *J. Appl. Meteor.*, **16**, 333-338.

Brandes, E.A., 1978: Mesocyclone evolution and tornadogenesis: Some observations. *Mon. Wea. Rev.*, **106**, 995-1011.

Brandes, E.A., 1984a: Relationships between radar-derived thermodynamic variables and tornadogenesis. *Mon. Wea. Rev.*, **112**, 1033-1052.

Brandes, E.A., 1984b: Vertical vorticity generation and mesocyclone sustenance in tornadic thunderstorms: The observational evidence. *Mon. Wea. Rev.*, **112**, 2253-2269.

Brewster, K.A., K.W. Thomas, J. Brotzge, Y. Wang, D. Weber, and M. Xue, 2007: High resolution data assimilation of CASA X-band radar data for thunderstorm forecasting. Preprints, *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 1B.1.

Brooks, H.E., and R.B. Wilhelmson, 1993: Hodograph curvature and updraft intensity in numerically modeled supercells. *J. Atmos. Sci.*, **50**, 1824-1833.

Brown, R.A., D.W. Burgess, and K.C. Crawford, 1973: Twin tornado cyclones within a severe thunderstorm: Single-Doppler radar observations. *Weatherwise*, **26**, 63-71.

Browning, K.A., 1964: Airflow and precipitation trajectories within severe local storms which travel to the right of the winds. *J. Atmos. Sci.*, **21**, 634-639.

Browning, K.A., 1965: Some inferences about the updraft within a severe local storm. *J. Atmos. Sci.*, **22**, 669-678.

Browning, K.A., 1968: The organization of severe local storms. *Weather*, **23**, 429-434.

Browning, K.A., and F.H. Ludlam, 1962: Airflow in convective storms. *Quart. J. Roy. Meteor. Soc.*, **88**, 117-135.

Browning, K.A., and R.J. Donaldson, 1963: Airflow and structure of a tornadic storm. *J. Atmos. Sci.*, **20**, 533-545.

Burgess, D. W., V. T. Wood, and R. A. Brown, 1982: Mesocyclone evolution statistics. Preprints, *12th Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 422–424**.**

Burgess, D.W., R.J. Donaldson, Jr., and P.R. Desrochers, 1993: Tornado detection and warning by radar. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophs. Monogr.*, No. 79, *Amer. Geophs. Union*, 203-221.

Burgess, D.W., L.D. Hennington, R.J. Doviak, and P.S. Ray, 1976: Multi-moment Doppler display for severe storm identification. *J. Appl. Meteor.*, **15**, 1302-1306.

Burgess, D.W., R.A. Brown, L.R. Lemon, and C.R. Safford, 1977: Evolution of a tornadic thunderstorm. Preprints, *10th Conf. on Severe Local Storms*, Omaha, NE, Amer. Meteor. Soc., 84-89.

Byers, H.R., 1942: Nonfrontal thunderstorms. *Misc. Rep. No. 3*. University of Chicago Press, 26 pp.

Byers, H.R., and R.R. Braham Jr., 1949: *The Thunderstorm*. Washington D.C., Government Printing Office, 287 pp.

Crum, T.D., R.E. Saffle, and J.W. Wilson, 1998: An update on the NEXRAD program and future WSR-88D support to operations. *Wea. Forecasting*, **13**, 253-262.

Chisholm, A.J., 1973: Alberta hailstorms. Part I: Radar case studies and airflow models. *Alberta Hailstorms, Meteor. Monogr.*, No. 36, Amer. Meteor. Soc., 1-36.

Chisholm, A.J., and J.H. Renick, 1972: The kinematics of multicell and supercell Alberta hailstorms. Alberta Hail Studies, 1972, Research Council of Alberta Hail Studies Rep. 72-2, 24-31.

Darkow, G.L., and D.W. McCann, 1977: Relative environmental winds for 121 tornado bearing storms. Preprints, *10th Conf. on Severe Local Storms*, Omaha, NE, Amer. Meteor. Soc., 413-417.

Davies, J.M., and R.H. Johns, 1993: Some wind and instability parameters associated with strong and violent tornadoes: 1. Wind shear and helicity. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr.*, No. 79, Amer. Geophys. Union, 573-582.

Davies-Jones, R.P., 1982a: A new look at the vorticity equation with application to tornadogenesis. Preprints, *12th Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 249-252.

Davies-Jones, R.P., 1982b: Observational and theoretical aspects of tornadogenesis. *Intense Atmospheric Vortices*, L. Bengtsson and J. Lighthill, Eds. Springer-Verlag, 175-189.

Davies-Jones, R.P., 1984: Streamwise vorticity: The origin of updraft rotation in supercell storms, *J. Atmos. Sci.*, **41**, 2991-3006.

Davies-Jones, R.P., 1986: Tornado dynamics. *Thunderstorm Morphology and Dynamics*, 2nd ed., E. Kessler, Ed., University of Oklahoma Press, 197-236.

Davies-Jones, R.P., and H.E. Brooks, 1993: Mesocyclogenesis from a theoretical perspective. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr*., No. 79, Amer. Geophys. Union, 105-114.

Davies-Jones, R.P., D.W. Burgess, and M. Foster, 1990: Test of helicity as a forecast parameter. Preprints, *16th Conf. on Severe Local Storms*, Kananaskis Park, Alberta, Amer. Meteor. Soc., 588-592.

Davies-Jones, R.P., R.J. Trapp, and H.B. Bluestein, 2001: Tornadoes and tornadic storms. *Severe Convective Storms, Meteor. Monogr.*, No. 50, Amer. Meteor. Soc., 167-221.

Donaldson, R.J., 1970: Vortex signature recognition by a Doppler radar. *J. Appl. Meteor.*, **9**, 661-670.

Doswell, C.A., III, and D.W. Burgess, 1993: Tornadoes and tornadic storms: A review of conceptual models. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr.*, No. 79, Amer. Geophys. Union, 161-172.

Doswell, C.A., III, A.R. Moller, and H.E. Brooks, 1999: Storm spotting and public awareness since the first tornado forecasts of 1948. *Wea. Forecasting*, **14**, 544-557.

Dowell, D.C., and H.B. Bluestein, 1997: The Arcadia, Oklahoma storm of 17 May 1981: Analysis of a supercell during tornadogenesis. *Mon. Wea. Rev.*, **125**, 2562-2582.

Dowell, D.C., and H.B. Bluestein, 2002: The 8 June 1995 McLean, Texas, storm. Part I: Observations of cyclic tornadogenesis. *Mon. Wea. Rev.*, **130**, 2626-2648.

Droegemeier, K.K., S.M. Lazarus, and R. Davies-Jones, 1993: The influence of helicity on numerically simulated convective storms. *Mon. Wea. Rev.*, **121**, 2005-2029.

Fujita, T., 1958: Mesoanalysis of the Illinois tornadoes of 9 April 1953. *J. Meteor.*, **15**, 288-296.

Fujita, T.T., 1973: Proposed mechanism of tornado formation from rotating thunderstorms. Preprints, *Eighth Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., 191-196.

Golden, J.T., and D. Purcell, 1977: Photogrammetric velocities for the Great Bend, Kansas tornado of 30 August, 1974: Accelerations and asymmetrics. *Mon. Wea. Rev.*, **105**, 485-492.

Golden, J.T., and D. Purcell, 1978: Airflow characteristics around the Union City tornado. *Mon. Wea. Rev.*, **106**, 22-38.

Hand, D., H. Mannila, and P. Smyth, 2001: Principles of Data Mining. MIT Press, Cambridge, MA.

Hane, C.E., and P.S. Ray, 1985: Pressure and buoyancy fields derived from Doppler radar data in a tornadic thunderstorm. *J. Atmos. Sci.*, **42**, 18-35.

Hu, M., and M. Xue, 2007: Impact of configurations of rapid intermittent assimilation of WSR-88D radar data for the 8 May 2003 Oklahoma City tornadic thunderstorm case. *Mon. Wea. Rev.*, **135**, 507-525.

Johnson, K.W., P.S. Ray, B.C. Johnson, and R.P. Davies-Jones, 1987: Observations related to the rotational dynamics of the 20 May 1977 tornadic storms. *Mon. Wea. Rev.*, **115**, 2463-2478.

Keogh, E., J. Lin, and A. Fu, 2005: HOT SAX: Efficiently finding the most unusual time series subsequence. *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM 2005)*, Houston, TX, 226-233.

Klemp, J.B., 1987: Dynamics of tornadic thunderstorms. *Annu. Rev. Fluid. Mech.*, **19**, 369-402.

Klemp, J.B., and R. Rotunno, 1983: A study of the tornadic region within a supercell thunderstorm. *J. Atmos. Sci.*, **40**, 359-377.

Klemp, J.B., and M.L. Weisman, 1983: The dependence of convective precipitation patterns on vertical wind shear. Preprints, *21st Conf. on Radar Meteorology*, Edmonton, AB, Canada, Amer. Meteor. Soc., 44-49.

Lemon, L.R., 1976: The flanking line, a severe thunderstorm intensification source. *J. Atmos. Sci.*, **33**, 686-694.

Lemon, L.R., 1977: Severe thunderstorm evolution: Its use in a new technique for radar warnings. Preprints, *10th Conf. on Severe Local Storms*, Omaha, NE, Amer. Meteor. Soc., 77-83.

Lemon, L.R., 1980: Severe thunderstorm radar identification techniques and warning criteria: A preliminary report. NOAA Tech. Memo. NWS NSSFC-1, 60 pp. [NTIS PB273049.]

Lemon, L.R., and C.A. Doswell, 1979: Severe thunderstorm evolution and mesocyclone structure as related to tornadogenesis. *Mon. Wea. Rev.*, **107**, 1184-1197.

Lemon, L.R., D.W. Burgess, and R.A. Brown, 1975: Tornado production and storm sustenance. Preprints, *Ninth Conf. on Severe Local Storms*, Norman, OK, Amer. Meteor. Soc., 100-104.

Lemon, L.R., D.W. Burgess, and R.A. Brown, 1978: Tornadic storm airflow and morphology derived from single Doppler radar measurements. *Mon. Wea. Rev.*, **106**, 48-61.

Lemon, L.R., R.J. Donaldson, Jr., D.W. Burgess, and R.A. Brown, 1977: Doppler radar application to severe thunderstorm study and potential real-time warning. *Bull. Amer. Meteor. Soc.*, **58**, 1187-1193.

Lilly, D.K., 1982: The development and maintenance of rotation in convective storms, *Intense Atmospheric Vortices*, L. Bengtsson and J. Lighthill, Eds., Springer-Verlag, 149-160.

Lilly, D.K., 1983: Dynamics of rotating thunderstorms. *Mesoscale Meteorology, Theories, Observations, Models*. D.K. Lilly and T. Gal-Chen, Eds., Reidel, 531-544.

Lin, J., E. Keogh, S. Lonardi, and B. Chiu, 2003: A symbolic representation of time series, with implications for streaming algorithms. *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, San Diego, CA.

Maddox, R.A., 1976: An evaluation of tornado proximity wind and stability data. *Mon. Wea. Rev.*, **104**, 133-142.

Markowski, P.M., 2002: Hook echoes and rear-flank downdrafts: A review. *Mon. Wea. Rev.*, **130**, 852-876.

Markowski, P., E. Rasmussen, J. Straka, R. Davies-Jones, Y. Richardson, and R.J. Trapp, 2008: Vortex lines within low-level mesocyclones obtained from pseudo-dual-Doppler radar observations. *Mon. Wea. Rev.*, **136**, 3513-3535.

Marwitz, J.D., 1972a: The structure and motion of severe hailstorms. Part I: Supercell storms. *J. Appl. Meteor.*, **11**, 166-179.

Marwitz, J.D., 1972b: The structure and motion of severe hailstorms. Part III: Severely sheared storms. *J. Appl. Meteor.*, **11**, 189-201.

McGovern, A., N. Hiers, M Collier, D.J. Gagne, II, and R.A. Brown, 2008: Spatiotemporal relational probability trees. *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008)*, Pisa, Italy.

Moller, A.R., 2001: Severe local storms forecasting. *Severe Convective Storms, Meteor. Monogr.*, No. 50, Amer. Meteor. Soc., 433-480.

Moller, A., C. Doswell, J. McGinley, S. Tegtmeier, and R. Zipser, 1974: Field observations of the Union City tornado in Oklahoma. *Weatherwise*, **27**, 68-77.

Nelson, S.P., 1977: Rear flank downdraft: A hailstorm intensification mechanism. Preprints, *10th Conf. on Severe Local Storms*, Omaha, NE, Amer. Meteor. Soc., 521-525.

Newton, C.W., and S. Katz, 1958: Movement of large convective rain storms in relation to winds aloft. *Bull. Amer. Meteor. Soc.*, **32**, 129-136.

OFCM- Office of the Federal Coordinator for Meteorological Services and Supporting Research, 2006: Doppler radar meteorological observations, Part C: WSR-88D products and algorithms. Federal Meteorological Handbook No. 11. U.S. Dept of Commerce/NOAA, Washington, D.C.

Polger, P.D., B.S. Goldsmith, R.C. Przywarty, and J.R. Bocchieri, 1994: National Weather Service warning performance based on the WSR-88D. *Bull. Amer. Meteor. Soc.*, **75**, 203-214.

Ray, P.S., 1976: Vorticity and divergence fields within tornadic storms from dual-Doppler observations. *J. Appl. Meteor.*, **15**, 879-890.

Ray, P.S., R.J. Doviak, G.B. Walker, D. Sirmans, J. Carter, and B. Bumgarner, 1975: Dual-Doppler observation of a tornadic storm. *J. Appl. Meteor.*, **14**, 1521-1530.

Rotunno, R., 1981: On the evolution of thunderstorm rotation. *Mon. Wea. Rev.*, **109**, 577-586.

Rotunno, R., 1993: Supercell thunderstorm modeling and theory. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr.*, No. 79, Amer. Geophys. Union, 57-73.

Rotunno, R., and J.B. Klemp, 1982: The influence of the shear-induced pressure gradient on thunderstorm motion. *Mon. Wea. Rev.*, **110**, 136-151.

Rotunno, R., and J.B. Klemp, 1985: On the rotation and propagation of simulated supercell thunderstorms. *J. Atmos. Sci.*, **42**, 271-292.

Shabbott, C.J., and P.M. Markowski, 2006: Surface in situ observations within the outflow of forward-flank downdrafts of supercell thunderstorms. *Mon. Wea. Rev.*, **134**, 1422-1441.

Simmons, K.M., and D. Sutter, 2005: WSR-88D radar, tornado warnings, and tornado casualties. *Wea. and Forecasting*, **20**, 301-310.

Stout, G.E., and F.A. Huff, 1953: Radar records Illinois tornadogenesis. *Bull. Amer. Meteor. Soc.*, **34**, 281-284.

Straka, J.M., E.N. Rasmussen, R.P. Davies-Jones, and P.M. Markowski, 2007: An observational and idealized numerical examination of low-level counter-rotating vortices in the rear flank of supercells. *Electronic J. Severe Storms Meteor.*, **2** (8), 1-22.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiment. *Mon. Wea. Rev.*, **133**, 1789-1807.

Trapp, R.J., and B.H. Fiedler, 1995: Tornado-like vortexgenesis in a simplified numerical model. *J. Atmos. Sci.*, **52**, 3757-3778.

van Tassell, E.L., 1955: The North Platte Valley tornado outbreak of June 27, 1955. *Mon. Wea. Rev.*, **83**, 255-264.

Wakimoto, R.M., 2001: Convectively driven high wind events. *Severe Convective Storms, Meteor. Monogr.*, No. 50, Amer. Meteor. Soc., 255-298.

Wakimoto, R.M., W.-C. Lee, H.B. Bluestein, C.-H. Liu, and P.H. Hildebrand, 1996: ELDORA observations during VORTEX 95. *Bull. Amer. Meteor. Soc.*, **77**, 1465-1481.

Walko, R.L., 1993: Tornado spin-up beneath a convective cell: Required basic structure of the near-field boundary layer winds. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr.*, No. 79, Amer. Geophys. Union, 89-95.

WDTB - Warning Decision Training Branch, 2002: Tornado warning guidance. National Weather Service, <http://www.wdtb.noaa.gov/modules/twg02/>

Weisman, M.L., and J.B. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, **110**, 504-520.

Weisman, M.L., and J.B. Klemp, 1984: The structure and classification of numerically simulated convective storms in directionally varying wind shears. *Mon. Wea. Rev.*, **112**, 2479-2498.

Weisman, M.L., and J.B. Klemp, 1986: Characteristics of isolated convective storms. *Mesoscale Meteorology Forecasting*, P.S. Ray, Ed., Amer. Meteor. Soc., 331-358.

Wicker, L.J., and R.B. Wilhelmson, 1995: Simulation and analysis of tornado development and decay within a three-dimensional supercell thunderstorm. *J. Atmos. Sci.*, **52**, 2675-2703.

Wilhelmson, R.B., and J.B. Klemp, 1978: A three-dimensional numerical simulation of splitting that leads to long-lived storms. *J. Atmos. Sci.*, **35**, 1037-1063.

Wilks, D.S., 2006: Statistical Methods in the Atmospheric Sciences, 2nd ed. Elsevier Inc., 627 pp.

Wood, V.T., and R.A. Brown, 1997: Effects of radar sampling on single-Doppler velocity signatures of mesocyclones and tornadoes. *Wea. and Forecasting*, **12**, 928-938.

Xue, M., F. Kong, D. Weber, K.W. Thomas, Y. Wang, K. Brewster, K.K. Droegemeier, J.S. Kain, S.J. Weiss, D.R. Bright, M.S. Wandishin, M.C. Coniglio, and J. Du, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA hazardous weather testbed 2007 spring experiment. Preprints, *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 3B.1.

Xue, M., K.K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) – A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Physics.*, **75**, 161-193.

Xue, M., K.K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D.H. Wang, 2001: The Advanced Regional Prediction System (ARPS) – A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor Atmos. Physics.*, **76**, 143-165.

Xue, M., D.H. Wang, J.D. Gao, K. Brewster, and K.K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor Atmos. Physics.*, **82**, 139-170.

APPENDIX A

NUMERICAL MODEL SIMULATION EXPERIMENTS

Table A.1 provides a summary of the numerical model simulation experiments including initial environments and their derived parameters as well as the number of generated storms in each, labeled according to the development of strong low-level rotation.[25] The simulation name first identifies the Weisman and Klemp surface mixing ratio (i.e., "WK" 13, 14, 15, 16 or 17 g kg^{-1}) followed by the hodograph shape (i.e., half circle (half), quarter circle (qtr) or quarter circle with tail (qtr_tail)) and its radius (i.e., "r" 8, 10, 15, 20, 25, 30, 35 m s^{-1}). Then the tail length through 10 km, for quarter tail hodographs only, is listed (i.e., "t" 10, 20, 40 m s^{-1}) followed by the half or quarter circle turning depth for each (i.e., 1, 3, 4, 5, 6, 8, 10 "km"). . Convective available potential energy (CAPE) is then listed for calculations using temperature (T) only and then virtual temperature (T_v) and water loading (WL). Next the bulk Richardson number (BRN) shear term and BRN with both versions of CAPE are listed followed by storm relative helicity (SRH) from 0-1 km and 0-3 km. Each of these parameters are discussed further in Section 3.1.2. The number of "Positive," "Intermediate" and "Negative storms generated by each simulation are then listed.

[25] Five additional simulation experiments were originally intended to be included in the overall list of experiments but data loss during the transfer process required their omission. These experiments included the following: WK13_qtr_r30_5km, WK13_qtr_r25_3km, WK14_qtr_r10_3km, WK14_qtr_r8_1km, WK16_qtr_tail_r10_t10_3km.

Table A.1: Summary of simulation experiments. Initial environmental identifiers are listed (Simulation Name – with naming scheme provided in preceding paragraph) followed by CAPE (using T and then T_v and water loading in brackets), BRN shear, BRN (with CAPE calculated using T and then T_v and water loading in brackets), SRH (for 0-1 km and then for 0-3 km in brackets) and the number of "Positive," "Intermediate" and "Negative" storms identified in each simulation associated with the development of strong, weak and no low-level rotation respectively.

Simulation Name	CAPE ($J\ kg^{-1}$) T [T_v, WL]	BRN shear ($m\ s^{-1}$)	BRN T [T_v, WL]	SRH ($m^2\ s^{-2}$) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK13_half_r8_4km	1284 [686]	9.76	26.95 [14.39]	60.8 [141.7]			3
WK13_half_r8_6km	1284 [686]	8.06	39.58 [21.14]	35.29 [82.82]			3
WK13_half_r8_8km	1284 [686]	6.74	56.6 [30.22]	22.82 [53.35]			4
WK13_half_r8_10km	1284 [686]	5.67	79.98 [42.71]	15.99 [37.92]			1
WK13_half_r10_4km	1284 [686]	12.2	17.25 [9.21]	95 [221.41]			5
WK13_half_r10_6km	1284 [686]	10.07	25.33 [13.53]	55.14 [129.41]		1	4
WK13_half_r10_8km	1284 [686]	8.42	36.22 [19.34]	35.66 [83.37]			3
WK13_half_r10_10km	1284 [686]	7.08	51.19 [27.33]	24.68 [58.33]			2
WK13_half_r15_4km	1284 [686]	18.3	7.67 [4.09]	213.74 [498.17]		1	2
WK13_half_r15_6km	1284 [686]	15.1	11.26 [6.01]	122.27 [284.66]		2	2
WK13_half_r15_8km	1284 [686]	12.63	16.1 [8.6]	75.55 [172.85]		1	4
WK13_half_r15_10km	1284 [686]	10.63	22.75 [12.15]	50.54 [116.33]		2	1
WK13_half_r20_6km	1284 [686]	20.14	6.33 [3.38]	210.68 [482.98]		3	2
WK13_half_r20_8km	1284 [686]	16.84	9.06 [4.84]	125 [279.31]		1	3
WK13_half_r20_10km	1284 [686]	14.17	12.8 [6.83]	80.39 [179.6]			3
WK13_half_r25_6km	1284 [686]	25.17	4.05 [2.16]	317.7 [717.11]	1	1	
WK13_half_r25_8km	1284 [686]	21.05	5.8 [3.09]	179.76 [391.55]		2	
WK13_half_r25_10km	1284 [686]	17.71	8.19 [4.37]	115.67 [252.82]		1	2
WK13_half_r30_8km	1284 [686]	25.26	4.02 [2.15]	257.14 [558.98]	2	1	
WK13_half_r30_10km	1284 [686]	21.25	5.69 [3.04]	166.56 [364.06]	1	1	
WK13_qtr_r8_1km	1284 [686]	7.56	44.97 [24.01]	76.06 [77.26]			
WK13_qtr_r8_3km	1284 [686]	7.13	50.57 [27]	30.14 [62.56]			2
WK13_qtr_r8_5km	1284 [686]	5.58	82.35 [43.97]	15.71 [37.03]			1

Table A.1: Continued

Simulation Name	CAPE (J kg^{-1}) T [T_v, WL]	BRN shear (m s^{-1})	BRN T [T_v, WL]	SRH (m^2 s^{-2}) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK13_qtr_r10_1km	1284 [686]	9.45	28.78 [15.37]	115.7 [117.46]			
WK13_qtr_r10_3km	1284 [686]	8.91	32.36 [17.28]	46.59 [96.24]			5
WK13_qtr_r10_5km	1284 [686]	6.98	52.71 [28.14]	24.17 [56.69]			5
WK13_qtr_r15_1km	1284 [686]	14.17	12.79 [6.83]	234.06 [237.21]			4
WK13_qtr_r15_3km	1284 [686]	13.36	14.38 [7.68]	96.97 [193.59]			4
WK13_qtr_r15_5km	1284 [686]	10.47	23.43 [12.51]	49.18 [112.11]		1	2
WK13_qtr_r20_1km	1284 [686]	18.89	7.19 [3.84]	366.43 [370.61]			4
WK13_qtr_r20_3km	1284 [686]	17.82	8.09 [4.32]	157.15 [302.36]		3	1
WK13_qtr_r20_5km	1284 [686]	13.96	13.18 [7.04]	77.58 [171.16]		1	5
WK13_qtr_r25_1km	1284 [686]	23.62	4.6 [2.46]	549.28 [555.19]			5
WK13_qtr_r25_5km	1284 [686]	17.45	8.43 [4.5]	112.19 [242.39]		3	3
WK13_qtr_r30_3km	1284 [686]	26.73	3.6 [1.92]	330.97 [621.16]		4	1
WK13_qtr_r35_5km	1284 [686]	24.43	4.3 [2.3]	219.9 [475.08]		4	1
WK13_qtr_tail_r10_t10_1km	1284 [686]	11.15	20.67 [11.04]	136.74 [148.09]			2
WK13_qtr_tail_r10_t10_3km	1284 [686]	9.53	28.3 [15.11]	50.12 [106.34]			3
WK13_qtr_tail_r10_t20_1km	1284 [686]	12.94	15.34 [8.19]	156.94 [180.97]		2	
WK13_qtr_tail_r10_t20_3km	1284 [686]	10.18	24.77 [13.23]	53.54 [116.27]			3
WK13_qtr_tail_r10_t40_1km	1284 [686]	16.68	9.23 [4.93]	196.07 [252.33]		2	2
WK13_qtr_tail_r10_t40_3km	1284 [686]	11.58	19.15 [10.23]	60.04 [135.28]			4
WK13_qtr_tail_r15_t10_1km	1284 [686]	15.85	10.22 [5.46]	267.48 [281.62]			3
WK13_qtr_tail_r15_t10_3km	1284 [686]	13.97	13.15 [7.02]	102.98 [209.93]		1	2
WK13_qtr_tail_r15_t20_1km	1284 [686]	17.61	8.29 [4.42]	299.3 [327.3]		1	3
WK13_qtr_tail_r15_t20_3km	1284 [686]	14.61	12.03 [6.42]	108.82 [225.99]		1	3
WK13_qtr_tail_r15_t40_1km	1284 [686]	21.25	5.69 [3.04]	359.13 [420.96]		2	3
WK13_qtr_tail_r15_t40_3km	1284 [686]	15.96	10.09 [5.39]	120.07 [257.29]		3	
WK13_qtr_tail_r20_t10_1km	1284 [686]	20.57	6.07 [3.24]	412.99 [428.09]			3
WK13_qtr_tail_r20_t10_3km	1284 [686]	18.43	7.57 [4.04]	166.13 [325.55]			3

Table A.1: Continued

Simulation Name	CAPE ($J\ kg^{-1}$) T [T_v, WL]	BRN shear ($m\ s^{-1}$)	BRN T [T_v, WL]	SRH ($m^2\ s^{-2}$) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK13_qtr_tail_r20_t20_1km	1284 [686]	22.29	5.17 [2.76]	457.2 [485.71]		1	2
WK13_qtr_tail_r20_t20_3km	1284 [686]	19.05	7.07 [3.78]	174.89 [348.38]		1	3
WK13_qtr_tail_r20_t40_1km	1284 [686]	25.88	3.83 [2.05]	539.14 [599.65]		1	1
WK13_qtr_tail_r25_t10_1km	1284 [686]	25.28	4.02 [2.15]	605.28 [623.16]			5
WK13_qtr_tail_r25_t10_3km	1284 [686]	22.88	4.91 [2.62]	239.87 [456.17]			3
WK13_qtr_tail_r25_t20_1km	1284 [686]	27	3.52 [1.88]	661.32 [693.59]			3
WK13_qtr_tail_r25_t20_3km	1284 [686]	23.5	4.65 [2.48]	249.89 [481.03]		3	2
WK14_half_r8_4km	1769 [1033]	9.76	37.12 [21.68]	60.8 [141.71]		1	4
WK14_half_r8_6km	1769 [1033]	8.06	54.52 [31.84]	35.3 [82.83]			8
WK14_half_r8_8km	1769 [1033]	6.74	77.95 [45.53]	22.82 [53.36]			6
WK14_half_r10_4km	1769 [1033]	12.2	23.76 [13.88]	95.01 [221.43]	1	1	
WK14_half_r10_6km	1769 [1033]	10.07	34.89 [20.38]	55.15 [129.42]			5
WK14_half_r10_8km	1769 [1033]	8.42	49.89 [29.14]	35.66 [83.37]			7
WK14_half_r10_10km	1769 [1033]	7.08	70.5 [41.18]	24.68 [58.33]			4
WK14_half_r15_4km	1769 [1033]	18.31	10.56 [6.17]	213.76 [498.21]	1	1	2
WK14_half_r15_6km	1769 [1033]	15.11	15.51 [9.06]	122.28 [284.68]		1	3
WK14_half_r15_8km	1769 [1033]	12.63	22.17 [12.95]	75.55 [172.87]		1	2
WK14_half_r15_10km	1769 [1033]	10.63	31.33 [18.3]	50.54 [116.34]		1	3
WK14_half_r20_6km	1769 [1033]	20.14	8.72 [5.09]	210.7 [483.02]			3
WK14_half_r20_8km	1769 [1033]	16.84	12.47 [7.29]	125.02 [279.33]	1		2
WK14_half_r20_10km	1769 [1033]	14.17	17.63 [10.3]	80.4 [179.61]		2	3
WK14_half_r25_6km	1769 [1033]	25.18	5.58 [3.26]	317.74 [717.18]	2		
WK14_half_r25_8km	1769 [1033]	21.05	7.98 [4.66]	179.78 [391.59]	2	2	1
WK14_half_r25_10km	1769 [1033]	17.71	11.28 [6.59]	115.68 [252.84]		1	2
WK14_half_r30_8km	1769 [1033]	25.26	5.54 [3.24]	257.17 [559.04]	2	1	
WK14_qtr_r8_3km	1769 [1033]	7.13	69.65 [40.68]	30.14 [62.57]			3
WK14_qtr_r10_1km	1769 [1033]	9.45	39.64 [23.16]	115.7 [117.47]			2

Table A.1: Continued

Simulation Name	CAPE ($J\ kg^{-1}$) T [T_V, WL]	BRN shear ($m\ s^{-1}$)	BRN T [T_V, WL]	SRH ($m^2\ s^{-2}$) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK14_qtr_r10_5km	1769 [1033]	6.98	72.59 [42.4]	24.17 [56.7]			5
WK14_qtr_r15_1km	1769 [1033]	14.17	17.62 [10.29]	234.07 [237.22]			3
WK14_qtr_r15_3km	1769 [1033]	13.36	19.81 [11.57]	96.98 [193.61]		1	2
WK14_qtr_r15_5km	1769 [1033]	10.47	32.26 [18.85]	49.18 [112.12]		1	3
WK14_qtr_r20_1km	1769 [1033]	18.89	9.91 [5.79]	366.45 [370.63]			4
WK14_qtr_r20_3km	1769 [1033]	17.82	11.14 [6.51]	157.17 [302.38]		1	4
WK14_qtr_r20_5km	1769 [1033]	13.96	18.15 [10.6]	77.59 [171.18]		1	3
WK14_qtr_r25_1km	1769 [1033]	23.62	6.34 [3.7]	549.31 [555.23]			8
WK14_qtr_r25_3km	1769 [1033]	22.27	7.13 [4.17]	229.87 [431.4]		1	3
WK14_qtr_r25_5km	1769 [1033]	17.45	11.61 [6.78]	112.21 [242.42]		2	2
WK14_qtr_r30_3km	1769 [1033]	26.73	4.95 [2.89]	331.01 [621.21]		2	3
WK14_qtr_r30_5km	1769 [1033]	20.94	8.07 [4.71]	161.58 [349.08]		3	1
WK14_qtr_r35_5km	1769 [1033]	24.44	5.93 [3.46]	219.93 [475.13]		3	1
WK14_qtr_tail_r10_t10_1km	1769 [1033]	11.15	28.47 [16.63]	136.75 [148.1]			4
WK14_qtr_tail_r10_t10_3km	1769 [1033]	9.53	38.97 [22.77]	50.12 [106.35]			4
WK14_qtr_tail_r10_t20_1km	1769 [1033]	12.94	21.13 [12.34]	156.96 [180.99]		1	3
WK14_qtr_tail_r10_t20_3km	1769 [1033]	10.18	34.12 [19.93]	53.54 [116.28]			4
WK14_qtr_tail_r10_t40_1km	1769 [1033]	16.68	12.71 [7.43]	196.09 [252.35]		2	2
WK14_qtr_tail_r10_t40_3km	1769 [1033]	11.58	26.38 [15.41]	60.05 [135.29]		2	4
WK14_qtr_tail_r15_t10_1km	1769 [1033]	15.85	14.08 [8.22]	267.49 [281.64]			5
WK14_qtr_tail_r15_t10_3km	1769 [1033]	13.98	18.11 [10.58]	102.99 [209.94]		1	4
WK14_qtr_tail_r15_t20_1km	1769 [1033]	17.61	11.41 [6.67]	299.32 [327.32]			2
WK14_qtr_tail_r15_t20_3km	1769 [1033]	14.61	16.57 [9.68]	108.83 [226.01]		3	
WK14_qtr_tail_r15_t40_1km	1769 [1033]	21.26	7.83 [4.57]	359.16 [420.99]		1	2
WK14_qtr_tail_r15_t40_3km	1769 [1033]	15.96	13.9 [8.12]	120.08 [257.31]		3	2
WK14_qtr_tail_r20_t10_1km	1769 [1033]	20.57	8.36 [4.89]	413.01 [428.11]			2
WK14_qtr_tail_r20_t10_3km	1769 [1033]	18.43	10.42 [6.09]	166.15 [325.58]		2	2

Table A.1: Continued

Simulation Name	CAPE (J kg^{-1}) T [T_V, WL]	BRN shear (m s^{-1})	BRN T [T_V, WL]	SRH (m^2 s^{-2}) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK14_qtr_tail_r20_t20_1km	1769 [1033]	22.3	7.12 [4.16]	457.23 [485.73]		2	2
WK14_qtr_tail_r20_t20_3km	1769 [1033]	19.06	9.74 [5.69]	174.91 [348.41]		2	2
WK14_qtr_tail_r20_t40_1km	1769 [1033]	25.88	5.28 [3.09]	539.17 [599.68]		3	1
WK14_qtr_tail_r20_t40_3km	1769 [1033]	20.37	8.53 [4.98]	191.79 [392.93]	2	1	2
WK14_qtr_tail_r25_t10_1km	1769 [1033]	25.28	5.53 [3.23]	605.32 [623.2]			4
WK14_qtr_tail_r25_t10_3km	1769 [1033]	22.88	6.76 [3.95]	239.9 [456.21]		5	1
WK14_qtr_tail_r25_t20_1km	1769 [1033]	27	4.85 [2.84]	661.37 [693.64]		1	3
WK14_qtr_tail_r25_t20_3km	1769 [1033]	23.5	6.41 [3.74]	249.92 [481.08]		2	2
WK15_half_r8_4km	2307 [1427]	9.76	48.4 [29.93]	60.81 [141.72]			7
WK15_half_r8_6km	2307 [1427]	8.06	71.08 [43.96]	35.3 [82.84]			5
WK15_half_r10_4km	2307 [1427]	12.21	30.97 [19.16]	95.01 [221.45]	1	1	5
WK15_half_r10_6km	2307 [1427]	10.07	45.49 [28.14]	55.15 [129.43]			6
WK15_half_r10_8km	2307 [1427]	8.42	65.05 [40.23]	35.66 [83.38]			5
WK15_half_r15_4km	2307 [1427]	18.31	13.77 [8.51]	213.78 [498.25]		2	2
WK15_half_r15_6km	2307 [1427]	15.11	20.22 [12.5]	122.29 [284.7]		1	5
WK15_half_r15_8km	2307 [1427]	12.63	28.91 [17.88]	75.56 [172.88]		2	3
WK15_half_r15_10km	2307 [1427]	10.63	40.85 [25.27]	50.55 [116.35]		1	2
WK15_half_r20_6km	2307 [1427]	20.14	11.37 [7.03]	210.72 [483.06]		3	1
WK15_half_r20_8km	2307 [1427]	16.84	16.26 [10.06]	125.03 [279.36]		2	1
WK15_half_r20_10km	2307 [1427]	14.17	22.98 [14.21]	80.41 [179.63]	1	2	
WK15_half_r25_6km	2307 [1427]	25.18	7.28 [4.5]	317.77 [717.24]	2		
WK15_half_r25_8km	2307 [1427]	21.06	10.41 [6.44]	179.8 [391.62]	1	1	2
WK15_half_r25_10km	2307 [1427]	17.71	14.71 [9.1]	115.69 [252.87]	1	2	2
WK15_qtr_r8_1km	2307 [1427]	7.56	80.77 [49.95]	76.07 [77.27]			3
WK15_qtr_r10_1km	2307 [1427]	9.45	51.69 [31.97]	115.71 [117.47]			3
WK15_qtr_r10_3km	2307 [1427]	8.91	58.11 [35.94]	46.6 [96.26]			5
WK15_qtr_r15_1km	2307 [1427]	14.17	22.97 [14.21]	234.08 [237.23]			6

Table A.1: Continued

Simulation Name	CAPE (J kg^{-1}) T [T_v, WL]	BRN shear (m s^{-1})	BRN T [T_v, WL]	SRH (m^2 s^{-2}) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK15_qtr_r15_3km	2307 [1427]	13.37	25.83 [15.98]	96.99 [193.62]		2	3
WK15_qtr_r15_5km	2307 [1427]	10.47	42.06 [26.02]	49.19 [112.13]		2	2
WK15_qtr_r20_1km	2307 [1427]	18.9	12.92 [7.99]	366.46 [370.64]			4
WK15_qtr_r20_3km	2307 [1427]	17.82	14.53 [8.99]	157.18 [302.39]		2	3
WK15_qtr_r20_5km	2307 [1427]	13.96	23.66 [14.63]	77.59 [171.19]		4	1
WK15_qtr_r25_1km	2307 [1427]	23.62	8.27 [5.12]	549.34 [555.26]			4
WK15_qtr_r25_3km	2307 [1427]	22.28	9.3 [5.75]	229.89 [431.43]	1	5	3
WK15_qtr_r25_5km	2307 [1427]	17.46	15.14 [9.37]	112.22 [242.44]		3	1
WK15_qtr_r30_3km	2307 [1427]	26.73	6.46 [3.99]	331.04 [621.26]	3	3	
WK15_qtr_r30_5km	2307 [1427]	20.95	10.52 [6.5]	161.59 [349.11]		4	2
WK15_qtr_r35_5km	2307 [1427]	24.44	7.73 [4.78]	219.95 [475.18]		1	1
WK15_qtr_tail_r10_t10_1km	2307 [1427]	11.15	37.12 [22.96]	136.76 [148.1]			6
WK15_qtr_tail_r10_t10_3km	2307 [1427]	9.53	50.81 [31.43]	50.12 [106.35]			4
WK15_qtr_tail_r10_t20_1km	2307 [1427]	12.94	27.55 [17.04]	156.97 [181]		1	4
WK15_qtr_tail_r10_t20_3km	2307 [1427]	10.19	44.48 [27.51]	53.55 [116.29]		1	3
WK15_qtr_tail_r10_t40_1km	2307 [1427]	16.69	16.57 [10.25]	196.11 [252.37]		4	1
WK15_qtr_tail_r10_t40_3km	2307 [1427]	11.58	34.39 [21.27]	60.05 [135.3]		3	1
WK15_qtr_tail_r15_t10_1km	2307 [1427]	15.85	18.36 [11.35]	267.5 [281.65]			3
WK15_qtr_tail_r15_t10_3km	2307 [1427]	13.98	23.62 [14.61]	102.99 [209.96]		2	4
WK15_qtr_tail_r15_t20_1km	2307 [1427]	17.61	14.88 [9.2]	299.34 [327.33]	1		4
WK15_qtr_tail_r15_t20_3km	2307 [1427]	14.62	21.6 [13.36]	108.84 [226.02]		2	3
WK15_qtr_tail_r15_t40_1km	2307 [1427]	21.26	10.21 [6.32]	359.18 [421.01]		2	2
WK15_qtr_tail_r15_t40_3km	2307 [1427]	15.96	18.12 [11.21]	120.09 [257.33]		3	2
WK15_qtr_tail_r20_t10_1km	2307 [1427]	20.57	10.91 [6.75]	413.03 [428.13]		1	4
WK15_qtr_tail_r20_t10_3km	2307 [1427]	18.43	13.59 [8.4]	166.17 [325.6]		3	2
WK15_qtr_tail_r20_t20_1km	2307 [1427]	22.3	9.28 [5.74]	457.25 [485.76]		2	2
WK15_qtr_tail_r20_t20_3km	2307 [1427]	19.06	12.7 [7.86]	174.93 [348.43]		3	1

Table A.1: Continued

Simulation Name	CAPE (J kg^{-1}) T [T_v, WL]	BRN shear (m s^{-1})	BRN T [T_v, WL]	SRH (m^2 s^{-2}) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK15_qtr_tail_r20_t40_1km	2307 [1427]	25.89	6.89 [4.26]	539.21 [599.71]	1		2
WK15_qtr_tail_r20_t40_3km	2307 [1427]	20.37	11.12 [6.88]	191.81 [392.96]	2		1
WK15_qtr_tail_r25_t10_1km	2307 [1427]	25.29	7.22 [4.46]	605.36 [623.24]		2	3
WK15_qtr_tail_r25_t10_3km	2307 [1427]	22.88	8.81 [5.45]	239.92 [456.25]		5	1
WK15_qtr_tail_r25_t20_1km	2307 [1427]	27	6.33 [3.91]	661.42 [693.69]		1	5
WK15_qtr_tail_r25_t20_3km	2307 [1427]	23.51	8.35 [5.17]	249.95 [481.12]		3	1
WK16_half_r8_4km	2843 [1863]	9.77	59.63 [39.08]	60.81 [141.73]			5
WK16_half_r10_4km	2843 [1863]	12.21	38.16 [25.01]	95.02 [221.46]		2	4
WK16_half_r10_6km	2843 [1863]	10.07	56.05 [36.73]	55.16 [129.44]		2	3
WK16_half_r15_4km	2843 [1863]	18.31	16.96 [11.11]	213.8 [498.28]	1	1	
WK16_half_r15_6km	2843 [1863]	15.11	24.91 [16.32]	122.3 [284.72]		3	3
WK16_half_r15_8km	2843 [1863]	12.64	35.62 [23.34]	75.57 [172.89]		2	2
WK16_half_r15_10km	2843 [1863]	10.63	50.33 [32.98]	50.55 [116.36]		2	3
WK16_half_r20_6km	2843 [1863]	20.15	14.01 [9.18]	210.74 [483.09]	1	2	3
WK16_half_r20_8km	2843 [1863]	16.85	20.04 [13.13]	125.04 [279.37]	1		2
WK16_half_r20_10km	2843 [1863]	14.17	28.31 [18.55]	80.41 [179.64]		2	2
WK16_half_r25_6km	2843 [1863]	25.18	8.97 [5.88]	317.79 [717.28]	1	2	3
WK16_half_r25_8km	2843 [1863]	21.06	12.82 [8.4]	179.81 [391.64]	1	1	
WK16_half_r25_10km	2843 [1863]	17.71	18.12 [11.87]	115.7 [252.89]	1	3	1
WK16_half_r30_8km	2843 [1863]	25.27	8.9 [5.84]	257.22 [559.14]	2	2	
WK16_qtr_r10_1km	2843 [1863]	9.45	63.69 [41.74]	115.71 [117.48]			3
WK16_qtr_r10_3km	2843 [1863]	8.91	71.6 [46.92]	46.6 [96.26]		1	7
WK16_qtr_r15_1km	2843 [1863]	14.17	28.31 [18.55]	234.09 [237.24]			2
WK16_qtr_r15_3km	2843 [1863]	13.37	31.82 [20.85]	96.99 [193.63]		1	6
WK16_qtr_r15_5km	2843 [1863]	10.47	51.83 [33.96]	49.19 [112.14]		2	1
WK16_qtr_r20_1km	2843 [1863]	18.9	15.92 [10.43]	366.48 [370.66]			3
WK16_qtr_r20_3km	2843 [1863]	17.82	17.9 [11.73]	157.19 [302.41]	1	2	1

Table A.1: Continued

Simulation Name	CAPE (J kg^{-1}) T [T_V, WL]	BRN shear (m s^{-1})	BRN T [T_V, WL]	SRH (m^2 s^{-2}) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK16_qtr_r20_5km	2843 [1863]	13.97	29.15 [19.1]	77.6 [171.2]		2	3
WK16_qtr_r25_1km	2843 [1863]	23.62	10.19 [6.68]	549.37 [555.29]			5
WK16_qtr_r25_3km	2843 [1863]	22.28	11.46 [7.51]	229.91 [431.45]		4	2
WK16_qtr_r25_5km	2843 [1863]	17.46	18.66 [12.23]	112.23 [242.45]		1	3
WK16_qtr_r30_3km	2843 [1863]	26.73	7.96 [5.21]	331.07 [621.29]		5	2
WK16_qtr_r30_5km	2843 [1863]	20.95	12.96 [8.49]	161.61 [349.13]		4	3
WK16_qtr_r35_5km	2843 [1863]	24.44	9.52 [6.24]	219.97 [475.21]	1	1	4
WK16_qtr_tail_r10_t10_1km	2843 [1863]	11.15	45.74 [29.97]	136.76 [148.11]			4
WK16_qtr_tail_r10_t20_1km	2843 [1863]	12.94	33.94 [22.24]	156.97 [181.01]		2	3
WK16_qtr_tail_r10_t20_3km	2843 [1863]	10.19	54.8 [35.91]	53.55 [116.3]		2	4
WK16_qtr_tail_r10_t40_1km	2843 [1863]	16.69	20.42 [13.38]	196.12 [252.38]		4	
WK16_qtr_tail_r10_t40_3km	2843 [1863]	11.58	42.38 [27.77]	60.06 [135.31]	1	2	4
WK16_qtr_tail_r15_t10_1km	2843 [1863]	15.86	22.62 [14.82]	267.52 [281.66]		2	5
WK16_qtr_tail_r15_t10_3km	2843 [1863]	13.98	29.1 [19.07]	103 [209.97]		2	4
WK16_qtr_tail_r15_t20_1km	2843 [1863]	17.61	18.34 [12.02]	299.35 [327.35]			6
WK16_qtr_tail_r15_t20_3km	2843 [1863]	14.62	26.61 [17.44]	108.85 [226.03]		1	6
WK16_qtr_tail_r15_t40_1km	2843 [1863]	21.26	12.58 [8.24]	359.2 [421.03]	1	2	1
WK16_qtr_tail_r15_t40_3km	2843 [1863]	15.96	22.32 [14.63]	120.1 [257.34]	1	3	1
WK16_qtr_tail_r20_t10_1km	2843 [1863]	20.57	13.44 [8.81]	413.05 [428.15]		2	6
WK16_qtr_tail_r20_t10_3km	2843 [1863]	18.43	16.74 [10.97]	166.18 [325.61]		3	4
WK16_qtr_tail_r20_t20_1km	2843 [1863]	22.3	11.43 [7.49]	457.28 [485.78]		1	2
WK16_qtr_tail_r20_t20_3km	2843 [1863]	19.06	15.65 [10.26]	174.94 [348.45]		1	7
WK16_qtr_tail_r20_t40_1km	2843 [1863]	25.89	8.49 [5.56]	539.23 [599.73]	1	3	
WK16_qtr_tail_r25_t10_1km	2843 [1863]	25.29	8.89 [5.83]	605.4 [623.28]		1	6
WK16_qtr_tail_r25_t10_3km	2843 [1863]	22.89	10.86 [7.11]	239.94 [456.27]		4	1
WK16_qtr_tail_r25_t20_1km	2843 [1863]	27	7.8 [5.11]	661.47 [693.73]		3	4
WK16_qtr_tail_r25_t20_3km	2843 [1863]	23.51	10.29 [6.74]	249.97 [481.15]	1	3	2

Table A.1: Continued

Simulation Name	CAPE (J kg^{-1}) T [T_v, WL]	BRN shear (m s^{-1})	BRN T [T_v, WL]	SRH (m^2 s^{-2}) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK17_half_r8_4km	3407 [2338]	9.77	71.44 [49.03]	60.82 [141.74]			6
WK17_half_r10_4km	3407 [2338]	12.21	45.72 [31.38]	95.03 [221.47]		2	5
WK17_half_r10_6km	3407 [2338]	10.07	67.14 [46.09]	55.16 [129.45]		2	6
WK17_half_r15_4km	3407 [2338]	18.31	20.32 [13.95]	213.81 [498.3]		4	3
WK17_half_r15_6km	3407 [2338]	15.11	29.84 [20.48]	122.31 [284.74]		1	3
WK17_half_r15_8km	3407 [2338]	12.64	42.67 [29.29]	75.57 [172.9]		3	4
WK17_half_r15_10km	3407 [2338]	10.63	60.3 [41.39]	50.55 [116.37]	1	2	1
WK17_half_r20_6km	3407 [2338]	20.15	16.79 [11.52]	210.75 [483.12]	2	1	1
WK17_half_r20_8km	3407 [2338]	16.85	24 [16.47]	125.05 [279.39]		5	2
WK17_half_r20_10km	3407 [2338]	14.17	33.92 [23.28]	80.42 [179.65]		1	4
WK17_half_r25_6km	3407 [2338]	25.18	10.74 [7.37]	317.81 [717.32]	2		1
WK17_half_r25_8km	3407 [2338]	21.06	15.36 [10.54]	179.83 [391.66]	1	1	1
WK17_half_r25_10km	3407 [2338]	17.72	21.71 [14.9]	115.71 [252.9]	2	1	5
WK17_qtr_r15_1km	3407 [2338]	14.17	33.91 [23.28]	234.1 [237.24]			5
WK17_qtr_r15_3km	3407 [2338]	13.37	38.13 [26.17]	97 [193.64]		3	7
WK17_qtr_r15_5km	3407 [2338]	10.48	62.09 [42.62]	49.19 [112.14]		3	2
WK17_qtr_r20_1km	3407 [2338]	18.9	19.08 [13.09]	366.49 [370.67]			8
WK17_qtr_r20_3km	3407 [2338]	17.82	21.45 [14.72]	157.2 [302.42]	1	2	2
WK17_qtr_r20_5km	3407 [2338]	13.97	34.93 [23.97]	77.6 [171.21]		3	1
WK17_qtr_r25_1km	3407 [2338]	23.62	12.21 [8.38]	549.4 [555.32]		2	4
WK17_qtr_r25_3km	3407 [2338]	22.28	13.73 [9.42]	229.93 [431.47]		3	4
WK17_qtr_r25_5km	3407 [2338]	17.46	22.35 [15.34]	112.23 [242.47]		4	2
WK17_qtr_r30_3km	3407 [2338]	26.74	9.53 [6.54]	331.1 [621.32]	1	2	2
WK17_qtr_r30_5km	3407 [2338]	20.95	15.52 [10.65]	161.62 [349.16]		2	3
WK17_qtr_r35_5km	3407 [2338]	24.44	11.4 [7.83]	219.98 [475.24]	1	1	1
WK17_qtr_tail_r10_t10_1km	3407 [2338]	11.15	54.79 [37.61]	136.77 [148.12]			5
WK17_qtr_tail_r10_t20_1km	3407 [2338]	12.94	40.66 [27.91]	156.98 [181.01]		2	4

Table A.1: Continued

Simulation Name	CAPE (J kg^{-1}) T [T_V, WL]	BRN shear (m s^{-1})	BRN T [T_V, WL]	SRH (m^2 s^{-2}) 0-1km [0-3km]	Number of Generated "Storms" "Positive"	"Intermediate"	"Negative"
WK17_qtr_tail_r10_t20_3km	3407 [2338]	10.19	65.66 [45.06]	53.55 [116.3]		2	3
WK17_qtr_tail_r10_t40_1km	3407 [2338]	16.69	24.47 [16.79]	196.13 [252.39]		5	1
WK17_qtr_tail_r10_t40_3km	3407 [2338]	11.58	50.77 [34.84]	60.06 [135.32]	1	3	2
WK17_qtr_tail_r15_t10_1km	3407 [2338]	15.86	27.1 [18.6]	267.53 [281.67]		1	8
WK17_qtr_tail_r15_t10_3km	3407 [2338]	13.98	34.86 [23.93]	103.01 [209.98]		3	1
WK17_qtr_tail_r15_t20_1km	3407 [2338]	17.61	21.97 [15.08]	299.37 [327.36]		1	4
WK17_qtr_tail_r15_t20_3km	3407 [2338]	14.62	31.88 [21.88]	108.85 [226.04]		2	3
WK17_qtr_tail_r15_t40_1km	3407 [2338]	21.26	15.07 [10.35]	359.22 [421.04]	2	1	3
WK17_qtr_tail_r20_t10_1km	3407 [2338]	20.57	16.1 [11.05]	413.07 [428.16]		1	10
WK17_qtr_tail_r20_t10_3km	3407 [2338]	18.43	20.05 [13.76]	166.19 [325.62]		2	4
WK17_qtr_tail_r20_t20_1km	3407 [2338]	22.3	13.7 [9.4]	457.29 [485.79]		4	3
WK17_qtr_tail_r20_t20_3km	3407 [2338]	19.06	18.75 [12.87]	174.95 [348.46]		5	1
WK17_qtr_tail_r20_t40_1km	3407 [2338]	25.89	10.17 [6.98]	539.25 [599.75]	1	2	
WK17_qtr_tail_r25_t10_1km	3407 [2338]	25.29	10.65 [7.31]	605.43 [623.31]		3	2
WK17_qtr_tail_r25_t10_3km	3407 [2338]	22.89	13.01 [8.93]	239.96 [456.29]		4	2
WK17_qtr_tail_r25_t20_1km	3407 [2338]	27.01	9.34 [6.41]	661.5 [693.77]	1	3	4
WK17_qtr_tail_r25_t20_3km	3407 [2338]	19.06	18.75 [12.87]	174.95 [348.46]		5	1

APPENDIX B

CORRELATION CALCULATION

The following describes the method used to calculate correlations between vertical velocity (w) and a second meteorological quantity (ζ will be used for illustration purposes) within the storm extraction region's (SER_{5km}'s) 3-D gridded simulated environment. A correlation can be calculated for both updraft and downdraft, where updraft correlations are calculated using grid points having $w > 1$ m s^{-1} and downdraft correlations include grid points having $w < -1$ m s^{-1}.

The first step is to calculate the horizontal average of vertical velocity (updraft or downdraft) across each level in the vertical (i.e., every vertical level will have a single average value). The updraft correlation will be used to illustrate this process and therefore the average is calculated at every grid point in the horizontal with $w > 1$ m s^{-1}. The average value of w for w > 1 m s^{-1} over a specific vertical level (k) will be defined as $< w >_{(w>1),\,k}$ and can be calculated using the following:

$$< w >_{(w>1),\,k} = \frac{\sum_i \sum_j w_{(w>1),\,k\;ij}}{\left(x\#\text{gridpoints}_{(w>1),\,k} - 1\right)\left(y\#\text{gridpoints}_{(w>1),\,k} - 1\right)}, \qquad \text{(B.1)}$$

The next step is to calculate the horizontal average of the second meteorological quantity (ζ in this example) for the same grid points used to calculate $< w >_{(w>1),\,k}$ (e.g., grid points with w > 1 m s^{-1}). This term will be defined as:

$$< \zeta >_{(w>1),\,k} = \frac{\sum_i \sum_j \zeta_{(w>1),\,k\;ij}}{\left(x\#\text{gridpoints}_{(w>1),\,k} - 1\right)\left(y\#\text{gridpoints}_{(w>1),\,k} - 1\right)} \quad , \qquad \text{(B.2)}$$

Finally, the correlation is calculated between updraft and ζ for each level in the vertical using the following equation:

$$\rho(w,\zeta)_{(w>1),\,k} = \frac{\sum_{ij}\left(w_{ij} - < w >_{(w>1),\,k}\right)\left(\zeta_{ij} - < \zeta >_{(w>1),\,k}\right)}{\left[\sum_{ij}\left(w_{ij} - < w >_{(w>1),\,k}\right)^2\right]^{\frac{1}{2}} * \left[\sum_{ij}\left(\zeta_{ij} - < \zeta >_{(w>1),\,k}\right)^2\right]^{\frac{1}{2}}} \quad , \qquad \text{(B.3)}$$

in which the terms w_{ij} and ζ_{ij} are at the same grid points ($w > 1$ m s^{-1}) and level (k) as the average terms.

APPENDIX C

RADAR REFLECTIVITY CALCULATION

The following derivation is used to calculate radar reflectivity (ref) within the model.

$$ref = 10 * LOG_{10}(Ze) \qquad \text{(in dBZ)}$$

where, $Ze = Zer + Zes + Zeh$ (indicates contributions from rain, snow, and hail)

$$Zer = \left(\frac{k*720}{\pi^{1.75} * N_{0_{rain}}^{0.75} * rho_{rain}^{1.75}} \right) * (rho*qr)^{1.75}$$

$Zes = Zesnegf$ for "dry" snow (T < 0ºC)

or

$Zes = Zesposf$ for "wet" snow (T > 0ºC)

$$Zesnegf = \left(\frac{k * 720 * |k_{ice}|^2 * rho_{snow}^{0.25}}{\pi^{1.75} * |k_{water}|^2 * N_{0_{snow}}^{0.75} * rho_{ice}^2} \right) * (rho * qs)^{1.75}$$

$$Zesposf = \left(\frac{k * 720}{\pi^{1.75} * N_{0_{snow}}^{0.75} * rho_{snow}^{1.75}} \right) * (rho * qs)^{1.75}$$

$$Zeh = \left(\frac{k * 720}{\pi^{1.75} * N_{0_{hail}}^{0.75} * rho_{hail}^{1.75}} \right)^{0.95} * (rho * qh)^{1.6625}$$

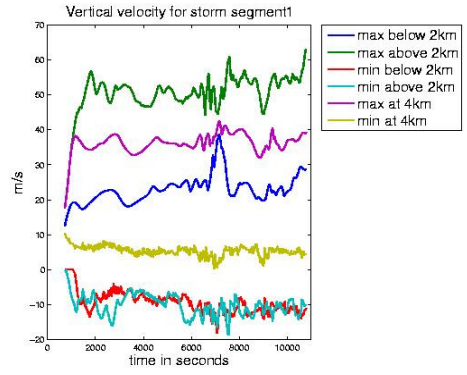
where,

rho	=	Air density (kg m^{-3})	
qr	=	Rain mixing ratio (kg kg^{-1})	
qs	=	Snow mixing ratio (kg kg^{-1})	
qh	=	Hail mixing ratio (kg kg^{-1})	
π	=	3.1415926	
k	=	1.0×10^{18}	=>Conversion factor from m^3 to mm^6 m^{-3}
rho_{ice}	=	917.0	=>Density of ice (kg m^{-3})
rho_{rain}	=	1000.0	=>Density of rain (kg m^{-3})
rho_{snow}	=	100.0	=>Density of snow (kg m^{-3})
rho_{hail}	=	913.0	=>Density of hail (kg m^{-3})
$N_{0\ rain}$	=	8.0×10^6	=>Intercept parameter in m^{-4} for rain.
$N_{0\ snow}$	=	3.0×10^6	=>Intercept parameter in m^{-4} for snow.
$N_{0\ hail}$	=	4.0×10^4	=>Intercept parameter in m^{-4} for hail.
$\lvert k_{water} \rvert^2$	=	0.93	=>Dielectric factor for water
$\lvert k_{ice} \rvert^2$	=	0.176	=>Dielectric factor for ice(if other than melted drop diameters are used)

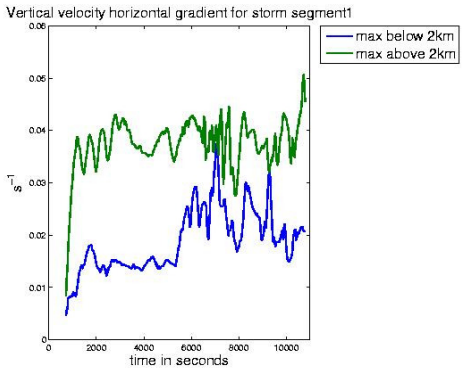
APPENDIX D

EXAMPLE FULL SET OF EXTRACTED QUANTITIES

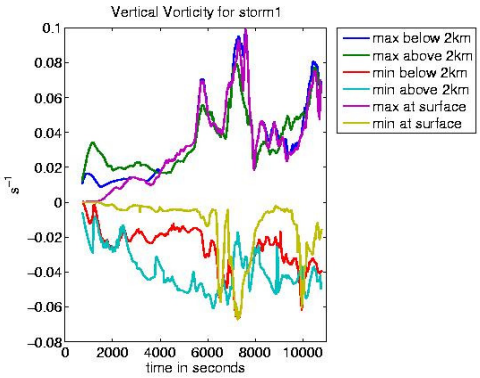
The following is the full set of extracted meteorological maximum and minimum quantities (refer to Table 3.2) for Storm 1 (blue) from the simulation experiment WK14_half_r15_4km (shown in Figure 4.1). The metadata extracted from different vertical levels for a specific meteorological quantity are plotted on a single figure with respect to simulation time. Quantity names are specified below each figure with extraction region and maximum or minimum identifier listed in legend.



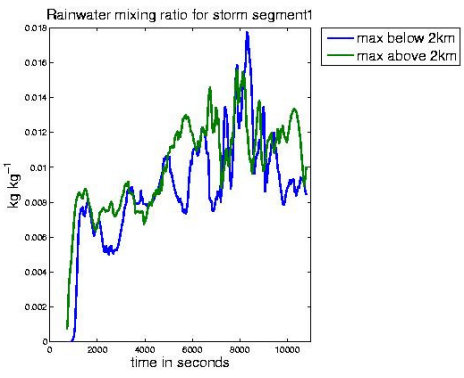
(*1*) Vertical Velocity



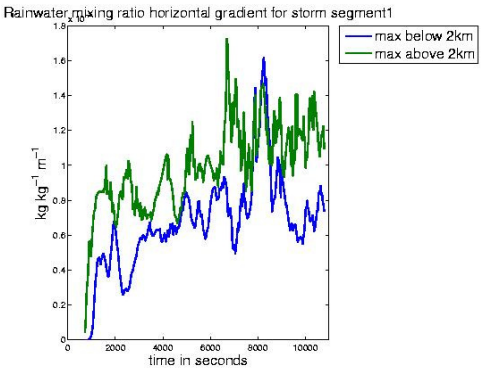
(*2*) Vertical Velocity Horizontal Gradient



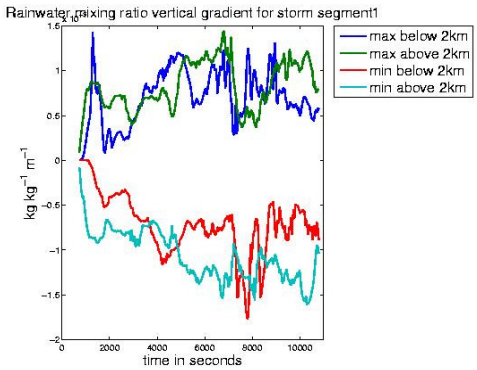
(*3*) Vertical Vorticity



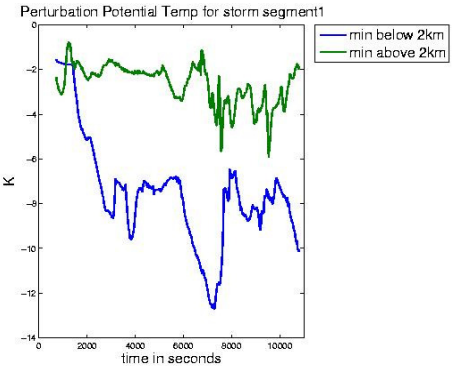
(*4*) Rainwater Mixing Ratio



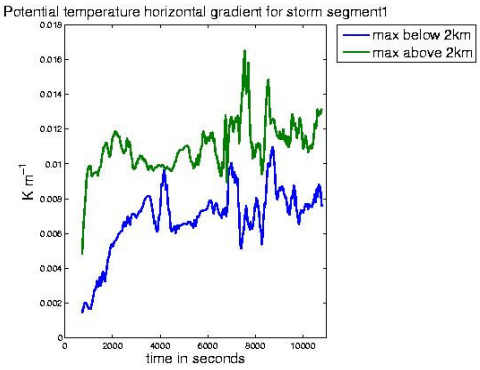
(*5*) Rainwater Mixing Ratio Horizontal Gradient



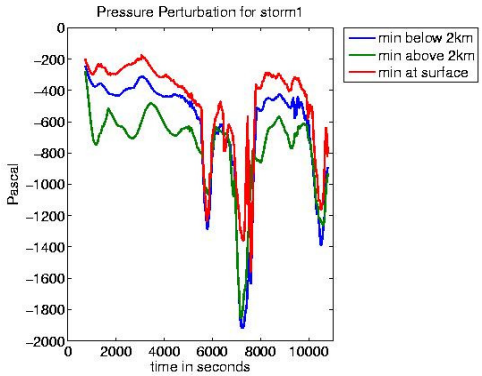
(*6*) Rainwater Mixing Ratio Vertical Gradient



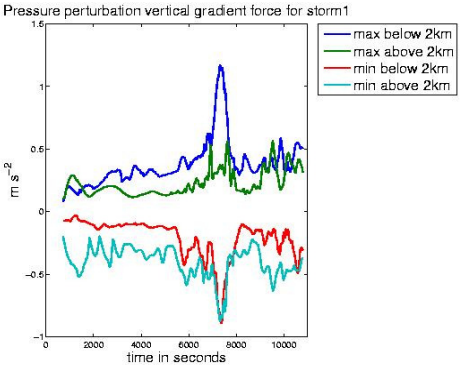
(*7*) Perturbation Potential Temperature



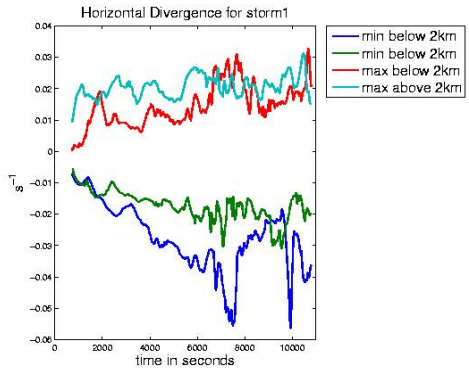
(*8*) Potential Temperature Horizontal Gradient



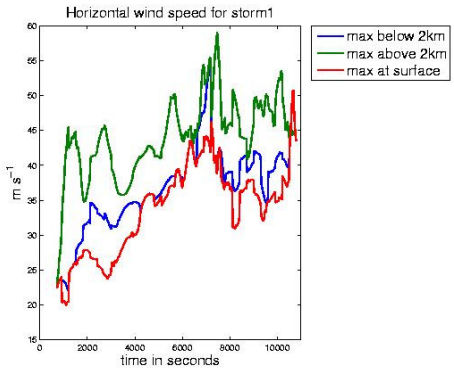
(*9*) Pressure Perturbation



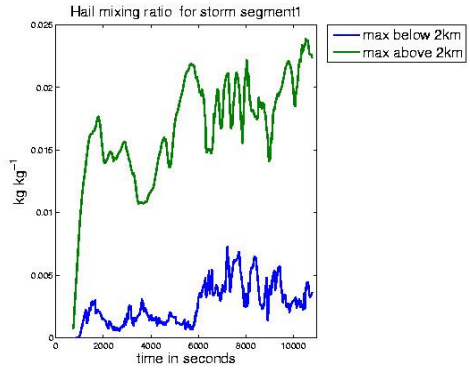
(*10*) Pressure Perturbation Vertical Gradient Force



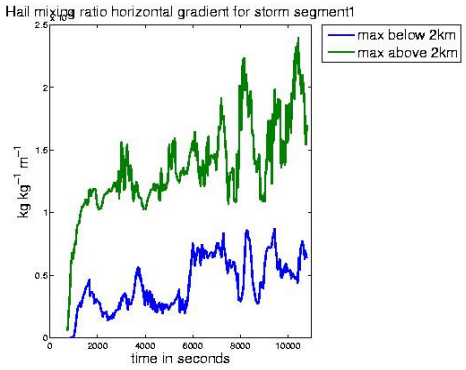
(*11*) Horizontal Divergence



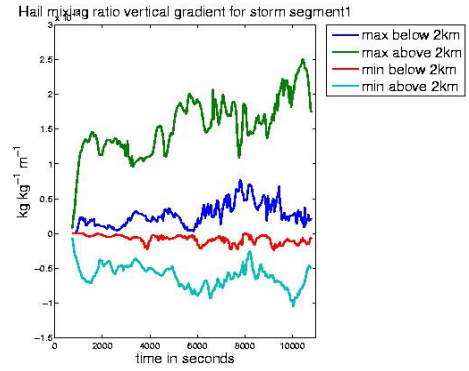
(*12*) Horizontal Wind Speed



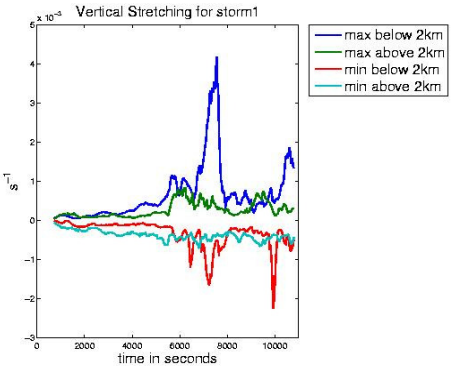
(*13*) Hail Mixing Ratio



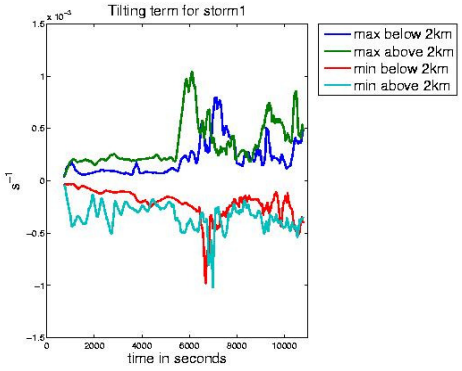
(*14*) Hail Mixing Ratio Horizontal Gradient



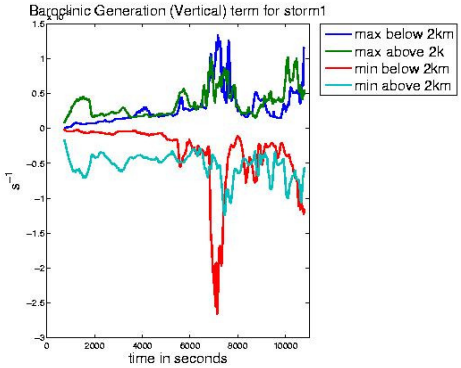
(*15*) Hail Mixing Ratio Vertical Gradient



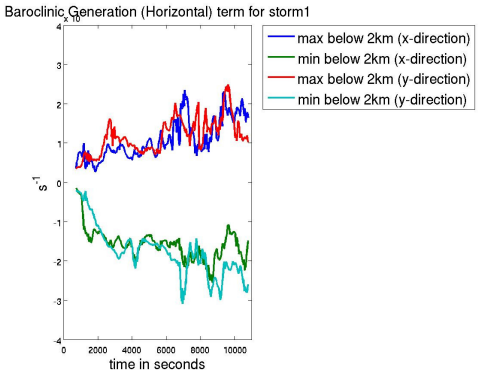
(*16*) Vertical Stretching Term



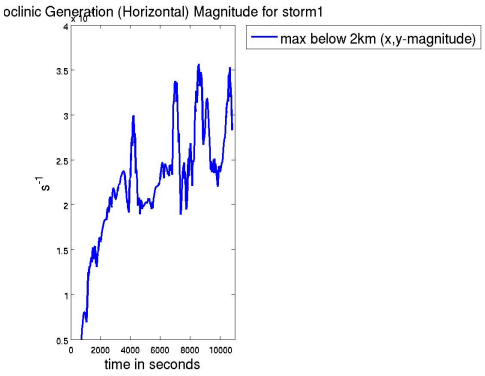
(*17*) Tilting Term



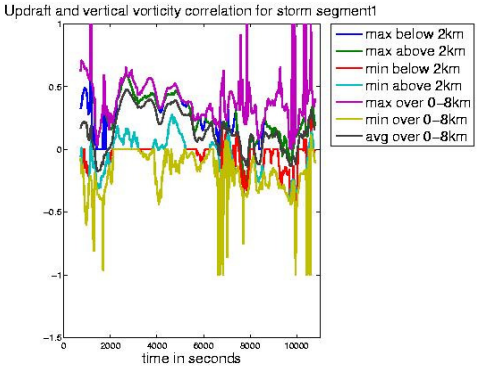
(*18*) Baroclinic Vorticity Generation (Vertical) Term



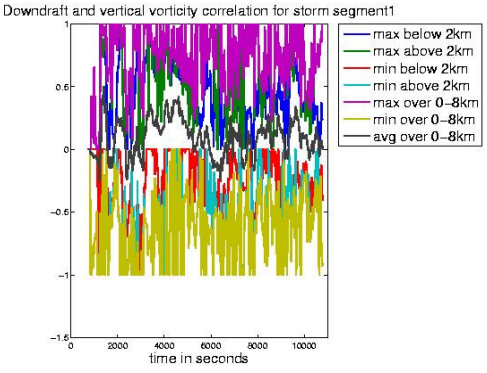
(*19*) Baroclinic Vorticity Generation (Horizontal) Term



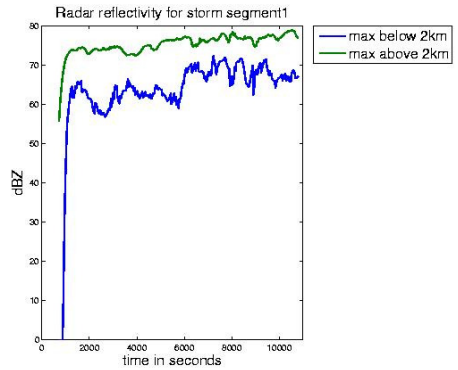
(*20*) Baroclinic Vorticity Generation (Horizontal x,y magnitude) Term



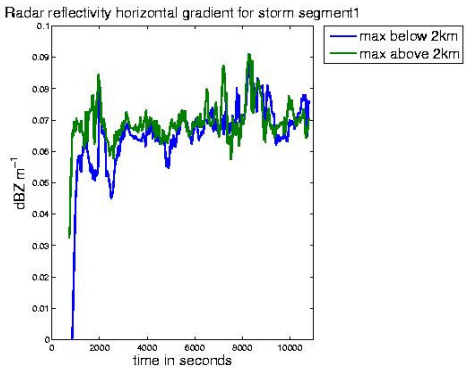
(*21*) Updraft and Vertical Vorticity Correlation



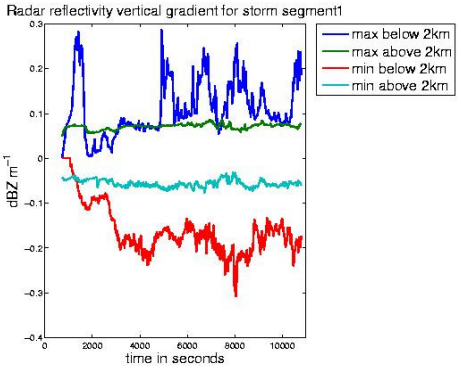
(*22*) Downdraft and Vertical Vorticity Correlation



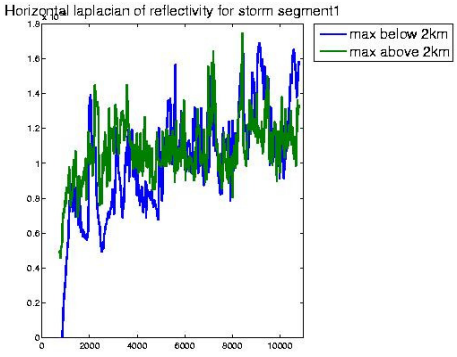
(*23*) Radar Reflectivity



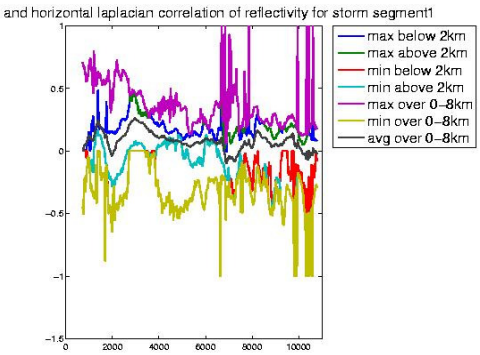
(*24*) Radar Reflectivity Horizontal Gradient



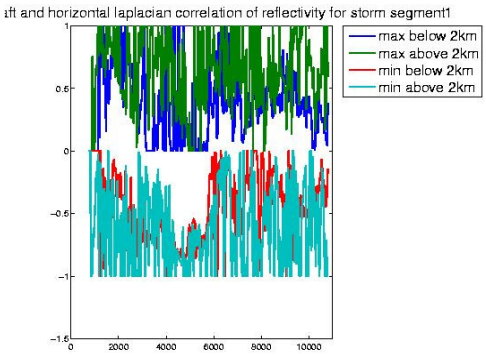
(*25*) Radar Reflectivity Vertical Gradient



(*26*) Horizontal Laplacian of Radar Reflectivity



(*27*) Updraft and Horizontal Laplacian of Radar Reflectivity Correlation



(*28*) Downdraft and Horizontal Laplacian of Radar Reflectivity Correlation