

**CELL IDENTIFICATION, VERIFICATION, AND CLASSIFICATION USING
SHAPE ANALYSIS TECHNIQUES**

A Dissertation Presented to the Faculty of the Graduate School
University of Missouri-Columbia

In Partial Fulfillment for the Degree
Doctor of Philosophy

by

STEVEN A. LACK

Dr. Neil I. Fox, Dissertation Supervisor

December 2007

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

CELL IDENTIFICATION, VERIFICATION, AND CLASSIFICATION USING
SHAPE ANALYSIS TECHNIQUES

presented by Steven A. Lack,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Assistant Professor Neil I. Fox

Associate Professor Patrick S. Market

Associate Professor Anthony R. Lupo

Professor Christopher K. Wikle

Associate Professor Athanasios Christou Micheas

Acknowledgements

I will begin by thanking research group which has allowed me to work on an exciting interdisciplinary project, Dr. Neil Fox, Dr. Chris Wikle, and Dr. Sakis Micheas. I am thankful for the opportunities I have had working with this group to gain outside professional contacts. I would also like to thank the rest of the faculty in the Atmospheric Science Program, including Dr. Tony Lupo, Dr. Pat Market, and Dr. Pat Guinan, for providing important insight to my work and providing interesting opportunities outside the realm of my current research, not to mention welcome breaks from academic endeavors.

Without the support of my family over the years, I would not have had the personal and professional success I enjoy. Special thanks also go to the rest of my family and friends, especially my fellow grad students, for all the encouragement and support I could ever hope for. I would especially like to thank Levi Blanchette for aiding in my research down the stretch and trips to the boat, Brian Pettegrew for welcome trips to the gym, and David Dale for excursions to the links. Additional thanks go to the OFW crew for a great five years of sports.

Last, but by no means least, I would like to thank Lindsay for her understanding and patience. I also thank her for introducing me to Dessie; without her we probably both would have lost our sanity. Through all the difficult times Lindsay has helped me see the light at the end of the tunnel.

This research was funded by a National Science Foundation grant, award number ATM-0434213.

Table of Contents

Acknowledgements	ii
Table of Contents.....	iii
List of Figures	v
List of Tables.....	xi
Chapter 1 Introduction.....	1
1.1 Outline of Dissertation	1
1.2 Objectives	2
Chapter 2 Verification Introduction.....	5
2.1 Review of Other Verification Methodologies	5
2.2 Review of Other Verification Methodologies	8
Chapter 3 Methodology	16
3.1 Original Procrustes Methodology	16
3.2 Adjustments to the Procrustes Methodology	22
Chapter 4 Results	25
4.1 Idealized Cases and Procrustes Scores	25
4.2 WDSS-II Nowcasts and Procrustes Results.....	38
4.2.1 Standard (Categorical) Skill Scores	43
4.2.2 Procrustes Verification	48
Chapter 5 Verification Conclusions	57
5.1 Procrustes Verification Conclusions and Future.....	57
Chapter 6 Cell Identification and Classification	66
6.1 Introduction to Cell Classification.....	66
6.2 Classification Background	67
Chapter 7 Cell Classification Methodology.....	73
7.1 WDSS-II Usage	73

7.2	Identification of Storm Cells.....	75
7.3	Expert Classification.....	76
7.4	Classification Tree Structure.....	79
Chapter 8 Cell Classification Cases.....		84
8.1	Sample Case Descriptions.....	84
Chapter 9 Cell Classification Results.....		91
9.1	Cell Identification.....	91
9.2	Classification Using Radar Only.....	96
9.3	Classification Using Radar and RUC-20 Data.....	102
9.4	Classification Using Uneven Populations.....	108
Chapter 10 Cell Classification Conclusions.....		114
10.1	Classification Conclusions.....	114
10.2	Classification Future Work.....	118
Appendix A- Using the Verification Tool.....		122
Appendix B- Using the Classification Tool.....		130
References.....		137
Vita Auctoris.....		141

List of Figures

Figure 3.1.1. WRF4NCAR precipitation forecast field (left) in hundredths of an inch. Cell identification as performed by the Procrustes scheme (right) at a 0.10" threshold and a minimum size of 10 km ²	21
Figure 3.1.2. Stage II precipitation analysis (left) in hundredths of an inch. Cell identification as performed by the Procrustes scheme (right) at a 0.10" threshold and a minimum size of 10 km ²	21
Figure 4.1.1. Idealized case 1 showing the differences in intensity while maintaining similar size and shape.....	26
Figure 4.1.2. Idealized case 2 showing the differences in intensity while maintaining similar size and shape without a normal distribution.	26
Figure 4.1.3. Idealized case 3 showing the differences in size while maintaining the same intensity.	27
Figure 4.1.4. Idealized case 4 showing the differences in rotating about the centroid 90° with similar size, shape, and intensity parameters.....	27
Figure 4.1.5. Idealized case 5 showing a significant but smaller translation error while holding the intensity constant.....	28
Figure 4.1.6. Idealized case 6 showing a large translation error while holding the intensity constant.	28
Figure 4.1.7. Idealized case 7 showing a significant translation error combined with 90° rotation while holding the intensity constant.	29
Figure 4.1.8. Idealized case 8 showing a slight translation error combined with 90° rotation while holding the intensity constant.	29
Figure 4.1.9. Idealized case 9 showing the differences in translation from observed object to forecast object while blowing up the forecast object so	

there is some slight overlap with the observed object. Intensity structure remains the same.....	29
Figure 4.1.10. Idealized case 10 having similar intensity structure with different shapes only.....	30
Figure 4.1.11. Idealized case 11 with areas of high intensity peaks in the forecast object rectangle with uniform intensity in the observed rectangle.	30
Figure 4.1.12. Idealized case 12 having circular object with a normal distribution with the forecast object being noisy.	31
Figure 4.1.13. Idealized case 13 with three observed objects being simply misplaced in the forecast object. There is no change in object size, shape, or intensity.....	31
Figure 4.1.14. Idealized case 14 with the objects being mirrored in intensity.....	31
Figure 4.2.1. The 1000Z composite (maximum) centered over KEWX. The pink dot is the radar location and the main cell of interest lies to the NE of the radar.....	40
Figure 4.2.2. Reflectivity from 0900Z (left) with clusters identified at scale 0 by numbers 1 and 2 with the resulting motion vectors and intensity for a 60-min nowcast with cluster identifiers 1 and 2 for the 30 to 60-dBZ range.....	41
Figure 4.2.3. Reflectivity from 0900Z (left) with clusters identified at scale 0 by numbers 1 and 2 with the resulting motion vectors and intensity for a 60-min nowcast with cluster identifiers 1 and 2 for the 40 to 70-dBZ range.....	41
Figure 4.2.4. The comparison between growth rates for the 40 to 70-dBZ range (left) and the 30 to 60-dBZ range (right) overlaid with the forecast wind field. Growing tendencies are shown with red colors while decaying tendencies are shown in green.....	42
Figure 4.2.5. Skill scores (POD, FAR, CSI) for the 20 to 40-dBZ range (top left), the 20 to 60-dBZ range (top right), the 30 to 50-dBZ range (middle left), the	

30 to 60-dBZ range (middle right), the 40 to 70-dBZ range (bottom left), and the mean ensemble of the above 5 ranges used (bottom right) using a 20-dBZ verification threshold. Forecast lead time (x-axis) is in minutes.....	44
Figure 4.2.6. Skill scores (POD, FAR, CSI) for the 20 to 40-dBZ range (top left), the 20 to 60-dBZ range (top right), the 30 to 50-dBZ range (middle left), the 30 to 60-dBZ range (middle right), the 40 to 70-dBZ range (bottom left), and the mean ensemble of the above 5 ranges used (bottom right) using a 30-dBZ verification threshold. Forecast lead time (x-axis) is in minutes.....	45
Figure 4.2.7. Skill scores (POD, FAR, CSI) for the 20 to 40-dBZ range (top left), the 20 to 60-dBZ range (top right), the 30 to 50-dBZ range (middle left), the 30 to 60-dBZ range (middle right), the 40 to 70-dBZ range (bottom left), and the mean ensemble of the above 5 ranges used (bottom right) using a 40-dBZ verification threshold. Forecast lead time (x-axis) is in minutes.....	46
Figure 4.2.8. The percent of area receiving at least 10 mmhr ⁻¹ for the 20 to 40-dBZ range (top left), the 20 to 60-dBZ range (top right), the 30 to 50-dBZ range (middle left), the 30 to 60-dBZ range (middle right), the 40 to 70-dBZ range (bottom left), and the mean ensemble of the above 5 ranges used (bottom right) using a 20-dBZ verification threshold. Forecast lead time (x-axis) is in minutes.....	49
Figure 4.2.9. The adjusted Procrustes penalty function over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km ² minimum cell size (left) and a 30-dBZ threshold and a 40 km ² minimum cell size (right). Forecast lead time (x-axis) is in minutes.....	50
Figure 4.2.10. The cell of interest from 13 January 2007 at 1000Z for the 30-dBZ threshold identification (top) and the 20-dBZ threshold identification (bottom).....	52

Figure 4.2.11. The adjusted Procrustes penalty function for the main cell over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes. 52

Figure 4.2.12. The Procrustes translation penalties for the main cell over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes..... 54

Figure 4.2.13. The Procrustes dilation penalties for the main cell over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes..... 54

Figure 4.2.14. The Procrustes shape penalties for the main cell over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes..... 55

Figure 5.1.1. The precipitation rate (left) and the cell identification (right) using a 30-dBZ threshold valid at 1000Z on 13 January 2007 converted to reflectivity from the RUC-20 for comparison against the K-means nowcast members. . 61

Figure 8.1.1. Base reflectivity image from 12 March 2006 from Kansas City, MO (KEAX) at 2200Z. The southernmost supercell is producing a tornado near Sedalia, MO at this time. 85

Figure 8.1.2. The Pendleton, OR supercell located south-southwest of the radar (KPDT) on 19 July 2004 at 2155Z. The 1.45 elevation angle was used due to significant beam blockage at the lowest scan. 86

Figure 8.1.3. The beginnings of the derecho as it approaches the Louisville, KY radar (KLVX) at 0000Z on 14 July 2004..... 87

Figure 8.1.4. “Wake” pulses that formed near the KLVX radar behind the derecho as it progressed southward through Tennessee and northern Alabama from 0500Z on 14 July 2004..... 87

Figure 8.1.5. The severe linear convective system with embedded rotation that spawned tornadoes near the Columbus, MS (KGWX) radar at 0800Z. 88

Figure 9.1.1. The identified cells from 12 March 2006 composite reflectivity centered over the KEAX radar site at 2200Z. Four distinct objects were identified in this particular case with each cell labeled a-d..... 92

Figure 9.1.2. The maximum composite reflectivity image from 12 March 2006 centered over the KEAX radar site at 2200Z. 92

Figure 9.1.3. The MUCAPE field from 12 March 2006 centered over the KEAX radar site at 2200Z. At this time, MUCAPE is weakening to values less than 3000 Jkg^{-1} 93

Figure 9.1.4. The MUVGP field from 12 March 2006 centered over the KEAX radar site at 2200Z. Notice values exceeding 0.3 ms^{-2} throughout the domain which signifies the continuing threat for rotating storms..... 94

Figure 9.1.5. The MUVGP field from 12 March 2006 centered over the KEAX radar site at 2200Z. Notice values exceeding 0.3 ms^{-2} throughout the domain which signifies the continuing threat for rotating storms..... 95

Figure 9.2.1. The cell classification tree using reflectivity only..... 97

Figure 9.2.2. The misclassification error (cost) for the classification tree using reflectivity-derived products only. The resubstitution error is an idealized cost where the cross-validation curve shows the actual cost for the different terminal nodes..... 98

Figure 9.3.1. Classification tree for the RUC-20 fields combined with reflectivity data calculated from the WDSS-II system. The tree is pruned at level 2 of 8 to capture the eight classification types at the end-nodes. 104

Figure 9.3.2. The misclassification error (cost) for the classification tree using an even population with reflectivity-derived products combined with RUC-20 NSE fields. The resubstitution error is an idealized cost where the cross-validation curve shows the actual cost for the different terminal nodes..... 105

Figure 9.4.1. Classification tree for the uneven populations of storm types using RUC-20 fields calculated from the WDSS-II NSE algorithm and reflectivity-derived fields. 111

Figure 10.2.1. The azimuthal shear product for 12 March 2006 at 2200Z centered over KEAX at the lowest elevation scan. The southern most supercell has a peak in azimuthal shear of 0.01241 s^{-1} 119

Figure 10.2.2. The divergence product for 12 March 2006 at 2200Z centered over KEAX at the lowest elevation scan. There are significant couplets of convergence and divergence near the main updraft in the supercells present. 119

List of Tables

Table 2.1.1. The subjective evaluation results for the all dates used in the NCAR ICP (2007) study combining results from all models compiled by Dave Ahijevych at NCAR.	7
Table 2.1.2. The subjective evaluation results for the three models run for all dates compiled by Dave Ahijevych at NCAR (NCAR ICP, 2007).	8
Table 3.1.1. Shape analysis statistics for the 3 models on 13 May 2005. The bold numbers represent which model is performing the best for the given category. In this situation WRF4NCAR is performing the best in total error, but not shape error (RSS).	22
Table 3.1.2. A further breakdown (error per forecast object as a sum of squares) of the shape analysis and intensity statistics for the 3 models on 13 May 2005. The bold numbers represent which model is performing the best for the given category. In this situation WRF4NCAR is performing the best in all categories.	22
Table 4.1.1. Procrustes shape analysis total penalty for the 14 idealized cases using the penalty function described by (3.1.4) without the rotation fix and using RSS (shape) minimization for the match. The cases are listed in ascending order based on their total penalty.....	32
Table 4.1.2. Procrustes shape analysis total penalty for the 14 idealized cases using the penalty function described by (4.1.1). The cases are listed in ascending order based on their total penalty.....	36
Table 5.1.1. POD, FAR, and CSI for the two verification thresholds for the RUC-20 forecast valid at 1000Z on 13 January 2007.	61

Table 5.1.2. The forecast quality score for the entire domain (Total Penalty) and the main cell in the image located to the NE of the radar in Figure 4.2.1.....	63
Table 7.3.1. Eight classification types with abbreviations and brief descriptions used within this classification tree study.	78
Table 7.4.1. The variables used within the classification tree. The first 10 variables are radar-derived products while the rest of the parameters are RUC-derived quantities from the near storm environment (NSE) algorithm in WDSS-II.....	83
Table 8.1.1. Dates, locations, and cell type of the training set used for classification tree construction. Notice that several cases of the same date and same location have different storm types. These are due to the nature of development and morphology of storms over time for the given domain...	90
Table 9.1.1. Selected output fields for the 12 March 2006 case at 2200 including the “expert” classification, for units and category descriptions see Table 7.4.1.....	95
Table 9.2.1. Estimated probabilities at the unique supercell end-node from the classification tree (Figure 9.2.1) using reflectivity-derived parameters only.	100
Table 9.2.2. Skill scores for exact matches for the classification tree using reflectivity-derived parameters only.....	100
Table 9.2.3. Skill scores for near matches for the classification tree using reflectivity-derived parameters only.....	101
Table 9.2.4. Contingency table for the expert classified observation cell types (x-axis) and the output from the reflectivity-derived products only decision tree. The table was created for 365 total cells. Refer to Table 7.3.1 for abbreviations.	102

Table 9.3.1. Estimated probabilities at the pruned supercell end-node from the classification tree using reflectivity-derived parameters combined with RUC-20 NSE fields.	105
Table 9.3.2. Skill scores for exact matches for the classification tree using and even population of reflectivity-derived parameters combined with near storm environmental data from the RUC-20 model.	106
Table 9.3.3. Skill scores for near matches for the classification tree using and even population of reflectivity-derived parameters combined with near storm environmental data from the RUC-20 model.	106
Table 9.3.4. Contingency table for the expert classified observation cell types (x-axis) and the output from the even population RUC-20 model fields combined with reflectivity-derived products decision tree. The table was created for 365 total cells. Refer to Table 7.3.1 for abbreviations.	108
Table 9.4.1. Skill scores for near matches for the classification tree using and an uneven population of reflectivity-derived parameters combined with near storm environmental data from the RUC-20 model.	110

Chapter 1 Introduction

1.1 Outline of Dissertation

The use of object-oriented approaches in both the verification of meteorological variables (especially precipitation) and the short-term forecasting of storms (nowcasting) has recently expanded. This two-part dissertation addresses the issue of verification and the possible nowcasting application of using shape analysis. Part one deals with a newly developed object-oriented verification tool using Procrustes shape analysis methodology. Part two deals with the potential to classify convective storm cells based on shape characteristics combined with radar-derived products and near-storm environmental (NSE) data.

Part one utilizes the Procrustes verification tool for an objective assessment of the skill of an ensemble of pseudo-reflectivity nowcasts created by the Warning Decision Support System—Integrated Information (WDSS-II) K-means algorithm as well as some simple geometric figures to test the effectiveness of the scheme for certain attributes. Varying the range of reflectivity within the K-means algorithm allows for the advection of storm clusters by different motion vectors. Using the Procrustes verification tool, the examination of the errors of matched truth and forecast objects for the different nowcasts can be assessed via a penalty function. This penalty function is based on errors in translation, rotation, dilation, shape, and intensity of a forecast object

and, as these penalties can be assessed separately, the mode of the error in the forecast can be determined. This illustrates the power and effectiveness in such a verification tool.

Part two examines the possibility of creating an automated cell classification tool using the Procrustes verification technique to identify objects. Obtaining radar-derived products and NSE data from the WDSS-II software allows for data to be easily overlaid for an identifiable cell in a domain. This data can then be used in a classification tree architecture using a training set of storm cells from different geographic locations in the United States. Different classification trees were created using combinations of radar-derived fields and NSE fields. The ability to use classification trees with products available in WDSS-II may result in the ability to add physical information to a nowcast system for growth, decay, and morphology as well as provide grounds for correctly matching storm cells in an object-oriented verification technique.

1.2 Objectives

The overall objectives of this research are to explore the possibilities of using shape analysis techniques to address ongoing verification efforts and possible nowcasting solutions. For the verification effort, this is accomplished by examining an ensemble of K-means forecast with both standard, categorical skill scores with results from the robust Procrustes verification tool. The advantages of the Procrustes methodology are also shown by using idealized cases represented by simple geometric objects with obvious error characteristics. It will be clear that the robustness of this scheme illustrated by the real nowcasts and the idealized cases can be used operationally to evaluate forecast objects of

any variable. For the nowcasting effort, the classification of convective storm types by using classification, or decision, trees is explored by using different ways to construct trees using different training sets. If a given classification tree is successful it may provide necessary automated feedback to a nowcasting model for growth, decay, and morphology. Furthermore, it will be shown that the novel approach of using a combination of WDSS-II data sources, including RUC-20 model output fields of the near-storm environment and radar-derived quantities, enhances the ability to classify convective storm type.

Overall, the research herein will address the following bullet points.

- The Procrustes verification tool provides insight to object-oriented forecasts that standard (categorical) skill scores are unable to provide. Categorical skill scores may show no skill when in fact the forecast is useful to the forecaster.
- The adjustments made to the original version of the Procrustes verification methodology allows the verification on a cell-by-cell basis and takes into account rotation about the major axis of the storm.
- Cell-by-cell verification has the potential to enhance warning decision support by allowing the focus to be on cells meeting a certain criteria.
- Varying the range of reflectivity over which the motion is estimated in the K-means nowcaster allows for the creation of a poor-man's ensemble which can be examined by the Procrustes verification tool in order to assess the quality of individual members of the ensemble.
- Individual error components on a cell-by-cell basis or an entire domain of cells may be displayed and weighted differently for a variety of end-users.
- Automated cell classifiers based on decision trees have the capability of being used to inform a nowcaster with information on storm growth, decay and morphology.

- The combination of reflectivity-derived products of near-storm environmental parameters enhances the decision tree process over the decision trees created with reflectivity-derived products alone.
- The development of an automated cell classifier is dependent on a large database of storm cells being created with consistent expert classifications for the dataset.
- Overall this research illustrates novel approaches to using the Warning Decision Support System—Integrated Information software in terms of the using multiple reflectivity ranges to produce a suite of K-means nowcasts and combining near storm environmental fields with reflectivity-derived parameters to classify storm type.

Chapter 2 Verification Introduction

2.1 Review of Other Verification Methodologies

Verification in meteorology has three major goals. One is to simply assess the accuracy of forecasts of different types. The second is to make a comparison between different methodologies of observing the same phenomenon. The third goal is to provide feedback to a model developer or an end-user for real-time modifications of a forecasting model or observing platform. As models and observations have become increasingly more complex, covering finer and finer resolutions, the need for advances in verification techniques have become equally important. Older methodologies, i.e. standard skill scores, may give false representation of a high resolution forecast's value. It is therefore necessary to develop tools that will match an end-user's view of forecast value.

Subjective evaluations of forecasts are of some value; however, experts differ on the quality of forecasts based on their field of expertise. For example, a hydrologist may be more focused on whether or not precipitation intensity was handled properly, whereas, an aviation forecast may be primarily concerned on the location of hazardous weather and subsequently the timing of the event is important. An experimental expert subjective evaluation was performed during the National Center for Atmospheric Research (NCAR) Verification Intercomparison Project (NCAR ICP, 2007). A panel of "experts" evaluated selected Weather Research and Forecasting (WRF) model one-hourly

precipitation forecasts from various dates during the Storm Prediction Center's (SPC) Spring 2006 Experiment, in which three versions of the WRF were run and compared to Stage 2 radar data for ground truth. The three runs included NCAR's 4-km WRF, the National Center for Environmental Prediction's (NCEP) 4-km WRF, and a Center for Analysis and Prediction of Storms (CAPS) 2-km WRF product. The experts had a couple of minutes for each comparison and ranked the forecast value on a scale from 0 to 5 with 0 being complete failure and 5 being a perfect forecast. The experiment was performed twice in an attempt to see if the experts changed their opinion after seeing the results a second time. The results from the subjective evaluation are summarized in Tables 2.1.1 and 2.1.2.

Tables 2.1.1 and 2.1.2 show that, although the model solutions for different days can be drastically different, the average score yields a uniformly, average (3) rating performance from the panel of experts. The skill range for the majority of forecasts on any given day from any given model ranged from total lack of skill (0) to a perfect forecast (5) demonstrating little consistency of the expert assessment of performance. This illustrates the need for an automated verification scheme so that a baseline comparison of models used on various days can be usefully compared. Although the subjective evaluation varied quite significantly, from the Tables above it should also be noted that the verification scheme should make an effort to yield a result consistent with expert evaluation criteria.

The development of automated verification schemes have branched in many ways, including: advancement to standard skill scores, fuzzy logic (neighborhood) methods, scale-decomposition verification, and object-oriented approaches. A brief review of developments in the verification of fine-scale meteorological fields is given following an international collaboration focused

	<i>26 Apr</i>	<i>13 May</i>	<i>14 May</i>	<i>18 May</i>	<i>19 May</i>	<i>25 May</i>	<i>1 Jun</i>	<i>3 Jun</i>	<i>4 Jun</i>
Mean	3.266	3.184	2.591	3.277	2.434	2.758	3.342	3.040	2.582
Standard Error	0.068	0.082	0.073	0.078	0.083	0.069	0.074	0.068	0.072
Median	3	3	3	3	2.5	3	3	3	3
Mode	3	4	3	3	3	2	3	3	3
Standard Deviation	0.78	0.94	0.84	0.89	0.96	0.79	0.85	0.78	0.83
Sample Variance	0.61	0.89	0.71	0.80	0.91	0.63	0.72	0.61	0.69
Kurtosis	1.2	-0.4	-0.6	-0.1	-0.7	-0.5	-0.3	-0.6	-0.4
Skewness	-0.7	-0.4	-0.1	-0.3	0.0	0.4	-0.3	0.0	0.1
Range	4	4	3	4	4	4	4	4	3.5
Minimum	1	1	1	1	1	1	1	1	1
Maximum	5	5	4	5	5	5	5	5	4.5
Sum	431.1	420.3	342.0	432.6	321.3	364.0	441.2	401.3	338.3
Count	132	132	132	132	132	132	132	132	131
Confidence Lvl(95%)	0.134	0.162	0.145	0.154	0.164	0.137	0.146	0.134	0.143

Table 2.1.1. The subjective evaluation results for the all dates used in the NCAR ICP (2007) study combining results from all models compiled by Dave Ahijevych at NCAR.

Mean Score		
Model	1 st Try	2 nd Try
WRF2CAPS	2.98	2.93
WRF4NCAR	3.01	3.05
WRF4NCEP	2.84	2.84
ALL	2.94	2.94

Table 2.1.2. The subjective evaluation results for the three models run for all dates compiled by Dave Ahijevych at NCAR (NCAR ICP, 2007).

on intercomparing various verification tools (NCAR ICP, 2007). The research herein will primarily be focused on the methodology and application of the Procrustes shape analysis technique developed by Micheas *et al.* (2007). This work will show how altering the original scheme can yield promising results in terms of object-oriented approaches.

2.2 Review of Other Verification Methodologies

The current section will serve as a brief overview of a few verification techniques employed in the international meteorological community following discussion from the NCAR ICP (2007). The discussion will focus on fine-scale model solutions of precipitation fields, as quantitative precipitation forecasts (QPF) remain the toughest problem to good short-term forecasting.

One of the most widely used verification tools in precipitation forecast evaluation, still used to present date, is the equitable threat score (ETS) (2.2.1) (Rogers *et al.*, 1996), or the similar critical success index (CSI). It is generally understood that the higher the ETS the better the forecast; an ETS

$$ETS = \frac{Hits - E}{Hits + False\ Alarms + Misses + E}, \quad (2.2.1)$$

$$where\ E = \frac{Forecast\ Pts * Observed\ Pts}{Total\ Pts}$$

of 1 would be a perfect forecast. The ETS is often given along with the bias (2.2.2) for the particular forecasting event. For the instance of precipitation forecasts, over-producing precipitation results in a bias greater than 1,

$$Bias = \frac{No.\ of\ Forecast\ Pts}{No.\ of\ Observed\ Pts} \quad (2.2.2)$$

whereas under-producing precipitation results in a bias less than 1, where a perfect forecast would have bias equal to 1. From the calculation of ETS it is easily shown that there is a problem dealing with bias. For ETS to be optimized and give a useful verification measure the bias must be equal to 1 (Mesinger, 2007). A new method devised by Mesinger (2007) for bias correction to the ETS (BCETS) involves making two assumptions: increasing hits per unit increase in bias is proportional to the area remaining to be hit and that the ratio of increase in hits to false alarms must be proportional to the same area. This method highlights the use of altering traditional skill scores for the current problem of QPF verification.

Fuzzy verification refers to methodologies to take into account the combination of uncertainty in observations with the unpredictability associated with different scales of the feature. This is accomplished by comparing a point forecast to an observation neighborhood of points or comparing an entire neighborhood of forecast points to a neighborhood of observed points. The neighborhood of points can be a neighborhood of time and space, but for simplicity space is often used. One example of this comes from Roberts and Lean (2007), which involves the selection of a neighborhood of forecast and observed points and comparing the fractional coverage of

each. The result is the fractional skill score (FSS) defined by (2.2.3), where P represents the percentage of coverage of the forecast (P_{fcst}) or observation (P_{obs})

$$FSS = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (P_{fcst} - P_{obs})^2}{\frac{1}{N} \sum_{i=1}^N (P_{fcst})^2 + \frac{1}{N} \sum_{i=1}^N (P_{obs})^2} \quad (2.2.3)$$

within the neighborhood. The FSS is simply a measure of the accuracy of a forecast simulated as a coarser resolution forecast as individual points within the domain may exceed the scale of predictability. One of the advantages to this methodology is that a comparison can be made from forecasts of different resolutions without prior adjustments to scale the forecast fields. A minor drawback to using the fuzzy verification approach is the possibility for information overload. Many metrics are used and displayed at once and, unless the end-user already has a favorite metric to use given a certain event, making sense of the output could become quite tedious. For a complete overview of the fuzzy verification framework the reader should refer to Ebert (2007).

Along similar lines with the fuzzy verification methodology, spatial decomposition methods may also be used to give useful results on intensity type forecasts. Basically, the spatial decomposition methods decompose images using wavelet or Fourier decomposition techniques and forecast success is judged at the different decomposed scales. Casati *et al.* (2004) used wavelet techniques to judge the intensity of storms in a neighborhood-to-neighborhood matching framework. This technique is highly useful in verifying the intensity of precipitation with the knowledge of the scales of predictability the event. In other words, it makes the distinction between areas of high-intensity, low-predictability fields and larger areas of lower-

intensity fields that should have higher predictability. The system is robust and useful in spatial forecast outputs; however, the data must be on a 2^n sized domain.

Object-oriented approaches, also referred to as feature-based approaches, are methods that attempt to capture how a forecaster perceives the matching of objects (such as precipitation fields) in a forecast field to the observed field. The major differences in the current methodologies that deal with objects include: what defines an object, how they match objects, and what information the method provides the end-user (NCAR ICP, 2007). Object-oriented approaches that are commonly used include the Ebert-McBride Technique (EMT: Ebert and McBride, 2000) and the Method for Object-based Diagnostic Evaluation (MODE: Davis *et al.*, 2006). Variations of the EMT exist as well as a new object-oriented approach developed by Baldwin and Lakshmivarahan (2003).

The EMT is based on contiguous rain areas (CRA) where a precipitation area is defined by a user-defined threshold of minimum precipitation amount. A user-defined minimum area criterion for the CRA can also be used to filter out insignificant precipitation fields. The main goal of this work was to provide a framework to discover systematic errors in QPF for mesoscale models. The primary metrics obtained from the original EMT scheme include displacement, intensity (volume), and shape (pattern) errors in a mean squared error (MSE) framework where the total mean squared error of a particular set of forecast (f_i) and observed (o_i) objects can be decomposed into the aforementioned components (2.2.4).

$$MSE_{tot} = MSE_{displacement} + MSE_{volume} + MSE_{pattern},$$

$$\text{where } MSE = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2 \quad (2.2.4)$$

The displacement is found by drawing a bounding box around the CRA finding the best observed match and maximizing the correlation coefficient or by minimizing the MSE. An interesting twist in the methodology of the EMT comes in the form of the CRA not being contained entirely within the verification domain. In the instance of a CRA not entirely contained within the verification domain due to radar coverage or model boundaries, CRAs of various sizes are produced by a Monte Carlo simulation. Error statistics are then calculated using the results of the simulation to achieve high confidence in the error statistics of the CRA that was not entirely observed. Grams *et al.* (2006) have since modified the original EMT for use during the International H₂O Project (IHOP) of 2002 to allow for use in Midwest convective systems. The overall strengths of the EMT come in the fact that displacement errors are readily displayed in the output. It is also beneficial that error statistics are calculated before and after the displacement of the objects. The technique also readily displays information of complete misses and false alarms. The main weakness of the system is that there are too many user-defined (tunable) parameters which can drastically alter the end results. The forecasted and observed objects must also be contiguous or very close in proximity or matching may break down.

MODE (Davis *et al.*, 2006) can evaluate any model field that can be viewed as objects. It is similar to the EMT with differences in identifying and matching objects for the computation of verification metrics. The method focuses on large areas of precipitation areas; isolating stronger convection within mesoscale bands is not of primary interest.

The identification of cells in MODE is a three-step process. The first step includes smoothing the field using cylinders to fit the common shapes of rain features only keeping the mean rainfall amounts over the identified,

smoothed object. The second step applies a threshold to remove scales and intensities below a given rainfall amount. This step is the masking phase and produces smoothed objects within the domain. The final step simply overlays the original precipitation amounts over the masked objects from the previous step. The overall result yields objects in the domain of interest with complete intensity information.

After the objects are identified, properties for each are determined for later statistical analysis. These object properties include intensity, area, centroid, curvature, axis alignment, and aspect ratio. Once objects and their properties are available matching truth to forecast can begin. There are two strategies for matching in MODE. The simplest involves minimizing the distance between the observed object's centroid and the forecast object's centroid; however, the distance between the two centroids must be smaller than the sum of their sizes. The second, lesser used, strategy involves matching the object based on size and centroid distance so that two objects of similar size may be matched without having to overlap. Clearly, the matching methodology in MODE is more effective than the EMT, in that the CRAs in the EMT must be overlapping or adjacent.

The output for MODE includes basic skill scores of hits and misses for each of the identified objects. Therefore, the amount of overlap between the objects is still of importance in the evaluation of forecast performance. If more objects are identified in the forecast than available in the truth image, the system identifies these objects as false alarms. The system will then provide select statistics including intensity, orientation, and size differences as well as displacement errors for a variety of end-users.

The Baldwin approach (Baldwin and Lakshmivarahan, 2003) utilizes similar techniques for identifying objects with a slightly different twist. The

first step in identifying an object is to pick out all contiguous rain fields above a certain threshold. The areas identified are then blown up by 15-20% of their original size and if objects are within 20 km they are considered one merged object. Similar attribute vectors are created for each object as was done in the MODE approach. Statistics for each object and its different matches are done by comparing the attribute vectors using Euclidian distance in Cartesian (x,y) space, d (2.2.5).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.2.5)$$

The result of using the Euclidean distance with the vectors of attributes from identified objects causes multiple matching of truth objects and forecast objects. The main result is a number representative of the distance between two matched objects with the smaller the distance between the attribute vectors the better the forecast.

Another recent advance in object-oriented forecasting comes from Venugopal *et al.* (2005) and is referred to as the Forecast Quality Index (FQI). The FQI is similar to other object-oriented approaches; however, it uses the partial Hausdorff distance (PHD) to match the appropriate cells. "Surrogates" (ensemble forecasts based off of the observations) are created for each truth image that represents a range of possible forecast outcomes. The surrogate mean and the given forecast are then compared to the truth image to generate the FQI.

Overall, object-oriented approaches are more intuitive than other procedures when dealing with QPF in fine-scale models. They provide numerous metrics which may give the end-user a variety of useful parameters to judge specific attributes of given forecasts. In the case of the

EMT, the CRAs must be oriented in a way that the truth and forecast must overlap or be approximately adjacent. This does not bode well for false alarms and displacement information. MODE attempts to correct for this by using a distance weighted by size minimization to match truth and forecast precipitation entities. MODE also provides numerous other metrics besides hits, misses, and false alarms to quantify errors in orientation, displacement, and size. The Baldwin method is more complicated than the previous method but still uses the same basic principles to identify cells. The Baldwin method then compares vectors of attributes from the truth objects to the forecast object and calculates a minimum Euclidean distance to assess the skill of the forecast. The most complex method is the FQI which uses a PHD to match the cells and then assesses forecast skill based on the creation of surrogates that represent the range of possible forecasts for the domain. All methods apply a threshold to get rid of weak or small precipitation objects, which makes each method user-tunable. However, most of the metrics that result cannot be easily altered for the benefit of the end-user. There exists the possibility for counterintuitive matching to occur in all methods, but in general they all produce useful information on the performance of the forecast being evaluated. Finally, the methods all seem to give an overall assessment for all objects in the domain and may not breakdown the errors of the individual objects themselves without some modifications; thus, the methods described above may not be as robust as they could be.

Chapter 3 Methodology

3.1 Original Procrustes Methodology

The method used throughout the bulk of this research stems from yet another object-oriented verification technique using Procrustes shape analysis. The original framework of the scheme features a reduced dimensionality so that many ensemble members of a forecast may be handled at once and in a near real-time setting. The Procrustes method has strong roots in shape analysis and image processing (Dryden and Mardia, 1998) and has since been modified by Micheas *et al.* (2007) for use in QPF assessment. In particular, Micheas *et al.* (2007) produced a method which can evaluate short-term forecast (nowcast) ensembles in a robust and efficient manner. In the following paragraphs cell identification, matching, and verification metrics of the scheme will be presented in their original form.

Cells are identified using gridded ASCII data of rainfall amounts or reflectivity as input. A user-tunable threshold is then applied in terms of the selected variable to isolate the intensities of interest. A minimum size criterion is also selected to eliminate areas of intensity that are too small to be considered verifiable objects. Once the original image (truth or forecast) is filtered, adjacent areas of contiguous intensity are considered as individual objects allowing for gaps of one-pixel for smoothing purposes. For each identified object a centroid is found and minimum, mean, and maximum

intensities within the cell are saved for future use in the verification process. Starting with the centroid and using a tunable, user-specified fixed angular separation (A ; where $A \ll 360^\circ$), polar coordinates from the centroid to the edge of the cell are calculated. For example, using an angular separation of 10° ($A=10$) would give 36 points along the edge of the identified cell with corresponding range and azimuth. These points provide the landmarks on which the verification scheme is used to match cells and calculate verification quantities. Each cell in the domain is, consequently, identified as a vector of values consisting of maximum, mean, and minimum intensity, centroid location, and polar coordinates to the edge of the identified cells based on the fixed angle selected. This serves as a technique to reduce the dimensionality of the verification problem for ensemble forecasts.

There are two distinct methods for matching truth cells to forecast cells that the original Procrustes methodology for QPF verification outlined in Micheas *et al.* (2007) utilizes. The first is by finding a minimum distance between the centroids of the identified cells. The second is by matching objects of similar shape (Procrustes matching), in which case the shape errors between identified cells are minimized. If there are fewer forecast cells than truth cells, the same forecast cell may be matched to a number of truth cells by the criteria outlined above. The cells do not need to be adjacent or overlapping for matching. The drawback to both matching methodologies is that it does not ensure what a forecaster would deem a “proper” match; some counterintuitive matching can occur which is inherent to all object-oriented approaches. For example, when matching cells based on minimizing the residual error (shape), there is a possibility that improper matching would occur based on natural growth and decay at longer lead times. It is possible that future versions of the scheme may utilize a cost minimization technique

to combine distance and shape matching. In summary, there is some chance that the verification scheme would produce counterintuitive matches, as is the case with most object-oriented approaches.

Upon matching the observed and forecasted objects, penalties in dilation, rotation, and displacement must be assessed. These statistics are derived using the Procrustes methodology described in (Micheas *et al.*, 2007). The forecast object is essentially overlaid onto its corresponding observed field and a fit is performed using equation (3.1.1).

$$\hat{z}_j = \hat{c}_{jk} + \hat{r}_{jk} e^{i\hat{\phi}_{jk}} z_{kj} \quad (3.1.1)$$

Equation 3.1.1 is known as the Full Procrustes fit, the superimposition of z_{kj} (forecast) onto z_j (observation). Where the first component, c , is the translation term, r is the dilation term and ϕ is the rotational component. The estimators are shown in 3.1.2, where z_c represents the centered cell; $\overline{z^{kj}} = z^{kj} - z_c^{kj}$; $\overline{z^j} = z^j - z_c^j$; and z^*z represents the complex conjugate.

$$\begin{aligned} \hat{r}_{jk} &= \frac{|(z_c^{kj})^* z_c^j|}{(z_c^{kj})^* z_c^{kj}}, \\ \hat{\phi}_{jk} &= \arg((z_c^{kj})^* z_c^j), \\ \hat{c}_{jk} &= \overline{z^j} - \hat{r}_{jk} e^{i\hat{\phi}_{jk}} \overline{z^{kj}}, \end{aligned} \quad (3.1.2)$$

The residual sum of squares (RSS) between the observed cell and the forecasted cell becomes 3.1.3.

$$RSS_{jk} = (z^j - \hat{z}_k^j)^* (z^j - \hat{z}_k^j) \quad (3.1.3)$$

If the forecast is perfect, then $r=1$, $c=0$, and $\phi=0$ in 3.1.1. The terms in 3.1.1 are incorporated into the RSS term in the penalty function, D (3.1.4).

$$D = RSS_k + SS_{avg}^k + SS_{min}^k + SS_{max}^k \quad (3.1.4)$$

The other components in the penalty function are the errors based on intensity differences between the forecast and observed object summed for the entire domain, including average intensity (SS_{avg}), minimum intensity (SS_{min}), and maximum intensity (SS_{max}). It is easily shown that if the number of objects in the forecast field is less than that in the observed field, the penalty is increased due to the matching of multiple forecast objects to one truth object. However, a notable problem with this penalty function occurs when there are more forecast objects than truth objects. Any additional penalty is not assessed for false alarms in this case. The penalty function, in this framework, is only meaningful when multiple forecast products (different models at similar resolutions) are compared to the truth field. The lower the penalty, the better the forecast solution is. The penalty function may also be weighted accordingly to the user's specific needs. For example, a hydrologist using the data may value intensity errors highly and weight the penalty accordingly.

An example for the shape analysis verification scheme is shown for a sample case from the NCAR Spring 2005 project for an intercomparison of WRF forecast solutions (NCAR ICP, 2007). Three versions of the WRF were compared from 13 May 2005 and compared against a Stage 2 radar analysis. Figure 3.1.1 is an example of one of the three forecast products used; in this case the WRF4NCAR. Figure 3.1.2 is the Stage 2 radar analysis. For this particular date and example, the NCAR version of the WRF model outperformed the others (WRF2CAPS and WRF4NCEP). Although the basic skill scores for the date in question favored the NCAR version of the WRF, the skill score was not acceptable (CSI ~0.25). The shape analysis scheme for comparison between the models allows some insight into the differences (Tables 3.1.1 and 3.1.2) even when using the original penalty function (3.1.2).

Although the CSI was approximately the same for all cases for this date, it can be seen in the Table that using the Procrustes shape analysis verification can give more insight as to which model is performing the best using several criteria. However, when examining the Tables above the magnitudes of the values are not easy to understand. If only one WRF version were compared against the observed Stage 2 data using the Procrustes methodology, one could not immediately interpret the results. This methodology is primarily useful for a comparison of different models or ensemble members. Also of note is the fact that the total error in Table 3.1.1 does not include any information on the intensity error, and that there is the potential for bias based on number of cells detected in the identification process. Of final note is the absence of a rotation penalty for all cases shown and, the fact that the penalty is for the entire domain and is not broken down by identified objects. These problems will be addressed in the following section.

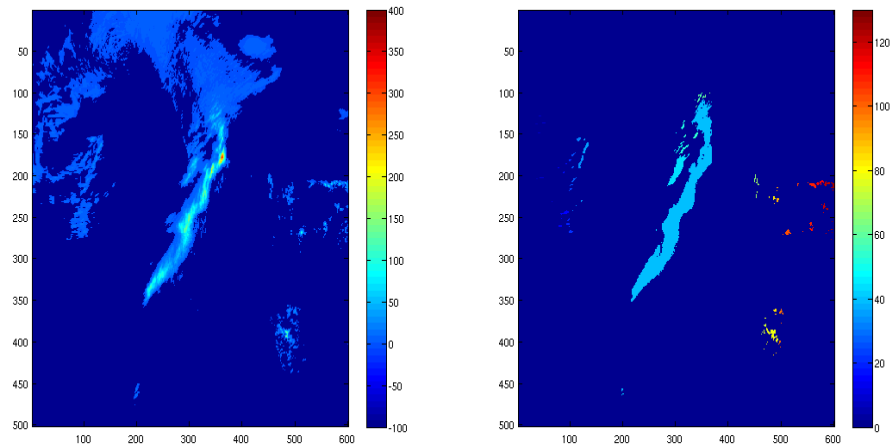


Figure 3.1.1. WRF4NCAR precipitation forecast field (left) in hundredths of an inch. Cell identification as performed by the Procrustes scheme (right) at a 0.10" threshold and a minimum size of 10 km².

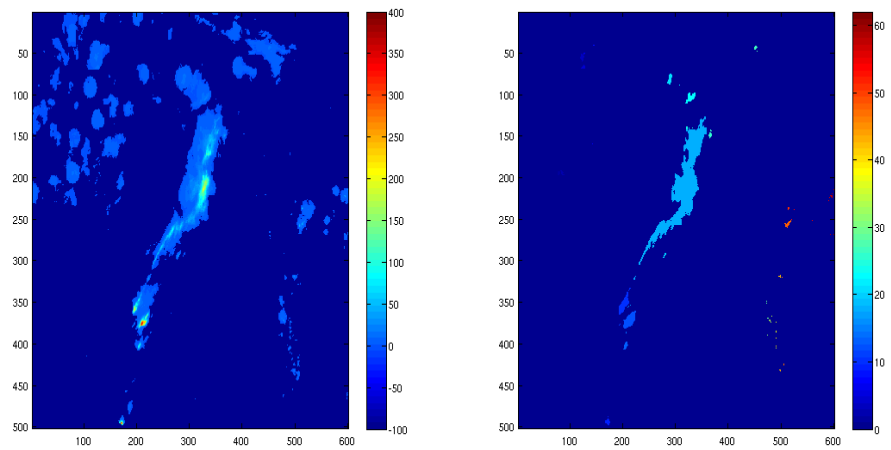


Figure 3.1.2. Stage II precipitation analysis (left) in hundredths of an inch. Cell identification as performed by the Procrustes scheme (right) at a 0.10" threshold and a minimum size of 10 km².

	WRF4NCAR	WRF4NCEP	WRF2CAPS
RSS	5915.1	5448.7	4860.3
SStot	1753.1	2308.3	1900.5
No. of Cells	26	61	39
Tot Error	266070	578580	284160

Table 3.1.1. Shape analysis statistics for the 3 models on 13 May 2005. The bold numbers represent which model is performing the best for the given category. In this situation WRF4NCAR is performing the best in total error, but not shape error (RSS).

	WRF4NCAR	WRF4NCEP	WRF2CAPS
Min Intensity	2	4	4
Mean Intensity	8926	26825	9079.9
Max Intensity	26	61	39
Dilation	4.6711	22.669	11.596
Rotation	0	0	0
Translation	428290	7736700	1176000

Table 3.1.2. A further breakdown (error per forecast object as a sum of squares) of the shape analysis and intensity statistics for the 3 models on 13 May 2005. The bold numbers represent which model is performing the best for the given category. In this situation WRF4NCAR is performing the best in all categories.

3.2 Adjustments to the Procrustes Methodology

The original framework of the Procrustes methodology allows for the comparison of different error components between different ensemble members (or models) which can be tailored to a specific end-user. However, problems arise in the interpretation with the comparison and the fact that

there seem to be no rotation errors in any of the sample forecast/observation pairs (Tables 3.1.1 and 3.1.2). In order to test these problems several idealized cases were used with known errors in the forecast/observation pairing, these will be summarized in the next section. The problem that gave the most puzzling result was the missing rotation error component. This was determined to be due to the improper matching of the forecast cell landmarks to the truth cell landmarks found along the cell boundary and by the fact the original Procrustes methodology defines rotation about the origin. When calculating the coordinates a fixed angle (A) is used, starting with 0° , and continuing a full 360° around the identified cell. Once two cells are matched using the minimum distance between centroids the cells may not be oriented the same; what may be deemed as the north edge of the forecasted cell will be matched to the north edge of the observed cell even though there may be significant rotation. The original framework of the scheme would interpret this error as a translation since the north edge of the cells were not in the same location as would be the case with the rest of the landmarks. Therefore, two modifications to the original Procrustes framework are made.

The first modification is to make sure the coordinates of the truth cell match to the forecast cell in the absence of rotation. This is accomplished by finding the major axis of each identified forecast cell in the domain with its corresponding observed cell match and aligning them along the y-axis while removing translation errors by overlaying each matched pair on the origin. Again, in the Procrustes scheme rotation is only valid around the origin. The angle through which each is rotated to align the major axis along the y-axis is recorded and the difference is taken to determine the actual rotation which must be less than or equal to 90° . As the observed/forecast pair is now aligned along a common axis the coordinates are found using the fixed-angle

methodology. They are then transformed back into their original space using the difference in the angles found during the previous step and by translating them back to their original centroid location; this ensures that the coordinates are matched between the forecast and observed object for the Procrustes analysis to be properly performed.

The second adjustment is the breakdown of error components by each matched forecast/observation pair as some cells may be of more importance than others for different end-users. This adjustment actually falls out of the previous correction as the rotation fix must be done on each pair of matched objects. A new total penalty for the entire domain must also be calculated which attempts to eliminate the bias of the number of cells identified. This also may give better insight of the usefulness of the forecast for the entire domain. A comparison of the original total penalty to the adjusted total penalty will be made in the following section when idealized cases are examined.

Chapter 4 Results

4.1 Idealized Cases and Procrustes Scores

Idealized cases were generated to test with the newly developed Procrustes shape analysis scheme. These cases run the gamut of the general types of forecasting or modeling flaws. Since this is a relatively new verification scheme it makes sense to test it on idealized cases, as well as real cases, to get an idea of its usefulness and potential areas for improvement. The idealized cases used in this study range from cases in which there could possibly be a timing error in the model yielding large translation errors to cases that may have no translation error but have significant intensity errors. Most cases used involve simple ellipsoidal objects with some given intensity. There were 14 idealized cases used in this study. The resulting total penalty for all idealized cases is summarized in Table 4.1.1 and 4.1.2 towards the end of this section. Total penalty is used in this instance since the original Procrustes methodology for QPF forecasts generates a total penalty based on all cells in the domain and does not handle rotation correctly which alters other individual error components. The adjustments made to the Procrustes scheme are done on a cell by cell basis and total error is summed over all cells as the final step.

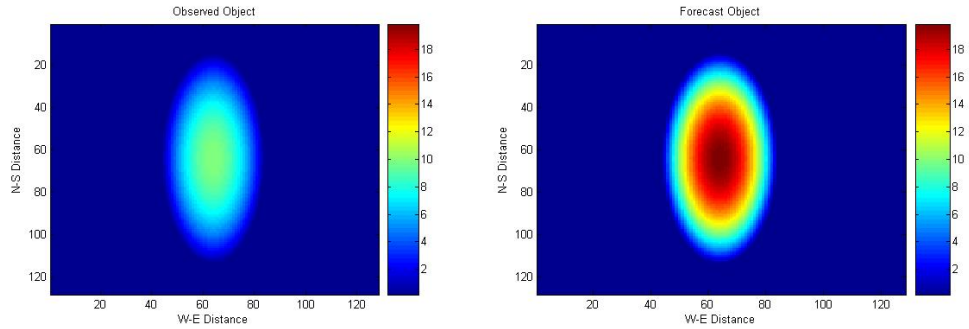


Figure 4.1.1. Idealized case 1 showing the differences in intensity while maintaining similar size and shape.

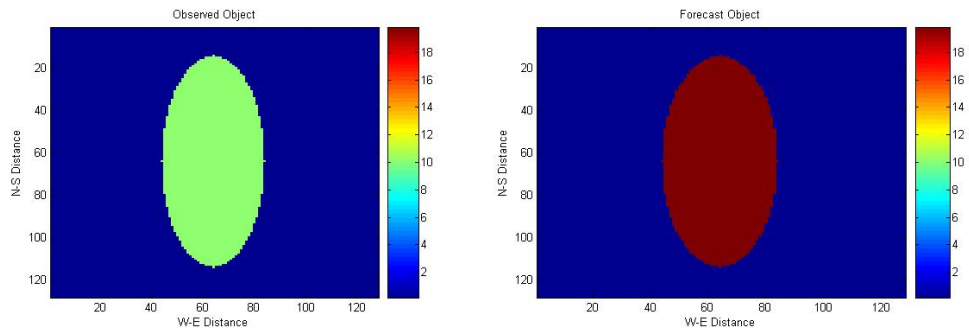


Figure 4.1.2. Idealized case 2 showing the differences in intensity while maintaining similar size and shape without a normal distribution.

The first case (case 1) examined dealt with a simplified ellipse with no translation (Figure 4.1.1). The observed object has a weak normalized intensity about its center, and the forecast object has a higher intensity which is normally distributed within the cell with its peak intensity in the middle. This case yields zero penalties for the dilation, rotation, and translation components with a near zero value for the RSS term of the penalty function, but a large error in terms of max intensity and a slight error in terms of mean intensity. Minimum intensity error is negligible in all idealized cases, but could be significant if there are peaks and valleys in the intensity field. This idealized case closely resembles case 2, in which there is no normalization of

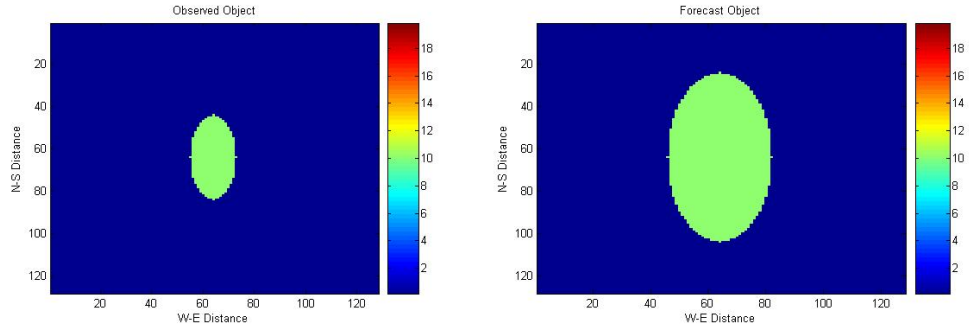


Figure 4.1.3. Idealized case 3 showing the differences in size while maintaining the same intensity.

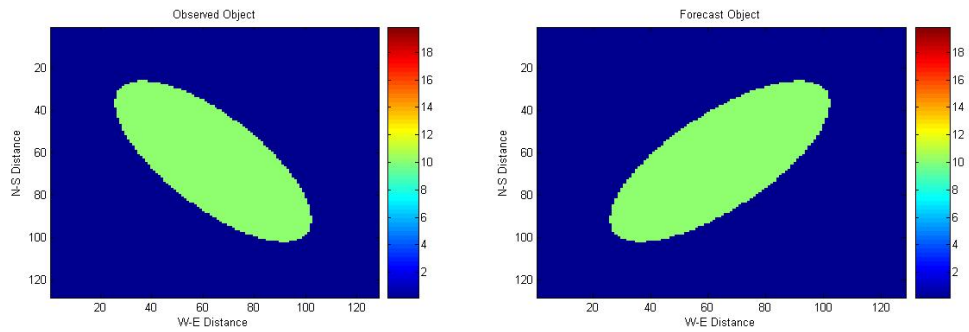


Figure 4.1.4. Idealized case 4 showing the differences in rotating about the centroid 90° with similar size, shape, and intensity parameters.

intensity over the ellipse (Figure 4.1.2). It simply has a higher uniform intensity. This case results in a large overall error due to both the mean and the max intensity producing a high penalty. Case 3 (Figure 4.1.3) simply involves two objects of similar reflectivity; however, the forecast object is simply double the size of the original object. Both cells share the same centroid therefore there is no translation error component. The only error existing is the dilation component from the Procrustes fit.

Case 4 is a unique case which deals with the forecast ellipse being rotated about its centroid 90° (Figure 4.1.4). There is no change in intensity, size, or shape of the ellipse. This case ends up yielding the highest error despite having uniform size, shape, and intensity structure. The prior version

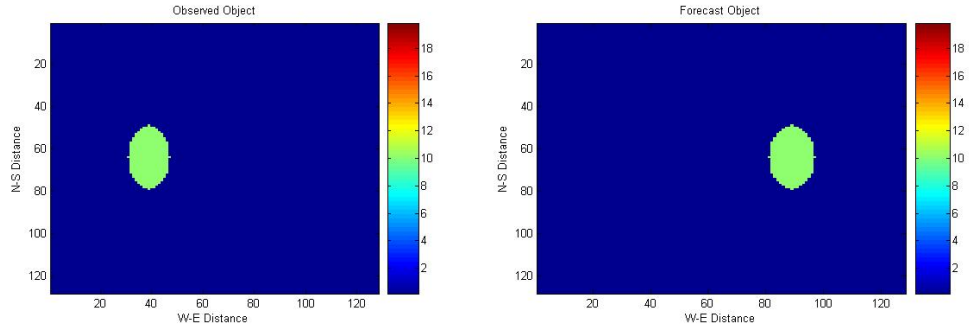


Figure 4.1.5. Idealized case 5 showing a significant but smaller translation error while holding the intensity constant.

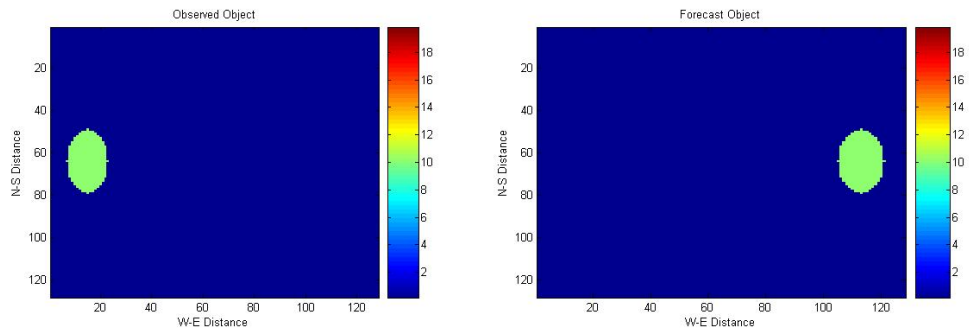


Figure 4.1.6. Idealized case 6 showing a large translation error while holding the intensity constant.

of the scheme had problems with rotation so that the overall fit was poor giving the high penalty value. This problem has since been corrected so that the values in Table 4.1.2 reveal more meaningful error characteristics.

Case 5 through case 9 represent translation error with a different twist each time. Case 5 (Figure 4.1.5) represents a small translation with ellipsoids of similar size, shape and intensity. Case 6 (Figure 4.1.6) simply increases the distance between the ellipsoids. Case 5 and 6 have the lowest combined penalty of all cases as the only thing that is wrong is the location (timing) of the system. Case 7 (Figure 4.1.7) deals with a translation error with objects of different sizes, while case 8 (Figure 4.1.8) deals with a translation and rotation component, and case 9 deals with a centroid translation with some overlap in

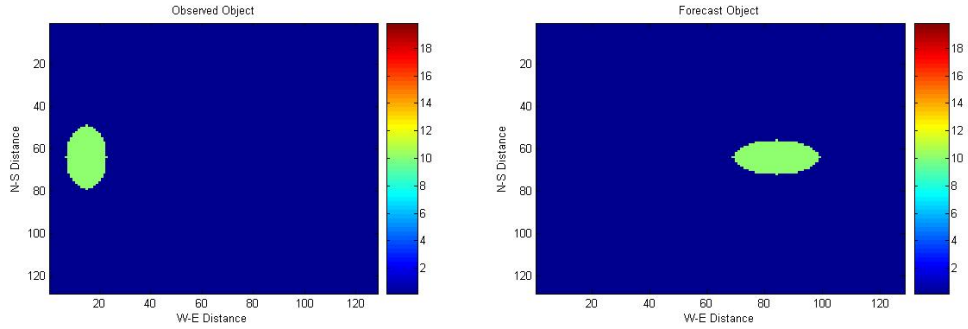


Figure 4.1.7. Idealized case 7 showing a significant translation error combined with 90° rotation while holding the intensity constant.

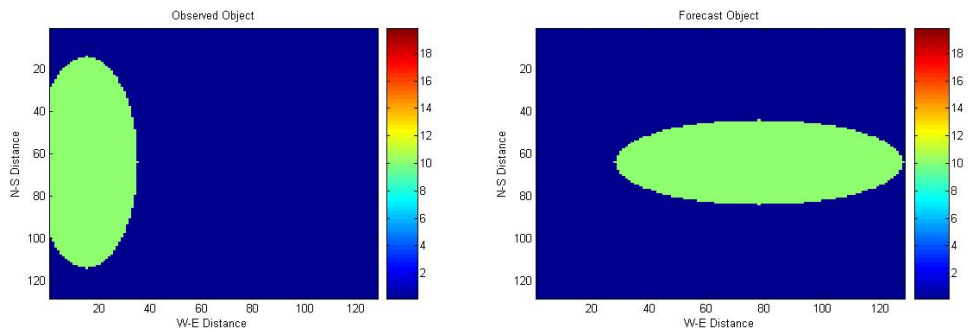


Figure 4.1.8. Idealized case 8 showing a slight translation error combined with 90° rotation while holding the intensity constant.

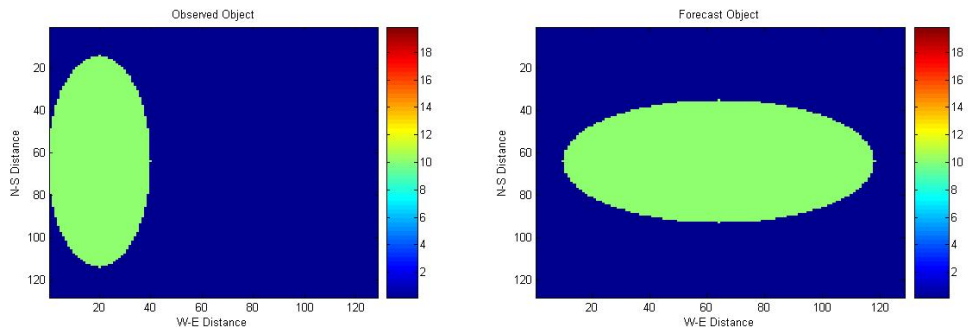


Figure 4.1.9. Idealized case 9 showing the differences in translation from observed object to forecast object while blowing up the forecast object so there is some slight overlap with the observed object. Intensity structure remains the same.

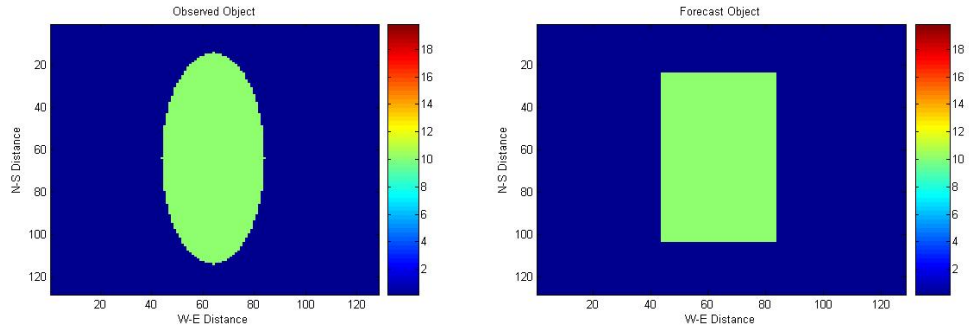


Figure 4.1.10. Idealized case 10 having similar intensity structure with different shapes only.

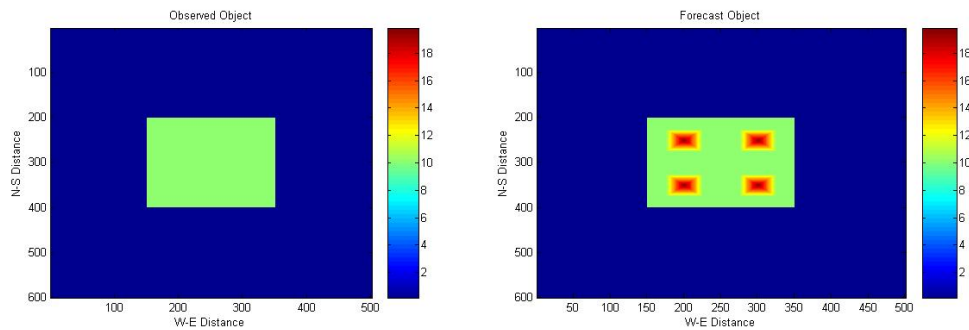


Figure 4.1.11. Idealized case 11 with areas of high intensity peaks in the forecast object rectangle with uniform intensity in the observed rectangle.

the different sized objects (Figure 4.1.9). It is no surprise that of all the idealized cases we find these having the highest penalties of all 14 idealized cases.

The last of the idealized cases are special cases that do not fit into a generic category, but still test the benefits of using the shape analysis scheme. Case 10 (Figure 4.1.10) simply concerns objects with the same centroid and intensity structure and even similar size; however, one object is a rectangle and one is an ellipse. Square objects may represent coarse model domains where precipitation realizations do not have smoothed features. Small errors exist in this case. Case 11 utilizes two square objects; however, the forecast object has peaks of reflectivity contained within its outer boundary (Figure 4.1.11). Case 12 (Figure 4.1.12) is a smoothed versus a noisy ellipsoid much

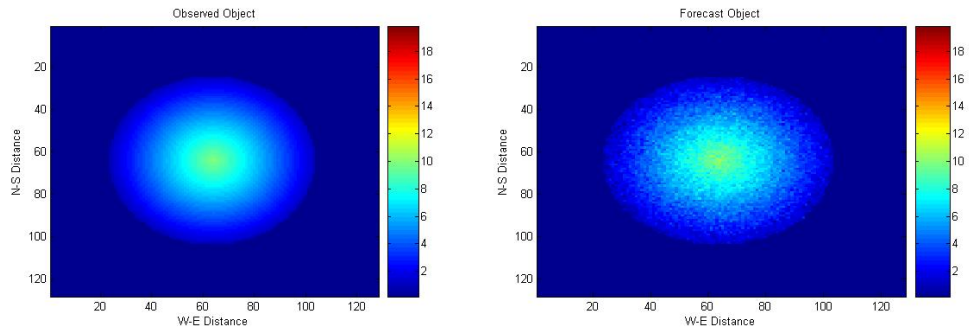


Figure 4.1.12. Idealized case 12 having circular object with a normal distribution with the forecast object being noisy.

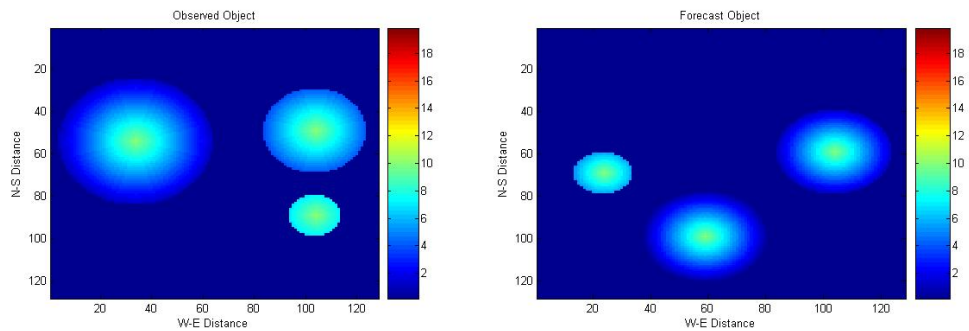


Figure 4.1.13. Idealized case 13 with three observed objects being simply misplaced in the forecast object. There is no change in object size, shape, or intensity.

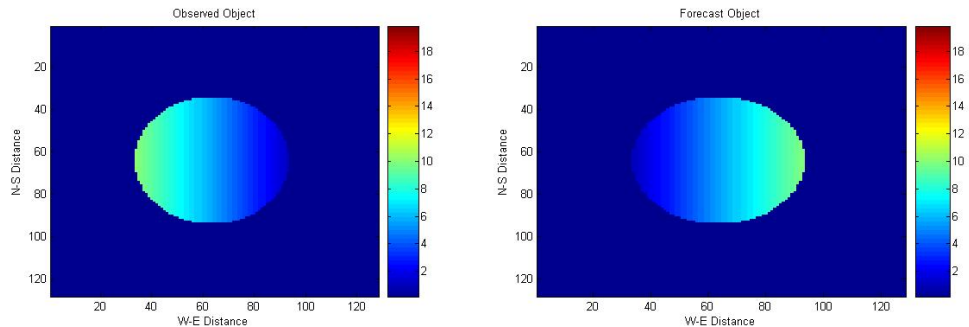


Figure 4.1.14. Idealized case 14 with the objects being mirrored in intensity.

Case #	Total Penalty from (3.1.4)
5	3.1706E-026
13	9.8632E-025
6	1.0835E-024
14	6.1000E-003
3	1.1475E+000
11	1.0069E+002
1	1.3419E+002
12	1.5497E+002
2	3.0000E+002
10	2.1305E+004
7	4.0677E+004
9	4.3809E+004
8	7.4343E+004
4	7.5281E+004

Table 4.1.1. Procrustes shape analysis total penalty for the 14 idealized cases using the penalty function described by (3.1.4) without the rotation fix and using RSS (shape) minimization for the match. The cases are listed in ascending order based on their total penalty.

like case 1 and also yields a moderate-sized penalty. Case 13 uses three objects in the observed field and misplaces the similarly shaped objects in the forecast field (Figure 4.1.13), which actually leads to one of the smallest penalties in the study based on (3.1.4). Finally, case 14 (Figure 4.1.14) includes ellipses that are simply mirrored objects. This type of error could be seen in a real situation in linear convection with either a trailing stratiform region or a leading stratiform region.

Idealized cases were utilized in order to test the applicability of the new Procrustes shape analysis verification tool to a variety of situations. This was done to examine the possible benefits of using such a system which can breakdown error components which include information of size, shape, and intensity. This was also done to see what improvements could be made to the

current system or to illustrate how end users may adjust the penalty function to fit their specific needs. Table 4.1.1 illustrates the differences in the overall penalty function to each of the idealized cases.

As a first examination of the penalties associated with various descriptions of the idealized cases may not make intuitive sense at first, a quick description of the breakdown of some of the idealized cases may help. The first question that arises is why idealized case 13 (Figure 4.1.6) has the second lowest penalty associated with it while the forecast looks totally worthless. The answer lies in the selection of how to match the objects within the verification scheme. This matching is based on minimizing the RSS portion of the penalty function instead of a minimum centroid distance matching. Since there are three objects that perfectly match in terms of shape, the match is made this way and yields only a translation error component for each cell. If one were to change the matching characteristic to a distance-weighted method a much larger error would result, as now there would be size, shape, and intensity errors to consider.

Another question arises in the form of the translation error cases having lower errors despite being separated by great distances to their counterparts. Again, in the current framework of the penalty function, the translation component is not utilized. This can be an instance where the end-user may wish to increase the penalty function based on translation if timing is the main concern as opposed to intensity. The aviation community comes to mind in this instance as a thunderstorm of any intensity is a threat; however, the timing of the event is the main player in keeping take-off and landing schedules intact.

To illustrate making the initial penalty function (3.1.4) more meaningful, consider altering the penalty function into components of similar

magnitude and including all possible errors including translation (SS_T), rotation (SS_R), dilation (SS_D), intensity (SS_{min} , SS_{max} , SS_{avg}) and shape (RSS) errors. In the adjusted version of the Procrustes scheme the errors are handled on a cell-by-cell basis, therefore a squared error measure is used for each match paired rather than a sum of squares. Once each matched pair of cells has an adjusted cell-based penalty, a total penalty is found by averaging the adjusted penalty for each cell found within the domain. For example, if there are three matched pairs (idealized case 13) in the domain the adjusted penalty 4.1.1 is calculated for each pair and then averaged for a total domain penalty (D_{domain} , 4.1.2). This is a mean penalty based upon the total number of observed cells (N_{obs}). An example of such an adjusted cell-based penalty function is given as (4.1.1), where SE represents the squared error in the residual (shape: RSE), minimum intensity (SE_{min}), maximum intensity (SE_{max}), average intensity (SE_{avg}), dilation (SE_D), rotation (SE_R), and translation (SE_T) for one cell match.

$$D = RSE^{0.5} + SE_{min} + SE_{max} + SE_{avg} + |1 - SE_D| * 100 + SE_R * 100 + SE_T^{0.5} \quad (4.1.1)$$

$$D_{domain} = \frac{\sum D}{N_{obs}} \quad (4.1.2)$$

The adjusted cell-based total penalty given each idealized case now makes more intuitive sense and the values based on the penalty function (4.1.1) can be seen in Table 4.1.2. The values in Table 4.1.2 are representative of the mean cell-based adjusted total in the given domain from 4.1.1. The advantage to using a cell-based penalty function allows the user to examine error characteristics of a particular cell of interest in the domain. Equation 4.1.1 is an example of how an end-user may choose to calculate a total penalty based on individual interest; this illustrates the robustness of the adjustment to the original Procrustes scheme from Micheas *et al.* (2007). The above

equation (4.1.1) attempts to bring all of the error components measured by the Procrustes shape analysis verification tool to a common order of magnitude. For example, the maximum rotation error for an individual cell is $\pi/2$ (1.57) for a maximum squared error of 2.47. For an object that needs no resizing (dilation = 1), the SE_D term in 4.1.1 becomes 0. A large squared error for translation may be on the order of 10000 (100 km) outlining the need for adjustments to the cell-based error to get everything close to the same order of magnitude; hence multiplication factors on dilation and rotation and the square roots of some other squared error components in 4.1.1. A perfect forecast results in $D=0$ in 4.1.1 for an individual cell, D may be considered a squared error term for a matched pair of cells; for more than one cell in a domain the total domain penalty would be a mean squared error.

It is now evident that the objects that must be rotated and translated to a large extent have the highest total penalty, as well as those cases that differ in intensities. Cases 7 and 8, respectively, deal with situations of the truth and forecast object having the same intensity structure but having signification size, rotation and translation error components. The errors would be magnified to a greater extent had the intensity components exhibited difference as well. Cases 12 and 14 exhibit the smallest errors as the objects have similar shape and size with the same centroid and have slightly different intensity structure. Case 14 is a mirrored case, which shows some weakness in the Procrustes scheme. However, if a different threshold was used in the mirrored case the objects would look as if they have different shapes and thus would increase the penalty function. Case 12 is just a random noise intensity field based on the true field so the result has minimal error.

Case #	Total Penalty (from 4.1.1)
14	2.43
12	22.63
5	102.56
1	129.00
10	129.46
11	197.48
2	197.99
4	246.74
3	384.86
6	396.35
9	486.15
13	528.83
7	562.95
8	640.74

Table 4.1.2. Procrustes shape analysis total penalty for the 14 idealized cases using the penalty function described by (4.1.1). The cases are listed in ascending order based on their total penalty.

A final note to using (4.1.2) in the current framework of the verification scheme involves the use of a penalty for the incorrect number of cells from forecast to truth. When there are fewer observed cells than forecast cells the only assessment from the original framework is to penalize the matches. The matches may give a low total penalty and the user would be left thinking that the forecast is good despite the false alarms. In the adjusted scheme each false alarm causes an increase in the total domain penalty (D_{domain}) by multiplying the ratio of forecasted cells (N_{fore}) to observed cells (N_{obs}) with the total penalty from all the matches in the case there are more forecasted cells than observed (4.1.3).

$$\text{if } N_{obs} < N_{fore} : D_{domain} = \frac{\sum D}{N_{obs}} \left(\frac{N_{fore}}{N_{obs}} \right) \quad (4.1.3)$$

Another interesting result comes in the examination of case 13 shown in Figure 4.1.5. In the original version of the Procrustes verification scheme the objects were matched on similar shape, which minimized the error in intensity which minimizes the total penalty in (3.1.4) simply because no translation error was accounted for in (3.1.4). An additional run of the Procrustes scheme was used to recalculate the values in Table 4.1.2 using a minimum distance matching approach. When examining the breakdown in error for idealized case 13, translation, dilation, and intensity errors are all greater than zero yielding a high total penalty. While utilizing the Procrustes scheme which minimizes the shape error (RSS), the breakdown in error changes by eliminating the intensity and dilation errors and leaves a different translation error. When using (4.1.1), the total penalty using the minimum distance scheme is 528.83 and the total penalty using the minimum RSS scheme is 263.38. It therefore seems best that the minimum distance version be used as it makes more intuitive sense until the schemes can be combined. Case 13 is the only sample case with multiple objects out of the idealized cases examined and it may be worthwhile to look at future cases with multiple objects for evaluation of the different matching schemes.

Overall, the usefulness of this particular verification scheme is readily apparent. It allows for the breakdown of error into components and the penalty function can be user-defined so that important errors become the dominant player in the end result. The use of these idealized cases also highlights the benefits of testing all verification schemes with some baseline set of cases. It may highlight limitations from one system to the next and provide insight to the developers on which areas need work in the particular scheme. It also allows an assessment of the meaning of the magnitudes of

errors produced in cases where the verification scheme uses an open-ended error scale.

4.2 WDSS-II Nowcasts and Procrustes Results

With the adjusted penalty and the rotation problems corrected, the Procrustes scheme will also be used to evaluate nowcasts made by varying the range of reflectivity used in the K-means nowcast product within the Warning Decision Support System—Integrated Information (WDSS-II) system (Lakshmanan *et al.*, 2007 and Hondl, 2002). WDSS-II is a platform primarily used for diagnosing and disseminating severe weather information. Included in the package are products that include hail, mesocyclone, and tornado detection algorithms along with rainfall estimation and other products useful for operational forecasting. A nowcasting tool (K-means) is also available in the package which gives the forecaster a pseudo-radar reflectivity image at user-specified lead times (Lakshmanan *et al.*, 2003).

The K-means product uses a hierarchical clustering technique to find storms at different spatial scales and estimate the motion at these various scales to provide a useful nowcast (Lakshmanan *et al.*, 2003). The tracking approach uses a combination of motion estimates for groups of storms rather than individual storms while estimating the motion of these groups of storms at various scales. The forecast images are calculated by a matrix of a smoothed mean absolute error and adjusted according to data based on the centroid (Lakshmanan *et al.*, 2003). The nowcast scheme also handles growth and decay by keeping information on the mean values of reflectivity of the

cluster. Within the WDSS-II software package the range of reflectivity used to identify clusters can be selected by the user.

The methodology for an alternate application for the WDSS-II nowcasting scheme is quite simple. The user can select different ranges of reflectivity on which to calculate the storm motion to generate the nowcast. The default storm motion calculation within the segmotion algorithm utilizes a range from 20- to 60-dBZ, which then gets divided into bins for storm motion calculation. By simply changing the lower and upper end of the range, the user can alter the bins, thus changing how the storm motion is calculated on different spatial scales. Systematically altering the range of reflectivity and running the nowcaster multiple times can be used to generate an ensemble product by taking the mean of the solutions. Although the ensemble image generated may not give the best skill scores, it can be used to determine the value of the individual members that make up the ensemble.

A sample case for this alternative WDSS-II K-means nowcast methodology is shown using data from 13 January 2007 centered over the New Braunfels, TX (KEWX) radar. The case involved a morning training event which caused major flooding north of San Antonio, and there were accompanying reports of damaging winds from possible downbursts. The example will illustrate the different threshold ranges for storm motion used in the K-means methodology and standard verification skill scores and Procrustes shape analysis penalties will be shown for the valid time of 1000Z (Figure 4.2.1). The images are composite images using a maximum composite technique found in the WDSS-II software. This technique finds the maximum reflectivity in a vertical column and fits it into a Cartesian (1 km x 1 km) planar view.

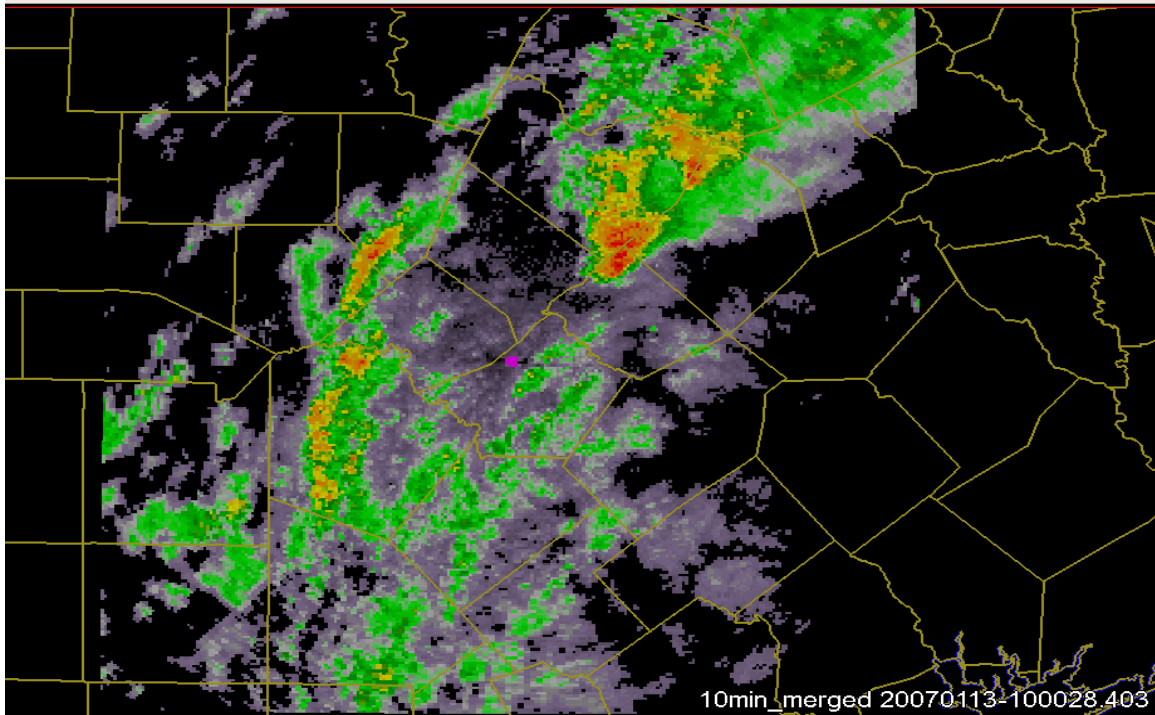


Figure 4.2.1. The 1000Z composite (maximum) centered over KEWX. The pink dot is the radar location and the main cell of interest lies to the NE of the radar.

The WDSS-II K-means nowcasting tool groups objects at different scales into clusters with clusters identified at different scales representative of the potential motion over different lead times (Lakshmanan *et al.*, 2003). There is a built-in growth/decay function that allows the radar echoes in an identified cluster to get stronger or weaker over time based on trends and possibly convergence of motion estimates. An example of the storm motion estimates from the WDSS-II nowcasts for varying reflectivity ranges shows the difference in both growth and decay and advection of reflectivity (Figures 4.2.2 and 4.2.3).

In the example, the resulting forecast images for the two reflectivity ranges with the same overall spread of 30 dBZ show two different have

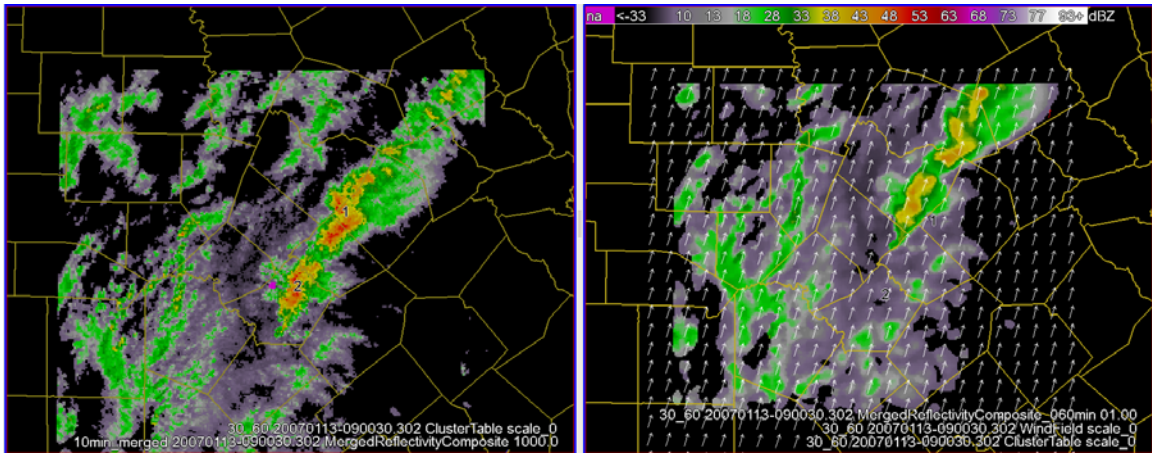


Figure 4.2.2. Reflectivity from 0900Z (left) with clusters identified at scale 0 by numbers 1 and 2 with the resulting motion vectors and intensity for a 60-min nowcast with cluster identifiers 1 and 2 for the 30 to 60-dBZ range.

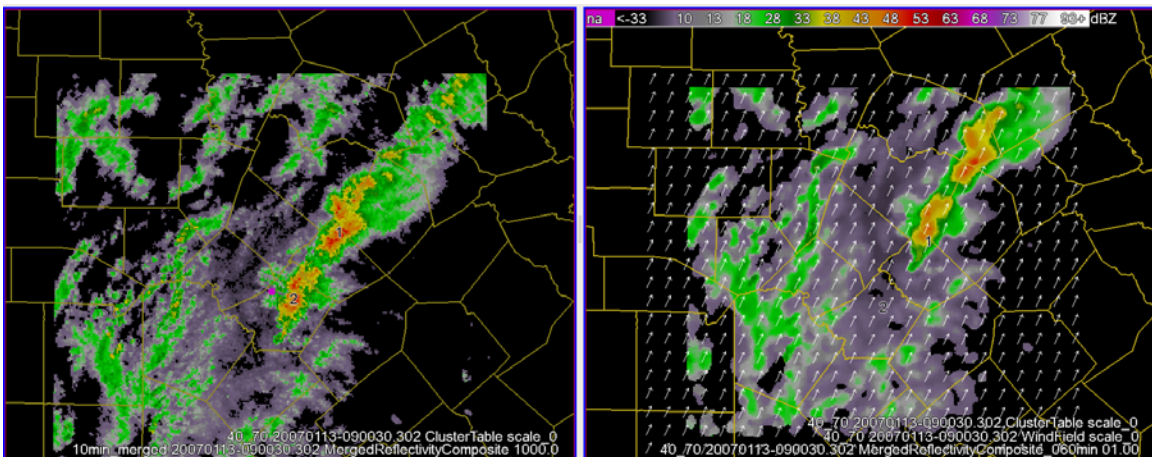


Figure 4.2.3. Reflectivity from 0900Z (left) with clusters identified at scale 0 by numbers 1 and 2 with the resulting motion vectors and intensity for a 60-min nowcast with cluster identifiers 1 and 2 for the 40 to 70-dBZ range.

slightly different shapes and intensity values. The 40- to 70-dBZ range (Figure 4.2.3) tends to have slight speed convergence in the winds near the strong reflectivity cores while the 30- to 60-dBZ tends to lack convergence of any kind near the strongest reflectivity. This may result in a strengthening

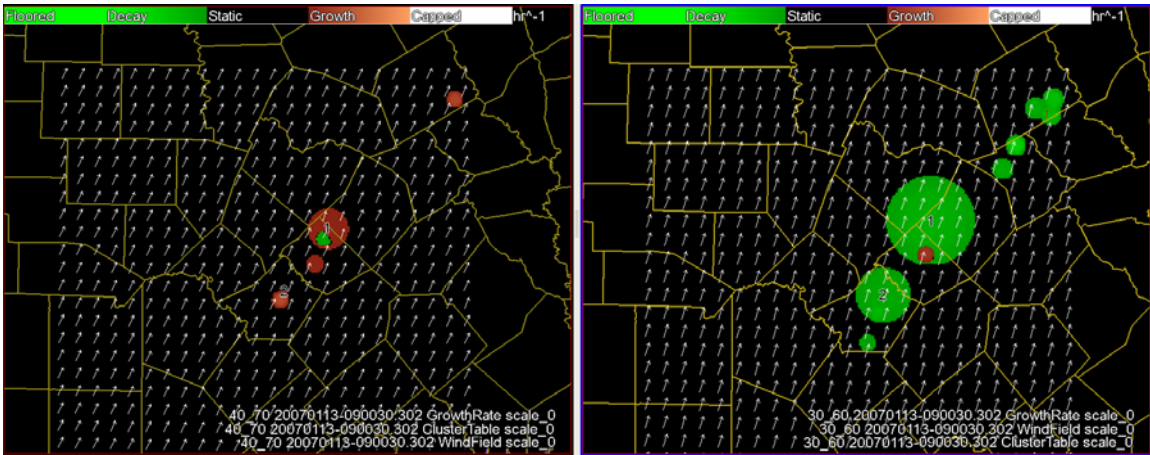


Figure 4.2.4. The comparison between growth rates for the 40 to 70-dBZ range (left) and the 30 to 60-dBZ range (right) overlaid with the forecast wind field. Growing tendencies are shown with red colors while decaying tendencies are shown in green.

trend in the 40- to 70-dBZ forecast over a 60-min lead-time forecast while the 30- to 60-dBZ forecast allows the forecast to decay. It is important to note that altering the range of reflectivity in the segmotion algorithm alters the motion vectors in the domain; the motion vectors then advect all reflectivities in the domain. The growth and decay rates for the two different ranges can be seen in Figure 4.2.4.

Throughout the sample case from 13 January 2007, five different reflectivity ranges were used to create nowcasts valid from 10 minutes to 60 minutes lead time every ten minutes. These reflectivity ranges were 20- to 40-, 30- to 50-, 20- to 60-, 30- to 60-, and 40- to 70-dBZ, as well as a mean ensemble created from the five different ranges. Comparisons of these forecasts are done using standard skill scores (POD, FAR, CSI), various intensity plots over the range of lead times, and the modified Procrustes shape analysis technique. Results of this will be summarized in the following paragraphs.

4.2.1 Standard (Categorical) Skill Scores

Standard skill scores, probability of detection (POD), false alarm rate (FAR), and critical success index (CSI) over the different WDSS-II K-means selected reflectivity ranges show what is expected of most forecasts: deterioration over time. However, even in the traditional skill scores differences can be seen by varying the range of reflectivity used in the K-means algorithm. Three different verification thresholds were used for assessment of the standard skill scores 20-dBZ (Figure 4.2.5), 30-dBZ (Figure 4.2.6), and 40-dBZ (Figure 4.2.7). These were selected to incorporate a range of spatial scales with the 40-dBZ threshold representing the smaller, more intense, less predictable scale.

From these figures it is apparent that by simply changing the threshold over which to verify will significantly alter the forecast skill. Therefore, in order to maximize the forecast skill one simply has to lower the threshold over which the forecast is to be verified to obtain the best possible score as inherently the areal coverage of the forecast and observation will increase leading to the possibility of more overlap. Conversely, if one wishes to illustrate poor performance, simply raising the threshold will push the observations and forecasts to a more unpredictable scale and thus yield poorer overall skill scores. It is also of note to point out the lead time at which the forecasts for the different reflectivity ranges of the K-means nowcast diverge. In the 20-dBZ verification images (Figure 4.2.5), the solutions begin to diverge at around 40-min lead time while the 30-dBZ verification threshold diverges near 30-min lead time and the 40-dBZ verification threshold forecast solutions diverge after the 20-min lead time.

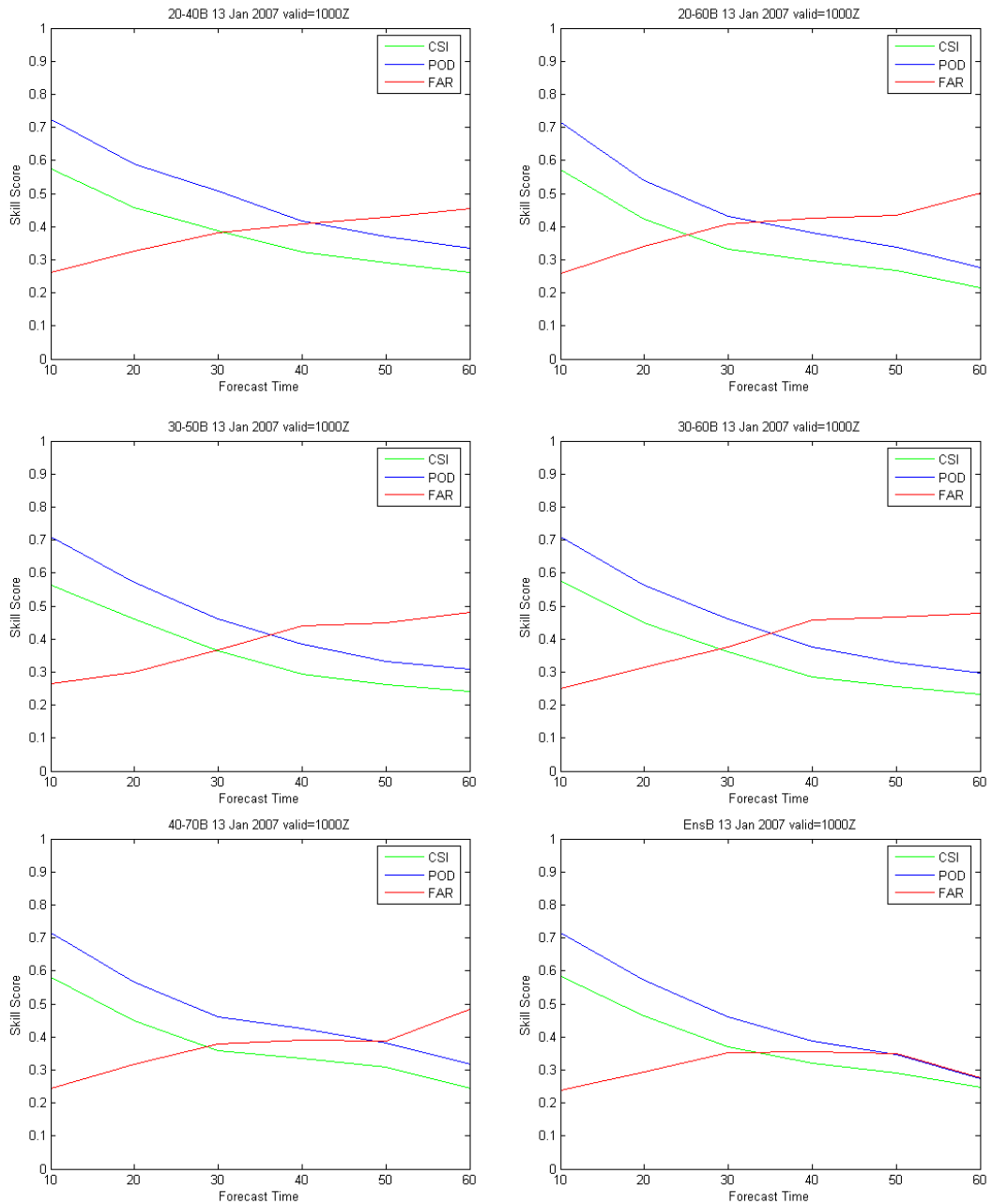


Figure 4.2.5. Skill scores (POD, FAR, CSI) for the 20 to 40-dBZ range (top left), the 20 to 60-dBZ range (top right), the 30 to 50-dBZ range (middle left), the 30 to 60-dBZ range (middle right), the 40 to 70-dBZ range (bottom left), and the mean ensemble of the above 5 ranges used (bottom right) using a 20-dBZ verification threshold. Forecast lead time (x-axis) is in minutes.

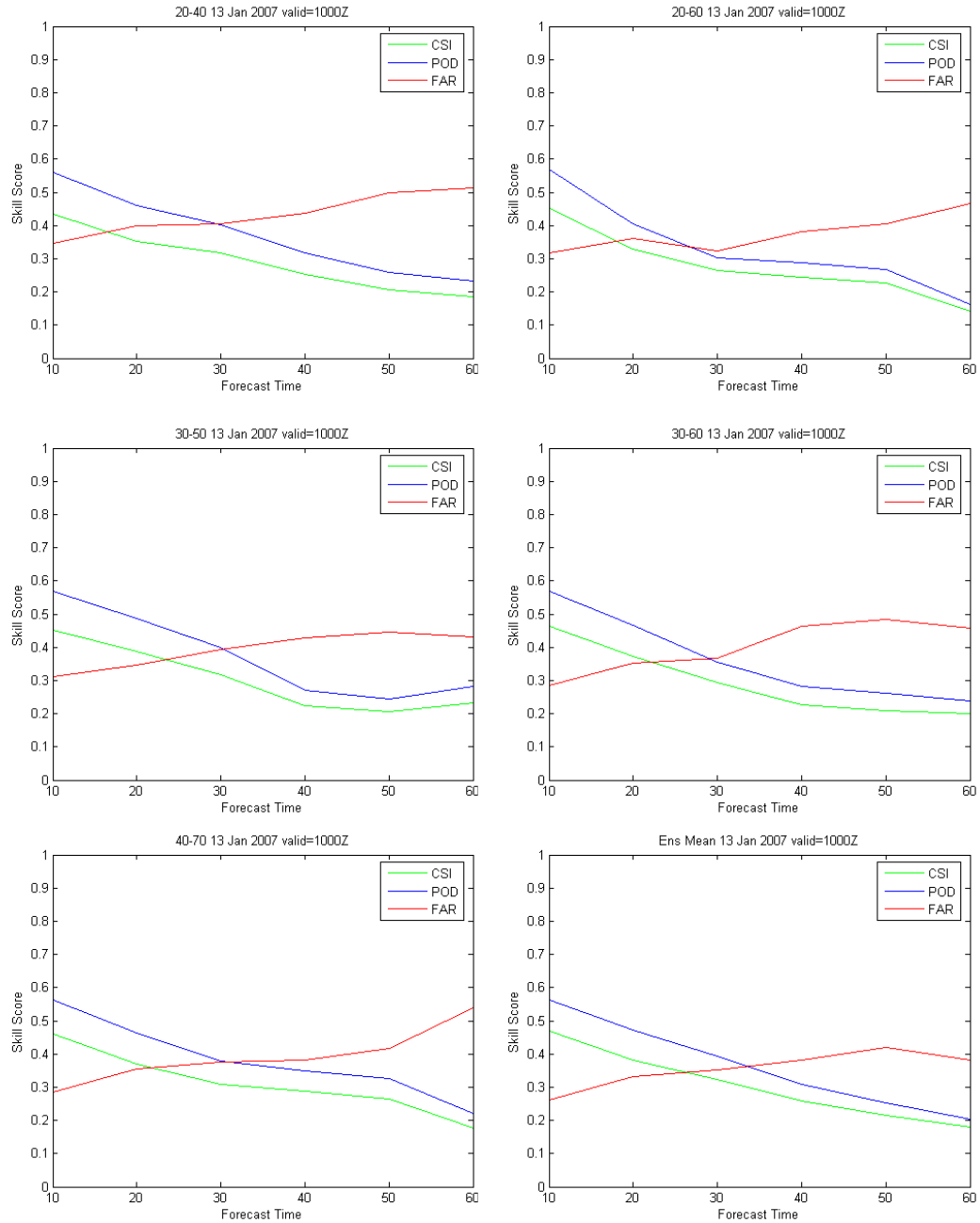


Figure 4.2.6. Skill scores (POD, FAR, CSI) for the 20 to 40-dBZ range (top left), the 20 to 60-dBZ range (top right), the 30 to 50-dBZ range (middle left), the 30 to 60-dBZ range (middle right), the 40 to 70-dBZ range (bottom left), and the mean ensemble of the above 5 ranges used (bottom right) using a 30-dBZ verification threshold. Forecast lead time (x-axis) is in minutes.

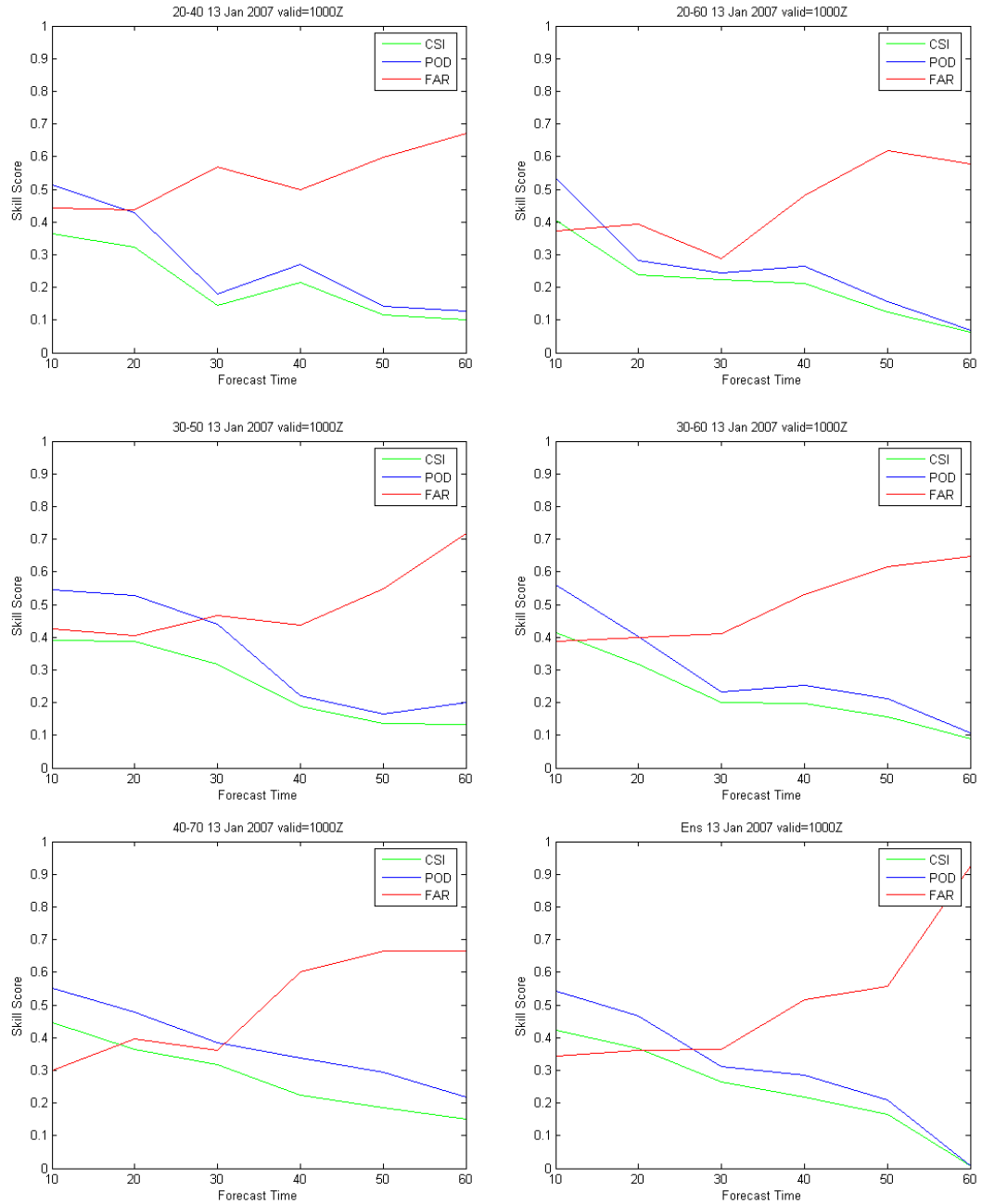


Figure 4.2.7. Skill scores (POD, FAR, CSI) for the 20 to 40-dBZ range (top left), the 20 to 60-dBZ range (top right), the 30 to 50-dBZ range (middle left), the 30 to 60-dBZ range (middle right), the 40 to 70-dBZ range (bottom left), and the mean ensemble of the above 5 ranges used (bottom right) using a 40-dBZ verification threshold. Forecast lead time (x-axis) is in minutes.

Of final note, in the figures above, it appears that the lower bounded reflectivity ranges of 20 dBZ for the K-means solution perform better when verifying against a 20-dBZ threshold, while the 30- to 50-dBZ range and the 40- to 70-dBZ range outperform the others at the 30- and 40-dBZ thresholds over the 60-min lead time.

The next image (Figure 4.2.8) shows overall trends in the area of reflectivity over a rainfall rate of 10 mm hr⁻¹ using the convective Z-R relationship of $Z=300R^{1.4}$. This attempts to examine the intensity trends of the forecasts over a 60-min lead time created by varying the reflectivity ranges in the WDSS-II K-means algorithm. The 20-dBZ verification threshold was used in creating the reflectivity trend figure. Figure 4.2.8 is viewed as the ratio of the number of pixels with over a 10 mmhr⁻¹ rainfall rate to the total number of (precipitation) pixels after the verification threshold is applied. From Figure 4.2.8 it can be shown that for all reflectivity ranges except for the 30- to 50-dBZ range and the 40- to 70-dBZ range the area is under-forecast as the 60-min lead-time is approached. In the case of the 30 to 50-dBZ and the 40 to 70-dBZ ranges the forecast area greater than 10 mm hr⁻¹ is actually over-forecast at the later lead-times. It may be possible that this is from a converging forecast motion solution for those thresholds leading, to an enhancement of precipitation after a 30-min lead time.

Although standard skill scores (Figures 4.2.5-4.2.7), along with a measure of forecast bias in intensity (Figure 4.2.8), can yield some interesting forecast trends. It is also apparent that one can simply increase the skill scores by lowering the threshold to achieve the desired result of a successful forecast or raise the threshold for greater penalty. These scores may be useful when looking at many forecasts over a long period of time in order to find systematic biases. In the case of the above, many cases from many dates and

times should be used in order to determine forecast trends over the different reflectivity ranges in the K-means nowcast algorithm in WDSS-II. However, in order to assess forecast skill as a snapshot to give feedback in a near-real time situation more metrics are needed. This situation is evident in the transportation industry, as timing and structures of incoming storms, need to be disseminated in a timely fashion. Understanding errors in shape, intensity, and timing of the storms is essential to the success of such operations. Therefore, for the above sample case, Procrustes shape analysis penalties will be examined over 60-minute lead times in order to determine the usefulness of such a system to breakdown error components and give a sense of overall skill for certain lead times.

4.2.2 Procrustes Verification

Like standard skill scores, the Procrustes technique requires a user-defined threshold as well as a user-defined minimum cell size. The assessment of penalties for the various reflectivity ranges of the K-means nowcasts will be displayed for different verification thresholds as was done for the standard skill scores. Plots will be constructed for the overall adjusted penalty for the entire image as well as error decomposition penalties for the main cell in the image. These results will be displayed as a function of lead-time as was done previously to examine the possibility of seeing any trends in the penalties derived from the Procrustes shape analysis technique. Figure 4.2.9 shows the adjusted penalties from (4.1.1) for each of the reflectivity range nowcasts from WDSS-II for a 30-dBZ threshold and a minimum size of 40 km² and for a 20-dBZ threshold at 200 km². The larger area was used for the 20-dBZ threshold in order to preserve meaningful precipitation producing cells.

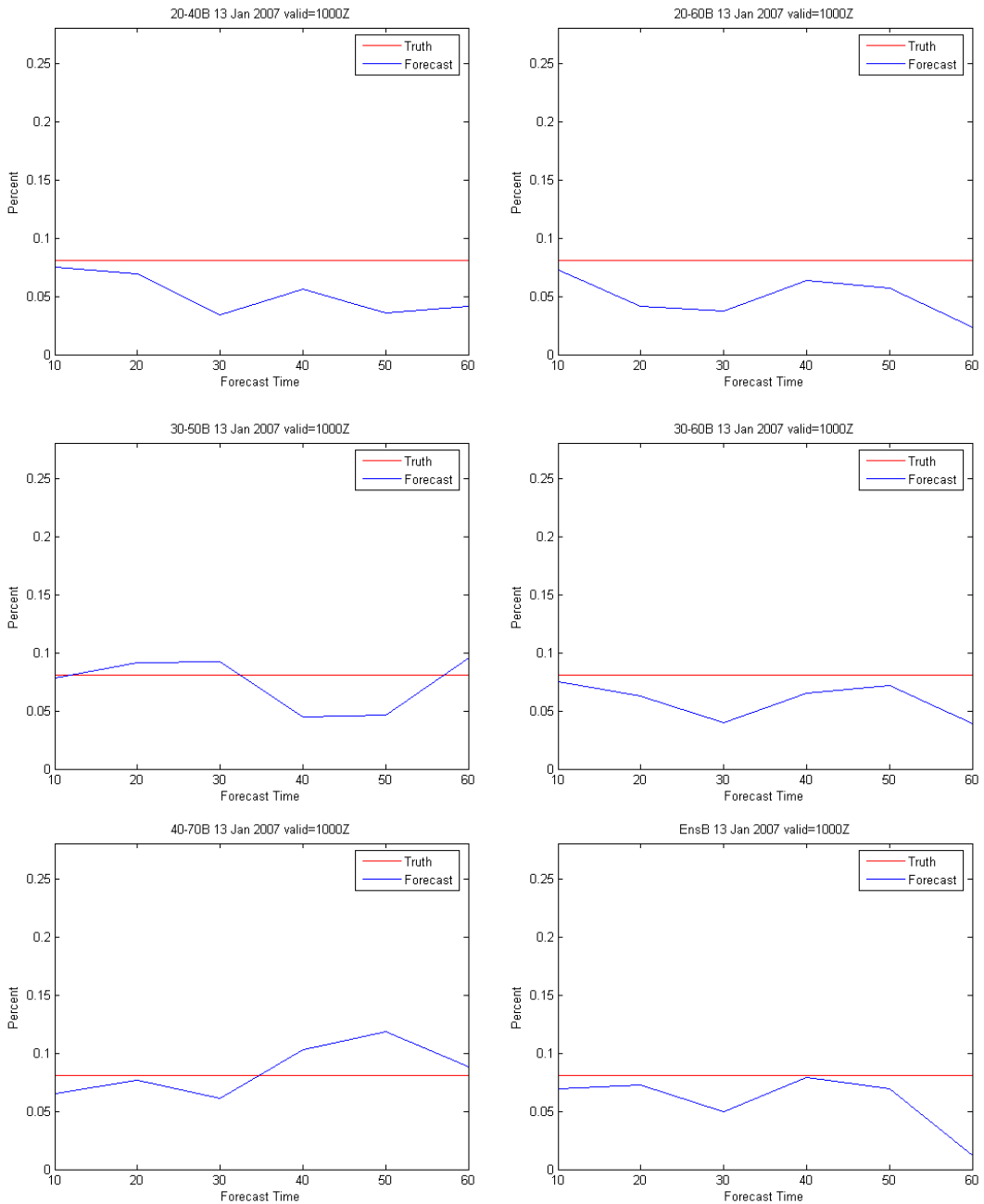


Figure 4.2.8. The percent of area receiving at least 10 mmhr⁻¹ for the 20 to 40-dBZ range (top left), the 20 to 60-dBZ range (top right), the 30 to 50-dBZ range (middle left), the 30 to 60-dBZ range (middle right), the 40 to 70-dBZ range (bottom left), and the mean ensemble of the above 5 ranges used (bottom right) using a 20-dBZ verification threshold. Forecast lead time (x-axis) is in minutes.

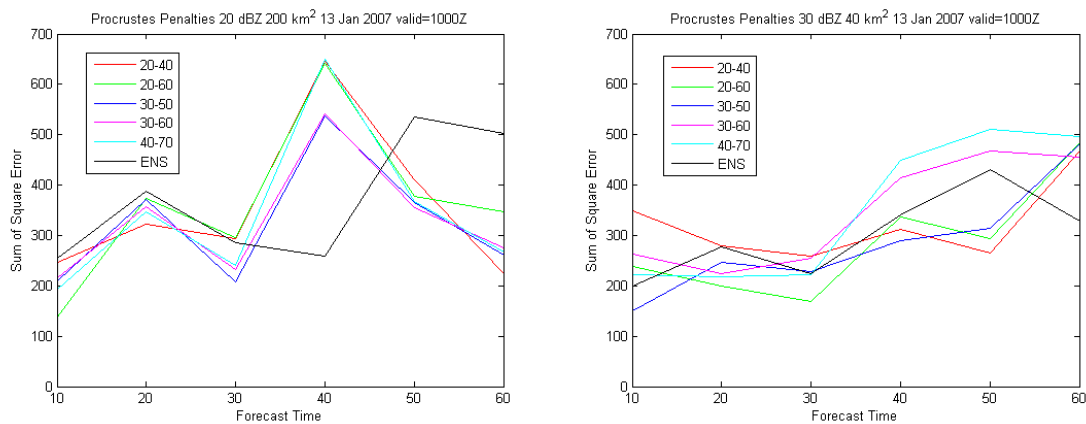


Figure 4.2.9. The adjusted Procrustes penalty function over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes.

From Figure 4.2.9 it can be shown that the general trend of both forecasts is that the penalty increases over time, with a slightly smoother increase in the 30-dBZ threshold verification. There is a noticeable penalty peak in the 20-dBZ threshold verification at the 40-min lead time. This peak is not uncommon in many different runs of the verification scheme due to the sensitivity of identifying cells at different thresholds. Identified cells may change size and shape due to the changing intensity of the forecast; this can also affect the cells that are matched from forecast to observation and may act to enhance or degrade the calculated penalty. Nonetheless, it is shown in this example that the problem is significant for the 20-dBZ identified cells at the 40-min lead time for all reflectivity ranges used to create the K-means nowcasts. The problem is corrected by taking the mean of all solutions as shown by the black mean ensemble curve in the above figure. Overall, the

penalty function is sensitive to the threshold selected to run the Procrustes scheme.

Although many trends can be shown from examining the penalty function over time for the entire domain, a strength of the Procrustes scheme comes from the examination of individual cells that make up the given forecast. The output table given from the scheme allows the comparison of a single identified cell over the 60-min lead times at different WDSS-II nowcasts derived from different reflectivity ranges. The main cell of interest for the 13 January 2007 at 1000Z is located in the NE quadrant of the image and takes different characteristics at the 20-dBZ and 30-dBZ thresholds for identifying objects (Figure 4.2.10).

Once a particular cell is identified the error breakdown of the total penalty for the cell from (4.1.1) and the components that make up the total error can be readily examined. Plots can be constructed for the cell for the entire series of 60-min lead time forecasts for each error component of interest. As shown by the previous figures, indicating the overall adjusted penalty for the entire domain, the 30-dBZ threshold and 40 km² minimum cell size has the smoothest increasing error over time. This is not the case for the main cell's overall Procrustes penalty and the individual components that make up this error for the main cell. The 20-dBZ threshold and the 200 km² minimum cell size seems to have the smoother increasing penalty traits for all reflectivity ranges used to create the K-means nowcasts, which may reflect good forecast continuity in the case of an accurate portrayal of forecast quality. For the overall Procrustes penalty of the main cell (Figure 4.2.11), there is a noticeable peak in the 20- to 60-dBZ range nowcast at a lead time of

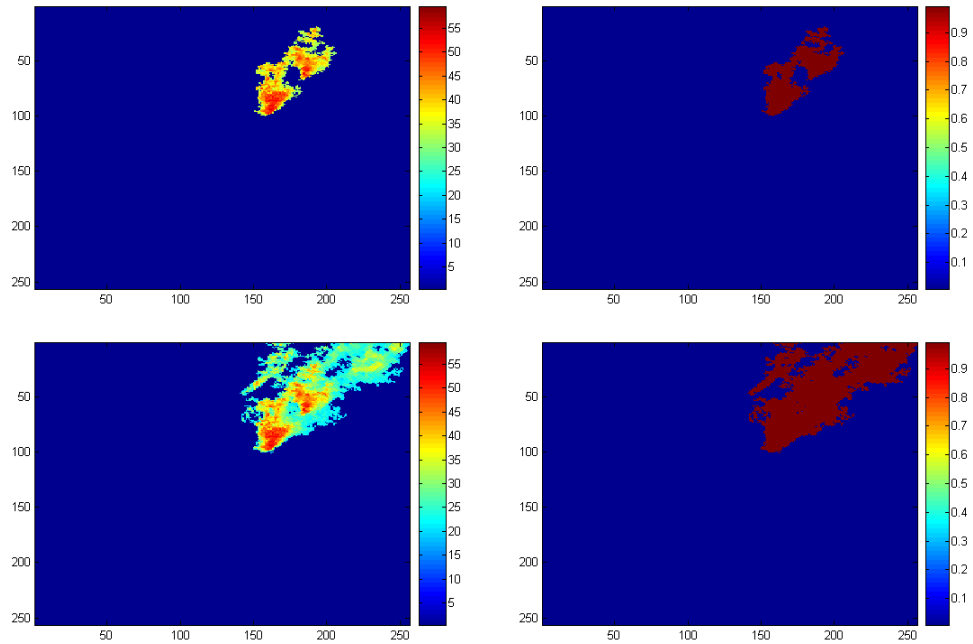


Figure 4.2.10. The cell of interest from 13 January 2007 at 1000Z for the 30-dBZ threshold identification (top) and the 20-dBZ threshold identification (bottom).

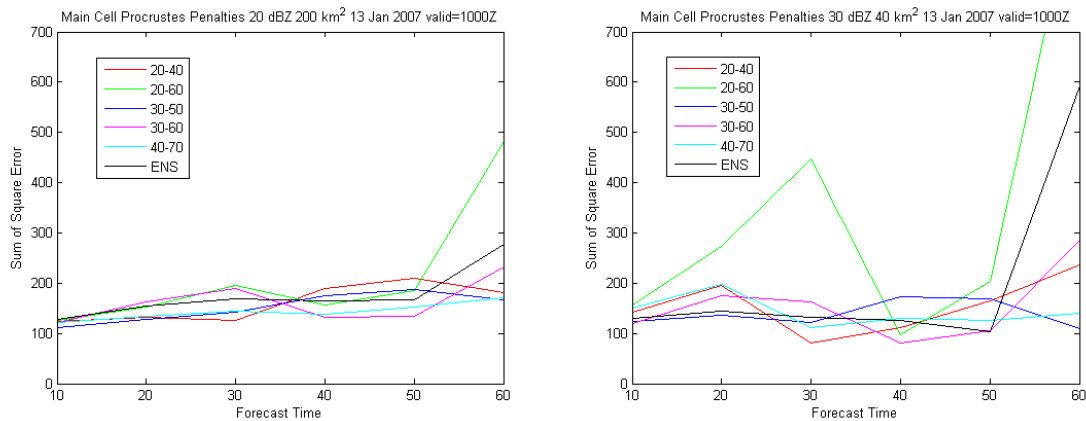


Figure 4.2.11. The adjusted Procrustes penalty function for the main cell over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes.

30-minutes in the 30-dBZ threshold case. The penalty for remaining nowcasts for both verification thresholds remains almost constant through the first 50 minutes of the nowcast, with a significant increase at the 60-min lead time. Most of the nowcast members exhibit slight differences, but as the 20- to 60-dBZ, the default threshold for K-means, is noticeably different (higher) the end-user may choose to examine the individual error components that may account for this variation.

The component which is of greatest value to many forecast users is the translation penalty (Figure 4.2.12). The translation, or displacement, penalty can be considered a timing or motion error in a forecast. Many communities such as aviation and emergency management value information on the correct location and timing of a forecast in cases of strong thunderstorms. Figure 4.2.12 shows once again a smooth trend in the translation error as the 60-min lead time is approached using the 20-dBZ verification threshold; however the error in translation is sporadic in the instance of the 30-dBZ verification threshold. The peak in the overall error at 30-min lead time in the 30-dBZ verification threshold image (Figure 4.2.11) does not appear to result from the translation error. Instead, as a whole, the translation errors for many of the ensemble members in the 30-dBZ verification seem to decrease through a lead time of 50 min.

Since the translation penalty did not provide enough insight into the overall penalty spike at the 30-min lead time of the 20 to 60-dBZ range nowcast verified at 30-dBZ, the dilation error can be examined (Figure 4.2.13). Right away the peak in dilation error is apparent at the 30-min lead time for the 30-dBZ verification. The squared error for the main cell in this case is near 3.5, which corresponds an almost doubling in size of the forecast object in order to fit it correctly to the truth object. Most of the other members of the

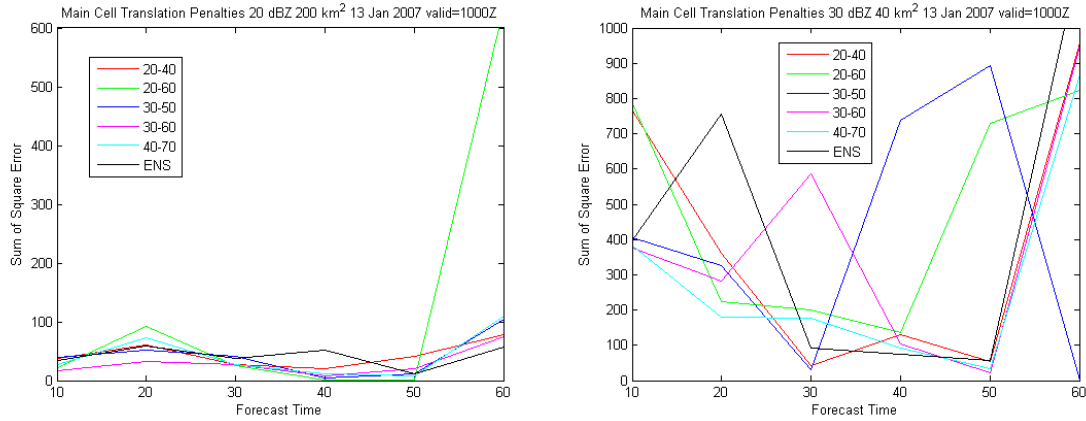


Figure 4.2.12. The Procrustes translation penalties for the main cell over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes.

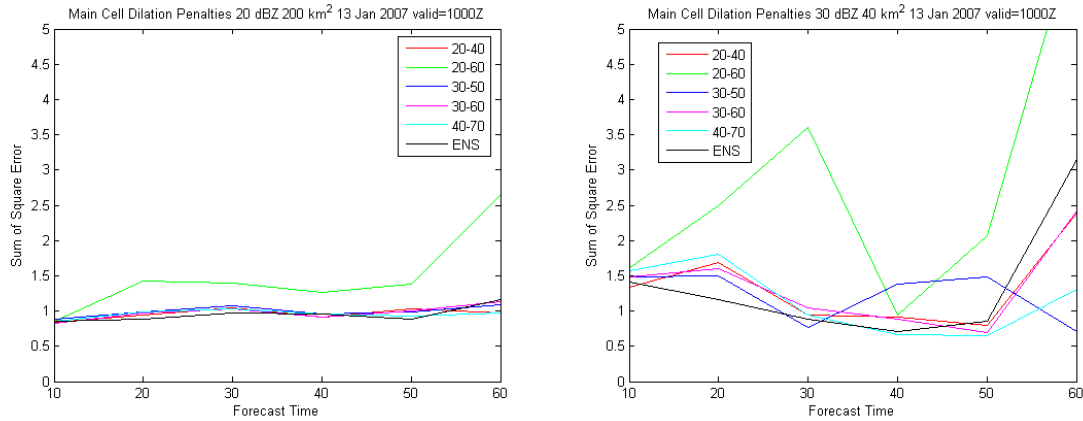


Figure 4.2.13. The Procrustes dilation penalties for the main cell over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes.

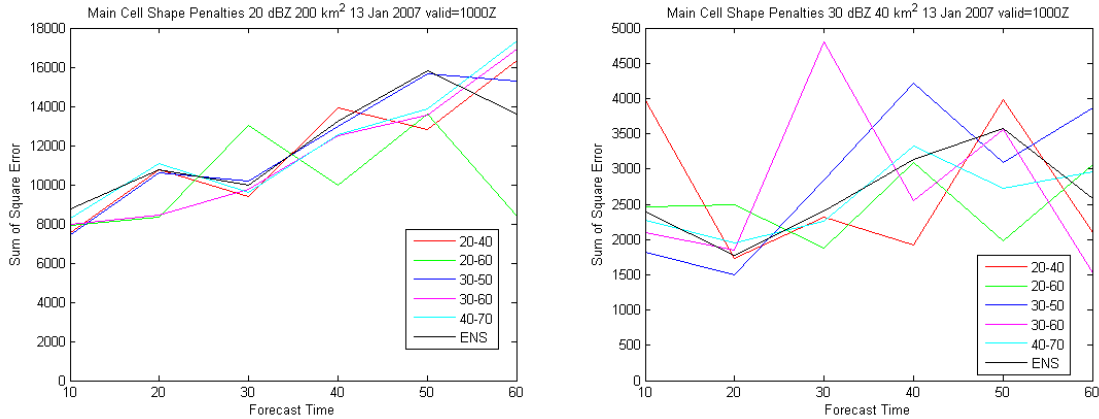


Figure 4.2.14. The Procrustes shape penalties for the main cell over a 60-min lead time for 13 Jan 2007 valid at 1000Z for a 20-dBZ threshold and a 200 km² minimum cell size (left) and a 30-dBZ threshold and a 40 km² minimum cell size (right). Forecast lead time (x-axis) is in minutes.

nowcast suite must also be enlarged at the 30-dBZ verification threshold, but only slightly in comparison to the 30-min lead time peak.

Although we can also examine the rotation and intensity aspects of the total Procrustes penalty, these were relatively negligible in terms of differences in the overall trends for the different nowcast members. The final error component to be examined is the shape (Figure 4.2.14). The shape is also considered a goodness of fit error component. Once the image is properly sized, translated, and rotated, the images are assessed in terms of how well their shapes match. This is accomplished by comparing the landmarks along the cells edges from the truth to the forecast object. In the case of the main cell in this example, the errors for both verification thresholds generally increase with increasing lead time. Once again, the 20-dBZ, 200 km² verification seems to have the smoother appearance with all nowcast members; however, the order of magnitude of the error is almost four times that of the 30-dBZ, 40 km² verification. In terms of the contribution

to the total Procrustes penalty for the main cell, the shape penalty contributes twice as much in the 20-dBZ verification than the 30-dBZ verification. This appears to be simply caused by the size of the objects compared. The larger the truth object, the more the shape penalty increases for the given threshold.

It is of note that the shape penalty may also be used as a matching tool. Finding the corresponding cells (observed to forecasted) that have the lowest shape penalty may be substituted for using a minimum distance. This is known as Procrustes matching. Methodology will soon be created to incorporate Procrustes matching with the minimum distance matching currently used in the verification tool.

Chapter 5 Verification Conclusions

5.1 Procrustes Verification Conclusions and Future

Overall, the total Procrustes penalty from (4.1.1) for the entire domain shows some usefulness in assessing overall skill when making a comparison to nowcasts on a similar scale (in this case different nowcast members created by the K-means algorithm). However, in the case of many end-users, we may only be interested in cells that pose an immediate threat to a particular region. This highlights the strength of the modified Procrustes scheme; the ability to breakdown the total Procrustes penalty for the identified cells within the domain. It can also breakdown the individual components of the total Procrustes penalty for the entire domain and the individual components for individual cells as illustrated in the above section. Although no conclusions can be drawn about the systematic errors in the different nowcast ranges in the K-means scheme, it is evident that the Procrustes verification scheme can highlight the differences between the nowcast members. It may be useful to keep logs of many forecasts created by the K-means scheme at different reflectivity ranges to assess possible situation in which one range works better than another. From the above example, it seems as though the standard 20- to 60-dBZ range may not be the best to use for a convective situation although many more cases would need to be run in order to justify this claim.

Although only two verification thresholds appear in the work above, three were explored. In addition to the 20-dBZ and 30-dBZ threshold a 40-dBZ threshold was tested and not shown in the results. An important feature of the 40-dBZ threshold was that at longer lead times for different nowcast members a forecast object was not found. In the current architecture of the Procrustes scheme the return is 'no cells in realization.' This makes comparing the error to the other verification thresholds nearly impossible at the longer lead times and occasionally even a lead time of 30 minutes. This is a problem that may need to be examined thoroughly to obtain a solution. A penalty must some how be quantified if there are cells in the truth and none in the forecast, or conversely, cells in the forecast and none in the truth. Although the solution to this problem is not readily apparent, varying the intensity thresholds and minimum cell sizes on which to verify the forecast shows that solutions diverge quickly as the scale of the forecast moves to more and more unpredictable levels (*i.e.* smaller areas of more intense reflectivity).

From examining the breakdown of error components for the main cell in the example above, it was also found that the different components are not independent. For example, the change in intensity over time causes there to be considerable location, dilation, and shape errors. Given the certain verification threshold, the size of the object may change, which in turn changes the dilation error and the translation error. One possible solution is to attempt to separate the shape and location penalties from the intensity penalties by utilizing constant reflectivity forecasts over lead-times to assess the location and shape errors first and then determining errors in possible intensity given the background reflectivity with the previously derived shapes. More work would have to be done to determine if this produces

meaningful results. However, either holding shapes constant from truth to forecast or holding intensity constant from truth to forecast may be a reasonable solution for eliminating the threshold problems inherent to all object-based verification approaches. It also may be possible that converting from reflectivity units into rainfall totals or rainfall rates may make the penalty function more reasonable and perhaps eliminate some threshold issues for QPF. Another possible application of this scheme would be to utilize both a distance- and shape-weighted scheme, as well as an intensity-weighted scheme so that cells of similar nature and life cycle are matched appropriately.

There are a few remaining problems with the current version of the Procrustes scheme. The first is that the dilation error component in the adjusted penalty from (4.1.1) should be altered. The current formulation does not penalize the same forecast for being twice as large or half as small as the truth object. A way to correct the current penalty for the dilation penalty is to use (5.1.1), where SE_{dil} is the squared error of dilation for a matched pair in a forecast domain and Dil would be substituting in equation 4.1.1 for the previous adjusted dilation error term. For a perfect fit for a single matched pair in the domain, with respect to size, $SE_{dil}=1$.

$$\begin{aligned} \text{if } SE_{dil} > 1, Dil &= (1 - \frac{1}{SE_{dil}}) * 100 \\ \text{if } SE_{dil} < 1, Dil &= (1 - SE_{dil}) * 100 \end{aligned} \tag{5.1.1}$$

The above equation ensures that a forecast cell that is twice the size of the truth cell is penalized the same as a forecast cell half the size of the truth. Overall, the strength of the Procrustes scheme comes in the robustness of the

penalty function. This allows for many different end-users to weight the different error components differently depending on their interests.

Another problem is that there should be an interface for the ease of evaluation. In its current framework a final array is created showing all of the identified truth cells and their corresponding matches with their individual error components with the total Procrustes penalty from (4.1.1). The scheme currently produces a number of images for each observed cell and its match and does not produce an overall image consisting of all the cells in the domain. This problem could be corrected to result in an overall image and statistics appearing on the same image. Significant errors can be flagged by setting end-user requirements for significant errors. An easy flag is to identify important cells in the domain by strength of intensity and the size of certain cells or even the location of a certain cell relative to a surface landmark (GIS potential).

Finally overall criticism of the scheme comes from the lack of comparing forecasts made on different times at different locations. In other words, there is no one measure of forecast quality that is understood without comparing two forecasts together for a similar time and place. An exploratory measure to fix this problem is the creation of a forecast quality skill score. This skill score uses a coarse numerical model forecast as a baseline comparison for overall forecast quality. In the case of a radar-based nowcaster, a numerical model at a coarser resolution should under-perform in most cases in terms of rainfall rate. Evaluating the Procrustes penalty for the numerical model and comparing to the Procrustes penalty of the radar-based nowcast product should yield a forecast quality based against an inferior product. An example of such a coarse model precipitation rate

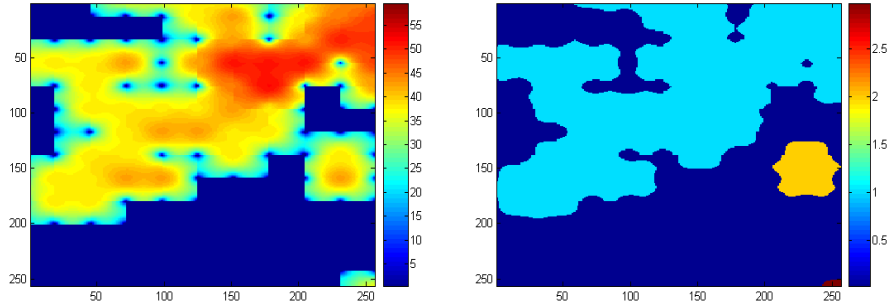


Figure 5.1.1. The precipitation rate (left) and the cell identification (right) using a 30-dBZ threshold valid at 1000Z on 13 January 2007 converted to reflectivity from the RUC-20 for comparison against the K-means nowcast members.

Threshold	CSI	FAR	POD
20 dBZ	0.22446	0.7619	0.7967
30 dBZ	0.10058	0.89795	0.87486

Table 5.1.1. POD, FAR, and CSI for the two verification thresholds for the RUC-20 forecast valid at 1000Z on 13 January 2007.

forecast converted to reflectivity is found using a Rapid Update Cycle 20-km run (RUC-20) (Figure 5.1.1).

The RUC-20 forecast is valid only at a 60-min lead-time which will therefore only be compared to the other 60-min lead-time nowcasts from the WDSS-II K-means scheme. Clearly when compared to the truth image (Figure 4.2.1) valid at 1000Z the RUC-20 forecast over-forecasts the precipitation area and intensity. This is evident in the increase in FAR and the decrease in CSI over the 20-dBZ and 30-dBZ verification thresholds (Table 5.1.1). Although POD is high for the RUC-20, it is necessary to examine the

error components from the Procrustes shape analysis for additional information on forecast quality compared to the WDSS-II K-means nowcasts.

A forecast quality score can be derived by comparing the different K-means nowcasts to the RUC-20. A forecast quality score (5.1.2) can be utilized for the average Procrustes penalty for all cells in the domain (average D for all matched cells from 4.1.1) or any of the components that make up of the penalty.

$$\begin{aligned} &\text{if } D_{RUC20} < D_{K\text{-means}}, \text{Quality} = 0 \\ &\text{if } D_{RUC20} > D_{K\text{-means}}, \text{Quality} = 1 - \left(\frac{D_{K\text{-means}}}{D_{RUC20}} \right) \end{aligned} \quad (5.1.2)$$

Table 5.1.2 shows the forecast quality score for the different K-means nowcasts created with the different reflectivity ranges over the 20-dBZ threshold at the 60-min lead time by comparing the total penalties from (4.1.1). From Table 5.1.2 it is evident that all of the K-means nowcasts, except the mean ensemble, exhibit skill over the RUC-20 forecast at a 60-min lead time for the 20-dBZ verification threshold. From the breakdown of the error for the main cell both the translation penalty and the average intensity forecast quality scores are near 1, which would be a perfect nowcast solution. The translation penalty comes as a result of the over-forecasting of the rainfall area in the RUC-20 forecast. Although the RUC-20 over-forecasts the area, the shape of the area matches more closely to the observed main cell in Figure 4.2.1 than any of the nowcast realizations, as illustrated by the zero scores in the shape column of Table 5.1.2. By varying the verification threshold one can assess improvement of forecast quality versus the RUC-20 for storms of differing intensity if that is of interest a particular forecaster. The method of the forecast quality score can easily be incorporated for the end-user to judge

Nowcast	Forecast Quality Score (20-dBZ)					
	Total Domain Score	Main Cell Score	Main Cell Max Int	Main Cell Avg Int	Main Cell Translation	Main Cell Shape
20-40 dBZ	0.5408	0.4186	0.0000	0.9242	0.9671	0.0000
20-60 dBZ	0.2904	0.0000	0.0000	0.9569	0.7310	0.0000
30-50 dBZ	0.4657	0.4654	0.4321	0.8634	0.9557	0.0000
30-60 dBZ	0.4399	0.2535	0.0000	0.9568	0.9686	0.0000
40-70 dBZ	0.4537	0.4527	0.0000	0.8931	0.9535	0.0000
Mean Ens	0.0000	0.1128	0.0000	0.9961	0.9755	0.0000

Table 5.1.2. The forecast quality score for the entire domain (Total Penalty) and the main cell in the image located to the NE of the radar in Figure 4.2.1.

the value of a forecast for the entire domain and of cells in the domain of importance to the given user. In future developments of the quality score it may be useful to allow the scores to be negative as a measure of how poorly the nowcast solution is compared to the coarse resolution model.

Verification of meteorological fields, especially in the case of precipitation, is a practice that is continually evolving as numerical prediction models are moving towards higher temporal and spatial resolutions. Although standard skill scores (POD, FAR, CSI, MAE, RMSE, and bias) may be used over time for the examination of systematic errors, they may not paint a complete picture for certain end-users. The development of object-oriented verification procedures allows more insight into forecast errors than categorical skill scores may illustrate. The use of object-oriented approaches makes it possible to examine certain cells that may be of particular interest, whereas standard skill scores incorporate the entire domain of a forecast. The research herein described a new Procrustes skill score and its utility to examine nowcast realizations created by the WDSS-II K-means algorithm. Not only was the utility of the Procrustes scheme illustrated; the K-means algorithm was shown to have an interesting ability in creating different

solutions by simply changing the range of reflectivity to calculate cell motion. Although only one example was worked through completely, the usefulness of both the K-means nowcaster and the Procrustes shape analysis verification tool could provide a framework for operation use in many different fields utilizes meteorological data. The points brought out by this research can be summarized in the following bullet points.

- The adjustment to the Procrustes original framework in terms of a total penalty function (4.1.1) shows the robustness of the scheme and the corrections to problems that arose while examining idealized cases.
- The utility of breaking down the total penalty into its various components (translation, rotation, dilation, and intensity) and by selecting individual cells in the domain make this scheme a powerful tool. Examining different cells may allow the user to find flaws in forecast motion at different locations in the domain.
- The Procrustes methodology can work with any type of spatial field, reflectivity, rainfall, VIL, and NWP model fields if desired.
- The WDSS-II K-means nowcast can produce a wide range of realizations by changing the range of reflectivity to define clusters. This results in motion differences and intensity difference perhaps caused by convergence in the wind field.
- The forecast quality score can be a useful tool for the end-user to quickly assess forecast skill when compared to a baseline forecast of coarser resolution.
- The Procrustes shape analysis scheme has a few weaknesses shared by other object-oriented approaches: using a threshold to define a cell, matching forecast to truth objects based on minimizing the centroid distance, and real-time efficiency.

- The Procrustes verification tool may be adjusted to match cells based on minimizing shape differences or centroid distances or some combination and these and other criteria may be used to enhance the matching step of the process.

Chapter 6 Cell Identification and Classification

6.1 Introduction to Cell Classification

When examining precipitation systems with radar, many different types of objects may be identified. The simplest distinction between types of precipitation objects is by separating convective precipitation areas from stratiform precipitation areas. Such an automated procedure for applying Z-R relationships from has been developed and employed by hydrometeorological users (*e.g.* Biggerstaff and Listemaa, 2000). In instances such as this, the primary reason for identifying cells as convective or stratiform leads to improvements in estimating rainfall rate from radar. Work has also been done to automate a process to distinguish simplified objects, linear versus cellular systems, given precipitation (Stage IV) analysis (Baldwin *et al.*, 2005). Both methods above provide a framework for improving quantitative precipitation forecasts (QPF) in some way and can give insight into mesoscale processes, but the process of identifying specific types of convective cells remains unexplored. The work herein provides a framework for an automated convective storm-cell classification algorithm, which includes linear convective systems, supercells, pulse thunderstorms and non-severe/severe discrimination. The algorithm uses composite reflectivity data combined with near-storm environmental (NSE) data derived from the RUC-20 gathered from the WDSS-II system.

The identification and classification of storm cells in this framework primarily is aimed at adjusting life cycle parameters in a nowcasting (0-2 hour forecast of reflectivity) framework. The underlying idea is that classifying a convective system and retaining information from prior time steps can lead to an understanding of growth, decay, and morphology of a particular storm. One may also choose to identify and classify convective storm elements for verification and climatological studies. This may be particularly useful in object-oriented verification procedures, described in the previous chapters, where matching objects from the truth to the forecast remains the biggest challenge.

6.2 Classification Background

Since the early days of radar, meteorologists have been categorizing radar echoes into storm types based on the shape and intensity structure present in a plan position indicator (PPI) scan, and have used information based on range height indicator (RHI) scans. From these storm types, work would be done to identify characteristics of the near-storm environment present in for each identified storm type. The work herein attempts to utilize NSE data from RUC-20 initialization fields combined with radar data to identify a storm type. Understanding different storm types and their physical mechanisms for formation and development can lead to forecast improvement. Developing an automated classification scheme for real-time application may assist forecasters in a nowcasting (0-2 h) timeframe. In particular this work is aimed at providing input parameters to constrain convective storm development and motion within a Bayesian hierarchical

nowcast scheme such as ones described in Xu *et al.* (2005) and Fox and Wikle (2005).

Early classification of mesoscale phenomenon included defining different observable phenomenon, primarily from WSR-57 PPI scans. Such phenomena included the squall line. Over time the squall line definition has broadened and many sub-classes of squall lines have been defined. An early study of classification of squall lines or mesoscale convective systems (MCS) from Bluestein and Jain (1985) utilized WSR-57 data combined with satellite observations of cloud extent. Their study consisted of 40 usable cases for classification of cells into five categories: broken-line squall lines, back-building squall lines, broken-area squall lines, embedded-area squall lines, and isolated supercells. The isolated supercell was used strictly for a baseline comparison. From the visual classification of PPI scans, the observed classes of squall lines were then compared against representative soundings derived from National Weather Service (NWS) synoptic observations and rawinsonde data. The goal of this study was simple; to understand the physical differences between the squall line types. Fields such as convective inhibition (CIN), convective available potential energy (CAPE), relative helicity (H), the bulk Richardson number (BRN), boundary layer moisture, and squall line orientation compared with shear were used. A conclusion to this study drew similarities to the isolated supercell and the back-building squall line. This important comparison will be used in later attempts to examine hybrid supercell/MCS systems later.

Parker and Johnson (2000; 2004) extended the earlier work from Bluestein and Jain (1985) to break down the squall lines into MCS archetypes. Three main archetypes were introduced including the trailing stratiform, leading stratiform, and parallel stratiform MCS (Parker and Johnson, 2000).

The parallel stratiform is similar in nature to the previously derived back-building squall line. From the three main archetypes, further discrimination depended on the relative direction that the storm was receiving inflow. The three main archetypes were then given forward-fed (FF) or rear-fed (RF) prefixes (Parker and Johnson, 2004). The motivation for this study was, again, to understand the physical environment of the MCS archetypes. The advances in modeling for this study allowed the authors to produce high resolution simulations and NSE parameters of the MCSs. The models included in that study included the non-hydrostatic Advanced Regional Prediction System (ARPS) developed by the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma. The simulations were found to differ mainly due to variations in the shear profiles and the magnitude of shear at the low-levels as well as the deep-shear magnitude. An additional observation from Parker and Johnson (2000) revealed that most MCSs will evolve into the classic trailing stratiform case which is most frequently observed. Overall the work reflected the earlier results found in Bluestein and Jain (1985).

Although MCSs represent a large portion of observable significant damage producing storms, numerous studies comparing radar and satellite observations to observed or numerically-derived near-storm environments have been done pertaining to the supercell. Like similar studies on MCSs, supercells have been found to have three main archetypes: classic; low-precipitation (LP); and high-precipitation (HP). Primarily research has focused on being able to predict not only general supercell formation given observations and modeled environments, but to determine what type of supercell to expect. One such study utilizing near-storm environment parameters in the case of supercells comes from Rasmussen and Blanchard

(1998). This paper serves as a baseline for supercell climatology and is based on over 6000 near-storm soundings. The breakdown of the supercells from Rasmussen and Blanchard (1998) included tornado-producing storms, 2" or greater hail-producing storms with or without weak tornadoes, and ordinary storms (non-supercells) producing no 2"-hail and no tornado with at least 10 cloud-to-ground (CG) lightning strokes. In addition to the parameters examined from Bluestein and Jain (1985), other shear, CAPE, CAPE-shear, and low-level thermodynamic parameters were examined as derived from the National Center for Atmospheric Research (NCAR) Mesoscale Model, version 5 (MM5). Parameters such as storm-relative helicity (SRH), the energy-helicity index (EHI), and the vorticity generation parameter were included in this study.

In the Rasmussen and Blanchard (1998) study, it was found that although VGP and EHI were very good discriminators between supercell types, especially the ordinary storm versus supercell instance, the false alarm rate from strictly looking at these variables is high. SRH was a relatively good discriminator as the likelihood of tornado producing thunderstorms increases as SRH increases; the same was true of deep-layer shear magnitudes. BRN, CAPE, and CIN were shown to not be a good discriminator of storm type in this study. Overall, the database of storms used combined with the variables studied in this research provide a practical outline for cell classification.

Currently, there are a few methods utilizing WSR-88D data for the identification and classification of storms. One such method devised by Stalker and Knupp (2001) utilizes multiple Doppler radars in order to achieve an estimate of vertical (updraft) velocity within convective cores of multicellular storms. This method relates the updraft strength with

corresponding storm-top heights observed by radar in order to identify the stage and subsequent growth, decay, and merging of convective elements within a cluster of cells. This method has shown promise in multicellular storms for isolating individual convective elements; however, the availability of multi-Doppler analysis to derive the vertical updraft profile can be a significant issue. An additional study by MacKeen *et al.* (1999) attempted to forecast storm longevity using radar reflectivity-derived quantities. This research attempted to answer the evolution of the storm in terms of growth and decay, which is an important aspect of nowcasting. The conclusions of the study indicated that reflectivity-derived fields alone cannot be used to assess the potential of storm longevity.

An additional study by Baldwin *et al.* (2005) attempted to automate the process of classification using radar rain estimates combined with rain gauges using shape and structure characteristics. The Baldwin *et al.* (2005) methodology relies on a hierarchical framework of classes of rainfall. The first stage in the classification separates convective versus stratiform rainfall classes. Obviously, the convective class of storm generally produces higher rainfall totals than the non-convective, or stratiform, type. The second stage of the classification procedure is to separate the linear types of storms from those that appear cellular in nature. The work of Baldwin *et al.* (2005) utilizes Stage 4 precipitation accumulation fields over 1-h periods instead of using a single radar image. The work utilized a training data set and an expert classification that only takes into account precipitation totals and the shape of the area of the rain region. Baldwin *et al.* (2005) also used an arbitrary 3 to 1 rectangular fit to delineate linear systems from cellular systems. The introduction of a 15% enlargement of the rain area was used to indicate a smoother “trace field” of rainfall accumulation. The research from Baldwin *et*

al. (2005) for classification utilized the General Method of Moments (GMM) to estimate the shape parameter and the scale parameter in the gamma distribution, commonly used for precipitation studies. From the training cases used, comprising of 48 members, estimates of the shape and scale parameters were made with the corresponding known “expert” classification. A test phase was accomplished and the parameters were fit and compared with the “expert” classification. The method proved to work well to distinguish between convective and stratiform precipitation, and linear versus cellular coverage; however, it does not discriminate between the various types of convective cells in a way which would be useful in the nowcasting problem.

Chapter 7 Cell Classification Methodology

The following section will be divided into several subsections describing the development of the automated classification scheme. This division includes methodology on the WDSS-II algorithms used, expert classification procedures, identification of storm cells, and the building of a classification tree.

7.1 WDSS-II Usage

The WDSS-II system was used to a large degree for the data collection stage of the cell classifier and is described by Lakshmanan *et al.* (2007). The first set of data obtained for the classifier was Level-II radar data for a given radar site. The Level-II data have the highest spatial and temporal resolutions for current operational radar networks. Although the resolution of the radar data used is 1 degree (azimuth) by 1 km (gate spacing) a Cartesian transformation was applied to obtain data at a 1 km x 1 km resolution (w2merger: Lakshmanan *et al.*, 2006). Although the merger algorithm in WDSS-II was primarily developed to merge multiple radars into one domain, merging data from a single radar results in a simple Cartesian grid of radar-derived variables. There are many merging strategies available within the algorithm and for this case the default exponential weighting scheme in space is used (Barnes analysis). All elevation angles for the given

data are merged resulting in constant altitude plan position indication (CAPPI) slices. From the CAPPI slices a composite radar image was created using the maximum reflectivity in a vertical column. This composite creation was chosen so that the product would maximize the rain rate for a given pixel, giving a worst-case scenario. This merging technique to a Cartesian grid allows for the data to be easily processed in future steps of the classification scheme.

As this work attempts to classify storm cells for use in determining growth, decay, maintenance and morphology it is important to incorporate storm environment data as well. For the problem of forecasting storm longevity on the short (nowcasting) time scale it was found that reflectivity alone cannot be used; however, with the inclusion of some environmental parameters forecasting trends of growth and decay may be possible (MacKeen *et al.*, 1999). For the problem of incorporating the environmental data in a nowcasting time scale the Rapid Update Cycle at 20-km resolution (RUC-20) was considered the best solution as it has high spatial and temporal resolution and is used operationally. In addition, the WDSS-II suite readily ingests RUC-20 data and outputs gridded NSE data. For details on the architecture and processes of the RUC-20 model the reader is referred to Benjamin *et al.* (2002). The RUC model has been widely used in proximity sounding studies that focus on storm type and error characteristics of variables used from proximity soundings are relatively well known (Thompson *et al.* 2003).

WDSS-II readily ingests and converts RUC model grib files into fields used by the system to display basic fields of temperature, height, and pressure. From the basic fields, one can also run a near-storm environment algorithm which calculates parameters such as CAPE, SRH, CIN, and many

other variables over different layers. Some of these fields will be discussed in more detail later. For classification purposes, the storm environmental variables were calculated using the initialization data at the top of each hour for the duration of the event examined. The nearest radar timestamp was used to match the model data initialization to the reflectivity image. As an operational note, it may be necessary to run the identification of cells from forecast products from the RUC, as one needs to account for model issuance times. The combination of running the near storm environment algorithm with the merging algorithm within WDSS-II results in a product with a resolution similar to the Cartesian product of reflectivity with the same domain.

7.2 Identification of Storm Cells

The identification of the individual storm cells was based on the object-oriented verification work from Micheas *et al.* (2007). In this framework, cells are identified based on a user-defined threshold of reflectivity (in this case 30 dBZ) and a minimum size threshold (in this case 40 km²). For all pixels in the domain greater than the threshold a binary image is created. From this binary image, individual cells are identified if there is sufficient separation from their nearest neighbor. In other words, adjacent pixels within a cluster of reflectivity are tagged with the same cell identification number within the domain. The cell identification number ranges from 1 to the maximum cell number identified in the domain above the minimum size criteria. Once the cluster of pixels is identified as an individual storm cell and a number is assigned, storm attributes are derived for each cell in the domain. This is

accomplished by overlaying matrices of near-storm environmental data and reflectivity-derived data to the matrix of identified cells. This results in storm attributes for each individually assigned storm cell within the given domain. For example, for a given identified cell, reflectivity, POH, CAPE, CIN, and SRH for all the pixels contained in the given cell are overlaid. From these data, values of maximum, minimum, and average (mean value) storm attributes are attained for all cells in the domain on which the cell classification will be based.

7.3 Expert Classification

In order to create a working dataset for classification, storm cells had to be identified individually and subjectively classified initially. The initial dataset consists of 15 different dates from 2004 to 2007 with over 360 individual identified cells covering various geographic regions; however, a majority came from the Midwest. The cases also span different seasons so that the classifier would be able to identify storm type independent on the time of year; however, cases deal only with warm precipitation (snow-producing storms are omitted). A majority of the cases also have severe storm reports associated with the given date. A severe report was not a necessity but allowed for dates and regions where convection was likely occurring to be easily identified.

From the initial dataset subjective (“expert”) classification of the individual convective storm cells commenced. Radar-derived products formed the foundation on which the initial storm type classification was built. These products included those from the storm cell identification and tracking (SCIT) algorithm (Johnson *et al.*, 1998) contained in WDSS-II along with radar

reflectivity- and velocity-derived products such as the hail detection algorithm (HDA: Witt *et al.*, 1998) and the mesocyclone detection algorithm (MDA: Stumpf *et al.*, 1998). Model data from the RUC-20 output run through the NSE algorithm in WDSS-II also provided additional information. An example of this expert classification involving a supercell requires the examination of the lowest available elevation scan for proper reflectivity structure (a possible hook), the detection of a mesocyclone from the MDA, strong probability of hail (POH) or severe hail (POSH), and a sheared environment with available potential energy that would additionally support the claim of a supercell. The ultimate goal, in the case of the supercell, was to properly identify the possibility of a tilted, rotating updraft from the given data.

Eight different classification types were used in this study and are summarized in Table 7.3.1. An important distinction is made between air mass thunderstorms and pulse thunderstorms. Although similar in structure, the pulse thunderstorm has characteristics that allow for the rapid formation of hail and/or the possibility for a severe downburst. Also within the classification types, the two small scale rotating types may seem redundant; however, it was thought that significant differences may exist between severe and non-severe varieties. The non-severe variety may be thought of as an early stage or developing supercell while the severe variety may be more representative of a mini-supercell. These small-scale rotating storm types have a slightly different physical makeup than the mature supercell. In addition, the severe versus non-severe discriminator was used in order to potentially identify storms which have the likelihood of carrying severe

Classification Types (Abbrev.)	Brief Description
Air Mass Thunderstorm (AM)	Low Shear Environment
Pulse Thunderstorm (PULSE)	Low Shear with Potential Severity
Small-scale Rotating Thunderstorm (SRS)	Organized Convection within Rotating Environment
Severe Small-scale Rotating Storms (SSRS)	Same as Above with High Reflectivity
Linear Convective System (LCS)	Squall-type System with Potential Severity
Linear Convective System (Rotation) (LCSROT)	Squall-type System within Rotating Environment
Supercell (SC)	Severe Storm with Significant Rotation
General Thunderstorm (GEN T)	Pre-frontal Non-severe and Other Convection

Table 7.3.1. Eight classification types with abbreviations and brief descriptions used within this classification tree study.

attributes throughout their lifecycle in order to use the information for nowcasting storm severity. In the expert classification process, the severe versus non-severe discriminator was identified by using the Storm Prediction Center's (SPC) storm reports and the National Climatic Data Center's (NCDC) storm archives. In addition, maximum reflectivity signatures, the HDA, and the MDA were used in determining severe storms. The NSE fields produced by WDSS-II were again used to discriminate the potential for severe activity from organized convection with limited severe threat. This was essential in the case of discriminating between severe linear convective systems and general thunderstorms. General thunderstorms can take many forms as they are the left-over class. Often they are organized in a linear fashion along boundaries with weak forcing, which reinforces the need for using model data to aid in classification. Notice there is no severe versus non-severe discriminator for supercells as supercell in this case implies a well-developed supercell that has significant potential for hail, damaging winds, and even tornadoes.

There is some inherent subjectivity on the grounds of dividing the convective cells into eight categories. The categories themselves are relatively

subjective as one may argue that there is some significant overlap of the divisions of convective cell (*i.e.* supercell versus severe small-scale rotating storm). Some of the subjectivity also pertains to storm reports or lack of storm reports. Most of the classifications of storms into the severe category were based upon unofficial SPC storm reports. It should be noted that report times do not always correspond to the actual events and that some reports are missed especially in sparsely populated areas. During the expert classification process storms without physical storm reports may be classified as severe based upon appropriate WDSS-II information. Fields such as probability of severe hail, maximum estimated hail size, as well as fields generated by the mesocyclone and tornadic vortex algorithm were used to supplement decisions in the classification process.

7.4 Classification Tree Structure

Once the storms were individually identified using expert classification, a table was generated complete with all storm attributes and tagged with one of the eight classification categories. From this information, a classification tree process was used to determine the optimum use of variables to automatically detect cell type. The classification tree results in a tree of information with nodes at each branch representative of the best split of the data. These nodes have the advantage of yielding a probability at that given certain criteria at the particular pruning stage that the identified cell belongs to a given class. The use of pruning to get probabilities of cell classification yields more meaningful results as it accounts for some degree of uncertainty in the classification; however, the tree eventually leads to a deterministic solution. This uncertainty is based on the equal numbers of identified cells of

each of the eight categories set for classification. The classification tree scheme will be briefly described below. For a more detailed review of creating and utilizing classification trees the reader is referred to Breiman *et al.* (1984). Classification and regression trees have been recently applied and have shown success in predicting lightning over the United States and Canada (Burrows, 2007).

The overall objective of using a classification tree is to predict the membership of a certain object given its characteristics compared to a prior set of known classifications given a similar dataset of characteristics. Classification trees are useful in that they contain a hierarchical framework for working with data. Each step of the decision process for classification consists of what eventually turns out to be a series of “if else” statements branching from one initial decision (“root” node). The methodology consists of finding the best split possible at each node that results in the optimum sized tree for deterministic classification.

Two splitting rules were available for use in this experiment including the Gini diversity index (GDI) and the twoing rule. The goal of the splitting rule is to make a split at a given node that minimizes the misclassification cost. In the case of GDI, we are interested in obtaining a measure of node impurity for splitting. Node impurity is given by equation 7.4.1, where $I(t)$ represents the impurity at node (t) , f is the relative frequency of cases in t that belong to class c_j .

$$I(t) = 1 - \sum_{j=1}^x f(c_j, t)^2 \quad (7.4.1)$$

This method involves a process that looks at which class to be categorized is largest. It then attempts to separate the classes by examining one class at a time. The mean squared error (MSE) is used to measure the performance of

the classification tree for the GDI. For each class j , let c_j be an indicator variable that is 1 if the class for the example e is j and 0 otherwise; the MSE for this case is given by 7.4.2, where the expectation (E_e) is over all examples and $P_j(e)$ represents the probability assigned to class j for example e .

$$MSE = E_e \left[\sum_{j=1}^x (c_j(e) - P_j(e))^2 \right] \quad (7.4.2)$$

The GDI minimizes the resubstitution estimate for the MSE. This is advantageous when looking at unequal populations of events; it is also the more efficient choice and will be used throughout the research.

The twoing rule is different than the GDI in that the classes are selected in a way that divides the data into two halves on each split. The result is a classification scheme that does not simply attempt to extract one class initially, and this method will pull out the variable that maximizes the split into two equal populations.

The pruning technique inherent to classification trees is called minimal cost complexity pruning (Breiman *et al.*, 1984). This process assumes that there is a linear increase in the bias of the resubstitution with the number of nodes for the tree. This cost (R_α) is assigned to a subtree and is composed of two terms including a resubstitution error for the given tree ($R(T)$) and the number of leaves (N) times a complexity factor α (7.4.3).

$$R_\alpha = R(T) + \alpha N \quad (7.4.3)$$

From the cost equation (7.4.3), a series of trees are created while varying α from 0 to infinity. These trees are nested and the tree that minimizes the overall cost is selected to represent the group.

Overall, the GDI method to grow classification trees involves four steps. The first involves using 7.4.1 to calculate the diversity index on the parent node. The second step calculates the resultant GDI for the two child

nodes from the split at the parent node. Weights are then arranged according to the proportion of the parent node in each child. Lastly, an improvement measure is calculated by minimizing the cost from each created subtree.

Input into the tree growth algorithm includes a matrix of a series of variables each having a known classification. Again, the classification was broken down into eight convective cell types. The different variables used with brief explanations as to the input to classification tree scheme is shown in Table 7.4.1. The data in Table 7.4.1 are an illustration of some of the variables used for classification. The variables were limited by availability in the WDSS-II scheme but follow closely with research previously done using proximity soundings with RUC-20 data (Thompson *et al.*, 2003; Coniglio *et al.*, 2006; Cohen *et al.*, 2006). From the input variables and the known classification a training set was created. Some of the cases used to create the training set will be discussed in the following section, while the results of the training will be found in the results section.

Variable	Abbreviation	Units	Usage
Cell Size	SIZE	km ²	size discriminator
Aspect Ratio	RATIO	none	shape discriminator
Maximum Reflectivity	MAXREF	dBZ	severity potential
Mean Reflectivity	AVGREF	dBZ	development trends
Standard Deviation in Reflectivity	STDREF	dBZ	severity potential
Probability of Severe Hail	POSH	%	severity potential
Echo Top Height	ECHOZ	m	life cycle stage
Maximum Estimated Hail Size	MESH	In	severity potential
Vertically Integrated Liquid	VIL	kg m ⁻²	precipitation/hail potential
Vertically Integrated Liquid Density	VILD	g m ⁻³	precipitation/hail potential
Mean 0-3km Storm Relative Helicity	SRH	m ² s ⁻²	rotation possibility
Mean Most Unstable Convective Available Potential Energy	MUCAPE	J kg ⁻¹	updraft strength any type
Mean Surface-based Convective Available Potential Energy	SBCAPE	J kg ⁻¹	surface based energy
Mean Most Unstable Convective Inhibition	MUCIN	J kg ⁻¹	cap strength
Mean Downdraft Convective Available Potential Energy	DCAPE	J kg ⁻¹	cold pool estimator
Mean Most Unstable Vorticity Generation Parameter	AVGVGP	m s ⁻²	tilting vorticity into vertical
Maximum Most Unstable Vorticity Generation Parameter	MAXVGP	m s ⁻²	tilting vorticity into vertical
Mean Normalized Convective Available Potential Energy	NCAPE	J kg ⁻¹	convective mode (fat vs. skinny)
Mean Surface Pressure Tendency	PTEND	mbar s ⁻¹	Development
Mean 0-3km Energy Helicity Index	EHI	none	rotation possibility
Mean Relative Humidity from Surface to LCL	RH_LCL	%	Storm Inflow
Mean Deep Shear Magnitude	SHEAR	kt	storm type discriminator
Mean Wet Bulb Zero Height	WETZ	m	hail potential

Table 7.4.1. The variables used within the classification tree. The first 10 variables are radar-derived products while the rest of the parameters are RUC-derived quantities from the near storm environment (NSE) algorithm in WDSS-II.

Chapter 8 Cell Classification Cases

8.1 Sample Case Descriptions

Several cases were examined to form a training data set within this study. These cases cover different synoptic and mesoscale environments as well as contain events from different geographic locations over different seasons. It is important to note that snow and sleet convective systems are ignored in this test bed, but the possibility for future use exists. For this study the database contains individual cells within a domain for each model initialization for the specified dates. This allows for many cells to be identified within a domain for a given date at different stages in their lifetime. A brief description of some of the cases that highlight specific types of classification will follow. A complete list of sample cases and times are presented in Table 8.1.1.

To get an idea on how the classification system would handle a widespread severe weather outbreak 12 March 2006 was selected. The radar domain selected was a 256 x 256 km grid centered over the Pleasant Hill, MO radar site (KEAX). The outbreak yielded 290 storm reports in Missouri alone, with 6 deaths and almost 50 injuries reported in the Pleasant Hill forecast area. In addition to the 17 confirmed tornado touchdowns from Pleasant Hill, there were numerous reports of golf-ball to tennis-ball hail and winds in excess of 80 mph. The synoptic environment was characterized by a rapidly deepening surface low over central Kansas. The outbreak began along a dry line across

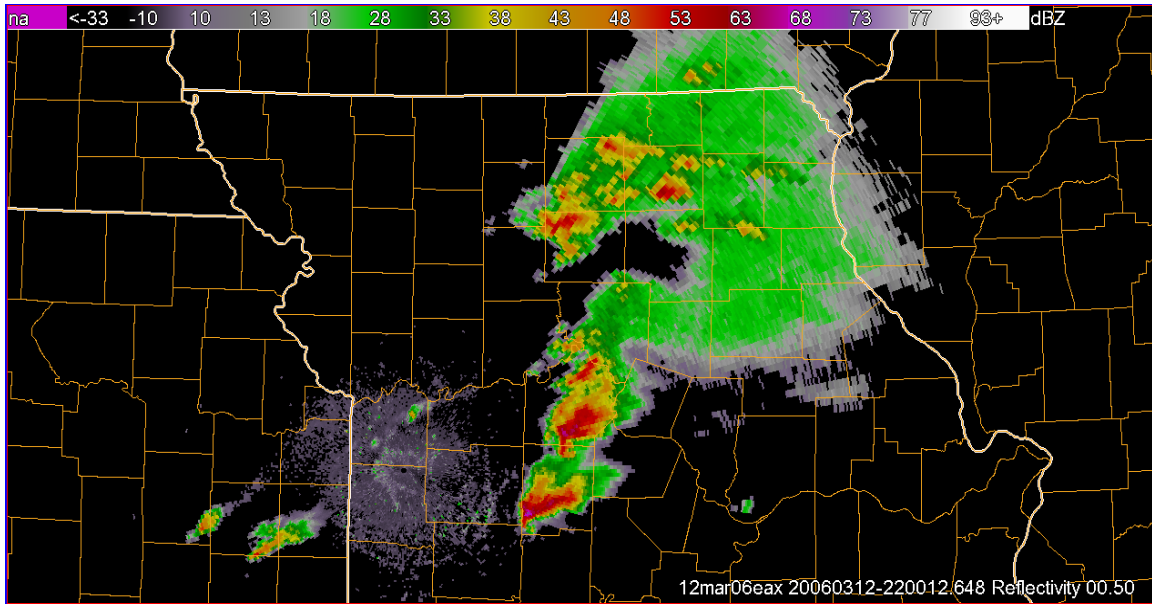


Figure 8.1.1. Base reflectivity image from 12 March 2006 from Kansas City, MO (KEAX) at 2200Z. The southernmost supercell is producing a tornado near Sedalia, MO at this time.

SE Kansas and the entire system progressed eastward keeping the bulk of the severe weather reports over Missouri. Supercells evolved quickly along the dry line bulge and moved rapidly. Radar imagery (Figure 8.1.1) showed several clusters of supercell thunderstorms across West Central Missouri. In the wake of the supercells, linear convective systems pushed through the same region later in the period with additional reports of severe wind gusts.

Another example of a sample case used in the testing of the classification scheme was an isolated supercell on 19 July 2004 in the Pendleton, OR (KPDT) radar domain. This case was unique because it took place in high elevation and, when using the lowest few elevation angles, the supercell was not immediately evident. Only in higher elevation scans or in composite mode did the isolated supercell reveal itself (Figure 8.1.2). There was one confirmed tornado touchdown and several reports of damaging winds in excess of 80 mph and

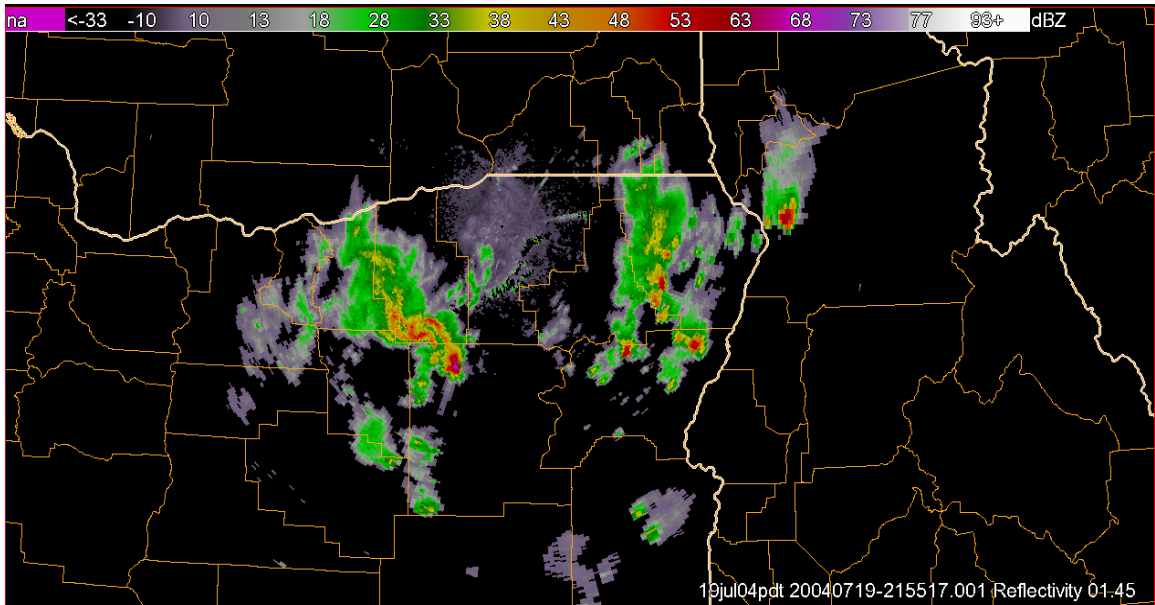


Figure 8.1.2. The Pendleton, OR supercell located south-southwest of the radar (KPDT) on 19 July 2004 at 2155Z. The 1.45 elevation angle was used due to significant beam blockage at the lowest scan.

reports of golf- ball-size hail. Heavy rain and flooding were also reported with the storms that fired up that day. Although nothing impressive was revealed on a surface analysis, there was plenty of CAPE (both surface-based and most unstable) along with favorable shear ($>0.30 \text{ ms}^{-2} \text{ VGP}$) for supercellular development.

13-14 July 2004 was a case used that actually gave a number of useful and different storm cell types from supercells to bow echoes to pulse thunderstorms. The event began as a large supercell in central Illinois (Figure 8.1.3) and transitioned into a derecho event as it moved through southern Illinois and into Kentucky and eventually to other points southward. In the wake of the derecho passage in Kentucky, several air mass-type and pulse-type thunderstorms developed and produced isolated hail reports (Figure 8.1.4). The synoptic environment began with a stationary front draped across Illinois and Indiana and points east with a deepening surface low approaching Minnesota from

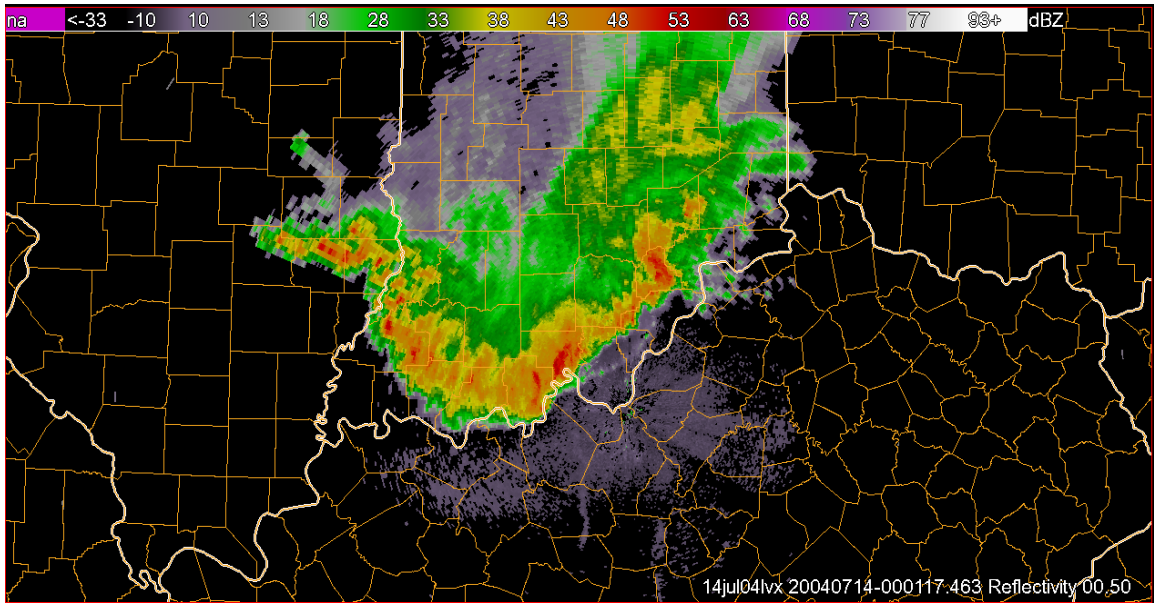


Figure 8.1.3. The beginnings of the derecho as it approaches the Louisville, KY radar (KLVX) at 0000Z on 14 July 2004.

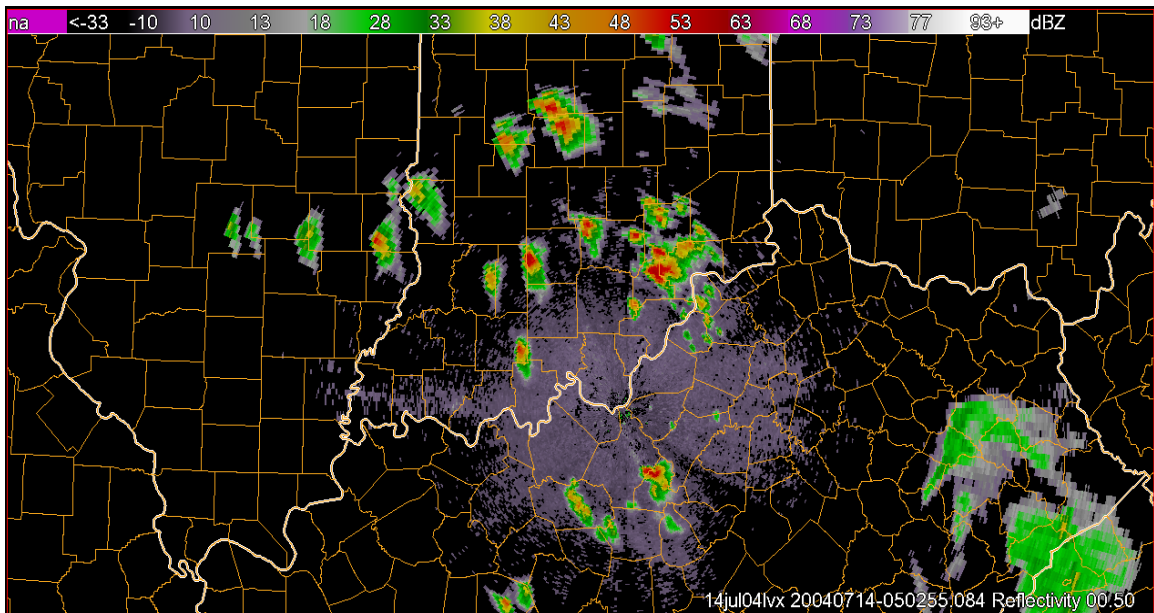


Figure 8.1.4. "Wake" pulses that formed near the KLVX radar behind the derecho as it progressed southward through Tennessee and northern Alabama from 0500Z on 14 July 2004.

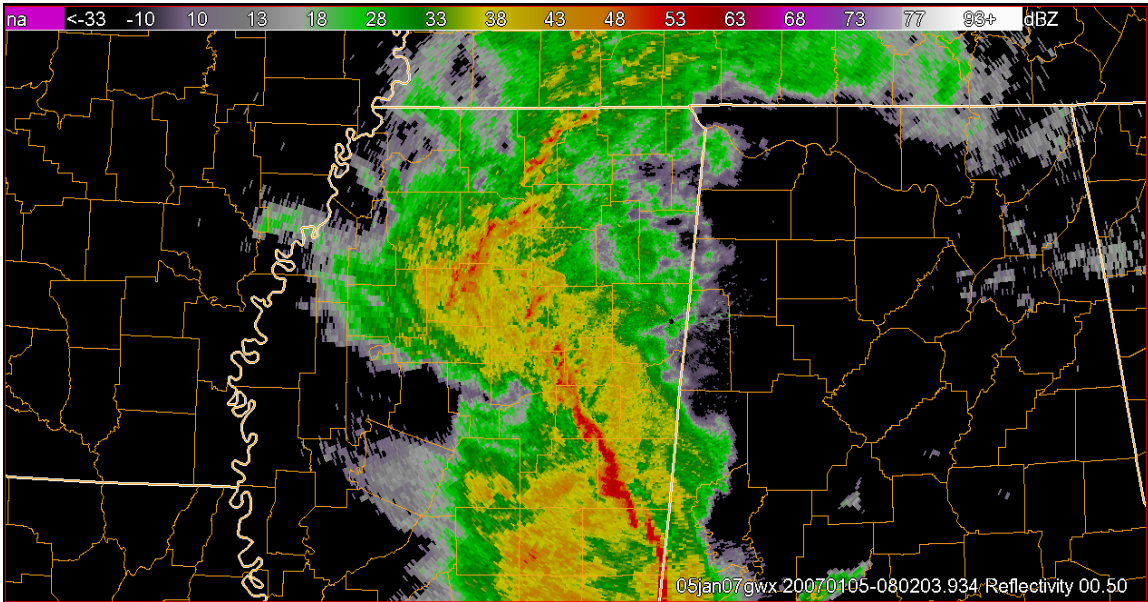


Figure 8.1.5. The severe linear convective system with embedded rotation that spawned tornadoes near the Columbus, MS (KGWX) radar at 0800Z.

central Canada. As the low deepened and occluded over Minnesota strong forcing south and east along a cold front sparked the initial supercells in central Illinois which morphed into the derecho that propagated southward.

Another type of system used in this study includes the 5 January 2007 case using the Columbus, MS radar (KGWX). Early in the period a surface cyclone developing along the Gulf Coast contributed to strong moisture convergence which allowed the formation of widespread stratiform precipitation with embedded heavier convection. As this cyclone developed and moved northeastward the dynamics favored the formation of a strong MCS in the warm sector. As this system progressed, the shear profiles began to become favorable for rotation to occur within the MCS. This caused the formation of severe squall lines with embedded rotation which sparked a few tornado reports in the Gulf States. These tornadoes occurred in the early morning hours of 5 January 2007 and a few deaths and several injuries were reported. Along with the tornado

reports the system was responsible for severe wind reports well in excess of 80 mph. Overall, this system was a good case as the situation transitioned from widespread stratiform with embedded convection to severe linear convective systems with embedded rotation. The sample reflectivity image (Figure 8.1.5) shows the messy situation which would make it hard to distinguish tornadic cells without additional velocity data.

Date	Radar Site	Abbr.	Storm Type	Date	Radar Site	Abbr.	Storm Type
12-Mar-2006	Kansas City, MO	EAX	Supercell	26-Jul-2005	Cleveland, OH	CLE	Severe Small Rotating Storm
13-Jul-2004	Lincoln, IL	ILX	Supercell	12-Mar-2006	Kansas City, MO	EAX	Small Rotating Storm
2-Feb-2007	Melbourne, FL	MLB	Supercell	13-Jul-2004	Lincoln, IL	ILX	Small Rotating Storm
5-Jun-2005	Midland, TX	MAF	Supercell	14-Jul-2004	Louisville, KY	LVX	Small Rotating Storm
6-Jun-2005	Midland, TX	MAF	Supercell	5-Jun-2005	Midland, TX	MAF	Small Rotating Storm
11-Jun-2005	Amarillo, TX	AMA	Supercell	6-Jun-2005	Midland, TX	MAF	Small Rotating Storm
12-Jun-2005	Amarillo, TX	AMA	Supercell	11-Jun-2005	Amarillo, TX	AMA	Small Rotating Storm
19-Jul-2004	Pendleton, OR	PDT	Supercell	26-Jul-2005	Cleveland, OH	CLE	Small Rotating Storm
7-Jun-2007	Kansas City, MO	EAX	Supercell	19-Jul-2004	Pendleton, OR	PDT	Small Rotating Storm
28-Nov-2005	Kansas City, MO	EAX	Supercell	6-Jun-2005	Midland, TX	MAF	Pulse
5-Jan-2007	Columbus, MS	GWX	Linear Convection with Rotation	12-Jun-2005	Amarillo, TX	AMA	Pulse
9-Jun-2005	Kansas City, MO	EAX	Linear Convection with Rotation	26-Jul-2005	Cleveland, OH	CLE	Pulse
2-Oct-2006	Chicago, IL	LOT	Linear Convection with Rotation	27-Jul-2005	Cleveland, OH	CLE	Pulse
3-Oct-2006	Chicago, IL	LOT	Linear Convection with Rotation	5-Jun-2005	Northern Indiana, IN	IWX	Pulse
28-Nov-2005	Kansas City, MO	EAX	Linear Convection with Rotation	6-Jun-2005	Northern Indiana, IN	IWX	Pulse
27-Aug-2006	Kansas City, MO	EAX	Linear Convection with Rotation	14-Jul-2007	El Paso, TX	EPZ	Pulse
2-Feb-2007	Melbourne, FL	MLB	Linear Convection with Rotation	27-Jul-2005	Cleveland, OH	CLE	Air Mass
14-Jul-2007	Louisville, KY	LVX	Linear Convection	5-Jan-2007	Columbus, MS	GWX	Air Mass
2-Feb-2007	Melbourne, FL	MLB	Linear Convection	25-Feb-2007	Kansas City, MO	EAX	Air Mass
12-Jun-2005	Amarillo, TX	AMA	Linear Convection	27-Nov-2005	Kansas City, MO	EAX	Air Mass
26-Jul-2005	Cleveland, OH	CLE	Linear Convection	27-Aug-2006	Kansas City, MO	EAX	Air Mass
27-Jul-2005	Cleveland, OH	CLE	Linear Convection	14-Jul-2007	El Paso, TX	EPZ	Air Mass
5-Jun-2005	Northern Indiana, IN	IWX	Linear Convection	27-Aug-2006	Kansas City, MO	EAX	Air Mass
6-Jun-2005	Northern Indiana, IN	IWX	Linear Convection	12-Jun-2005	Amarillo, TX	AMA	General Tstorm
12-Mar-2006	Kansas City, MO	EAX	Severe Small Rotating Storm	26-Jul-2005	Cleveland, OH	CLE	General Tstorm
13-Jul-2004	Lincoln, IL	ILX	Severe Small Rotating Storm	19-Jul-2004	Pendleton, OR	PDT	General Tstorm
14-Jul-2004	Louisville, KY	LVX	Severe Small Rotating Storm	20-Jul-2004	Pendleton, OR	PDT	General Tstorm
6-Jun-2005	Midland, TX	MAF	Severe Small Rotating Storm	14-Jul-2007	El Paso, TX	EPZ	General Tstorm
26-Jul-2005	Cleveland, OH	CLE	Severe Small Rotating Storm	5-Jan-2007	Columbus, MS	GWX	General Tstorm
19-Jul-2004	Pendleton, OR	PDT	Severe Small Rotating Storm	5-Jan-2007	Columbus, MS	GWX	General Tstorm

Table 8.1.1. Dates, locations, and cell type of the training set used for classification tree construction. Notice that several cases of the same date and same location have different storm types. These are due to the nature of development and morphology of storms over time for the given domain.

Chapter 9 Cell Classification Results

The results section will include a brief example on the cell identification portion of the code which runs prior to the classification tree algorithm in order to associate storm cells with environmental variables. Descriptions of different cell classification experiments will follow including an introduction to the skill of the cell classifier using different setups. Training data sets along with test cases will be examined.

9.1 Cell Identification

Cell identification is the first step in the process of the classification process. Identification of convective elements is done using a 30-dBZ threshold in reflectivity and a minimum size criterion of 40 km². The 30-dBZ criterion was used to keep a fair proportion of the stratiform area especially needed for embedded convection within a widespread stratiform region. The threshold in reflectivity and minimum size criterion may be altered for different applications of the software, especially when one wants to discriminate against the smaller convective cells.

The example for the cell classification was the severe weather outbreak over Missouri on 12 March 2006. The time period selected for this example was 2200Z (Figure 8.1.1). The domain for the radar image was centered over the KEAX radar

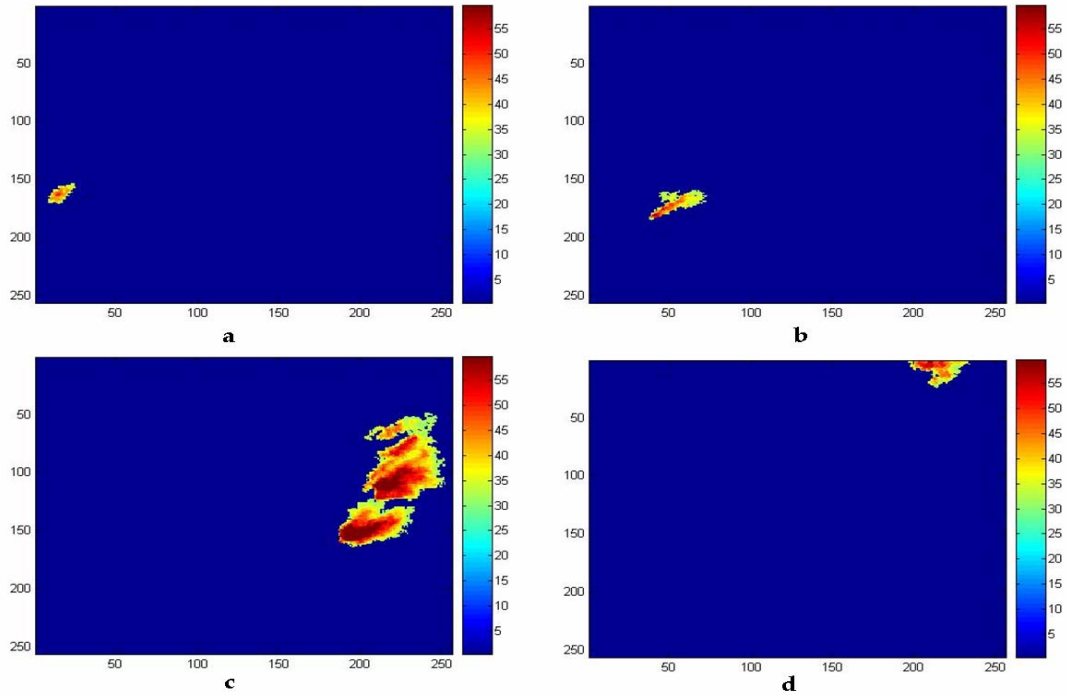


Figure 9.1.1. The identified cells from 12 March 2006 composite reflectivity centered over the KEAX radar site at 2200Z. Four distinct objects were identified in this particular case with each cell labeled a-d.

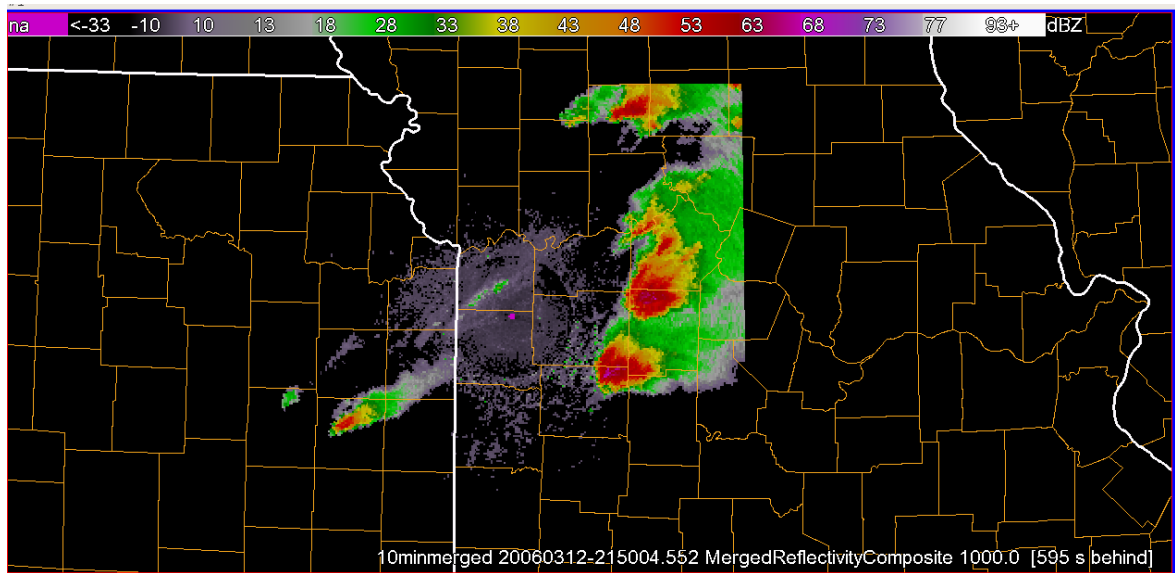


Figure 9.1.2. The maximum composite reflectivity image from 12 March 2006 centered over the KEAX radar site at 2200Z.

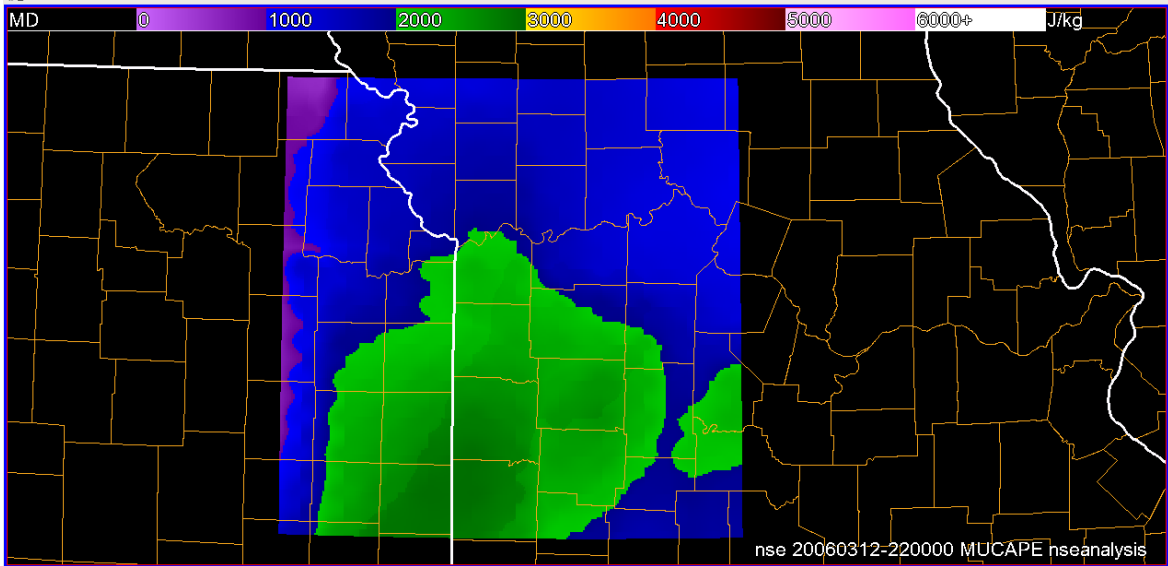


Figure 9.1.3. The MUCAPE field from 12 March 2006 centered over the KEAX radar site at 2200Z. At this time, MUCAPE is weakening to values less than 3000 Jkg^{-1} .

site and extends over a 256 x 256 km area. The radar image was created using a maximum composite scheme and then run through the cell identification scheme. Figure 9.1.1 shows the identification of the different cells in the domain after they have been isolated at the 30-dBZ threshold and have passed the minimum size criterion. The composite reflectivity image for the domain can be viewed for comparison in Figure 9.1.2. In Figure 9.1.1c, multiple supercells are identified as one cluster of supercells; this occurs as there is sufficient overlap of the 30-dBZ regions. As long as a part of the cluster contains supercell characteristics the entire identified cell will be considered a supercell.

Once the individual cells are tagged with identifying numbers, the program calculates and overlays reflectivity derived parameters along with selected NSE variables derived from the RUC-20 via the WDSS-II software. Figure 9.1.3 shows an example of the most unstable CAPE (MUCAPE) field that was overlaid on the composite reflectivity field, while Figure 9.1.4 shows the MUVGP and Figure 9.1.5

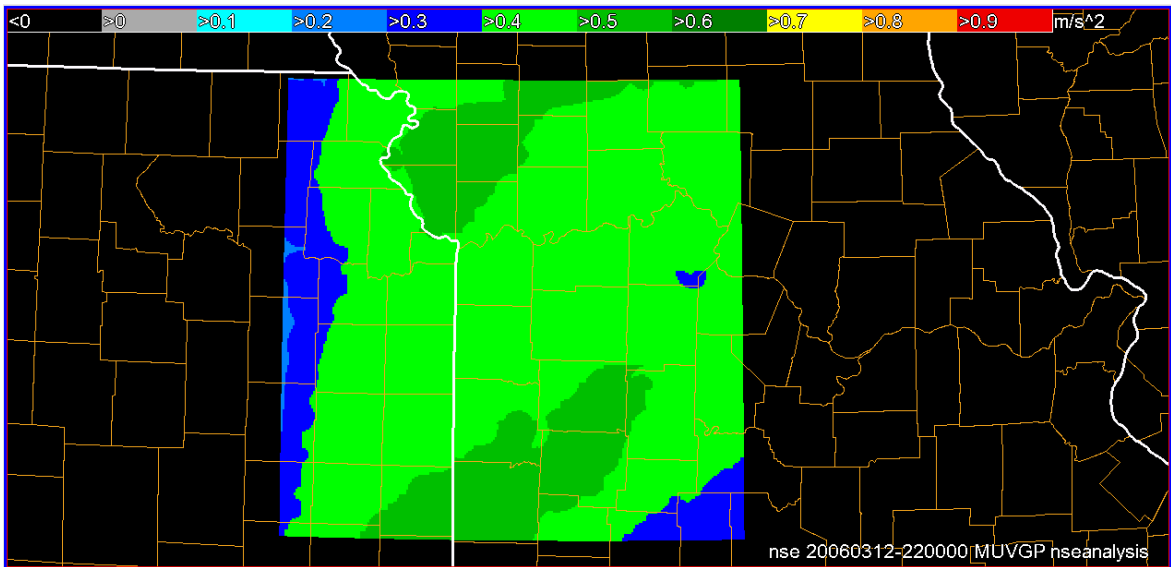


Figure 9.1.4. The MUVGP field from 12 March 2006 centered over the KEAX radar site at 2200Z. Notice values exceeding 0.3 ms^{-2} throughout the domain which signifies the continuing threat for rotating storms.

shows the SRH in the 0 to 3-km layer. These figures illustrate the variability in a small domain ($256 \times 256 \text{ km}$) at a 20-km resolution. Significant changes in these variables over such short distances can alter the evolution of the individually tagged storm from Figure 9.1.1. For each time, a table of storm attributes is created including the reflectivity-derived and the NSE variables which will be input into the classification tree. Table 9.1.1 includes some of the selected fields used for classification for the 12 March 2006 example from Figure 9.1.1. Accompanying the selected variables is the subjective “expert” classification that will fall into one of the eight classifications from Table 7.4.1 for input into the classification tree routine.

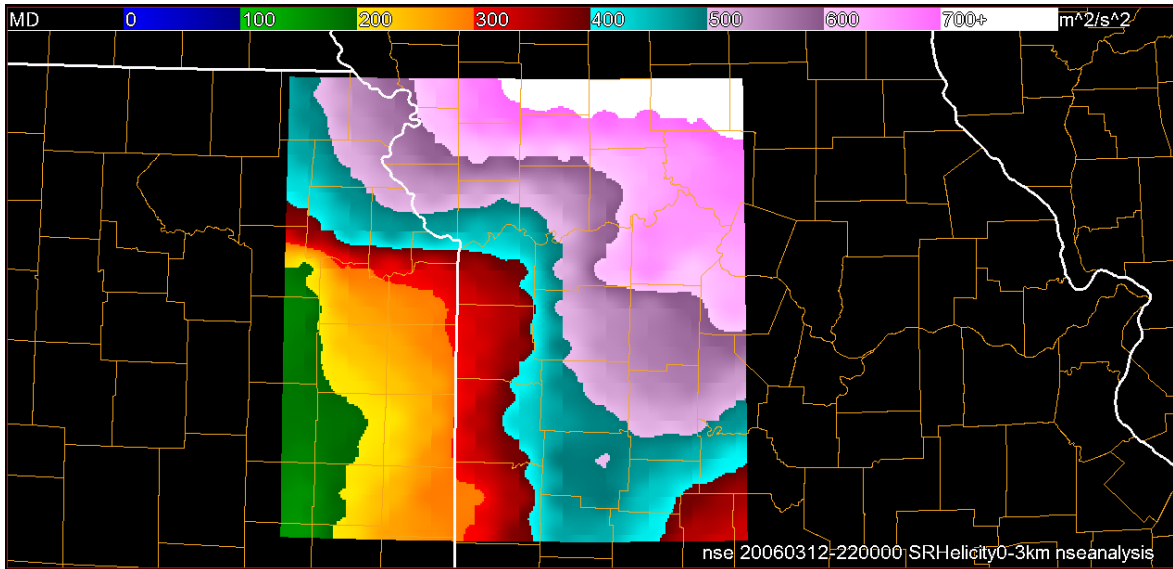


Figure 9.15. The MUVGP field from 12 March 2006 centered over the KEAX radar site at 2200Z. Notice values exceeding 0.3 ms^{-2} throughout the domain which signifies the continuing threat for rotating storms.

12 March 2006 at 2200Z Cell Array				
SIZE	3641	459	403	162
MAXREF	73.704	57.000	58.869	51.000
STDREF	10.058	7.039	6.359	4.938
ECHOZ	50081	41409	42962	36631
VIL	67.173	26.577	15.656	7.695
SRH	613.64	740.08	226.56	191.37
MUCAPE	1516.9	1322.8	2205.2	1327.9
DCAPE	800.73	11.17	825.69	746.17
AVGVGP	0.435	0.494	0.448	0.382
EH1	5.743	6.118	3.122	1.594
SHEAR	44.327	44.920	37.329	41.165
RH_LCL	78.634	97.904	70.539	58.252
RATIO	2.833	1.744	2.553	2.331
MU/SBCAPE	1.993	1322.800	1.636	1.143
ROT	TVSMeso	Circ	Circ	Circ
"EXPERT"	Supercell	Dying	Growing Rotating	Growing Rotating
TYPE	Cluster	Supercell	Storm	Storm

Table 9.1.1. Selected output fields for the 12 March 2006 case at 2200 including the "expert" classification, for units and category descriptions see Table 7.4.1.

9.2 Classification Using Radar Only

The classification tree used for this study first tests the hypothesis from MacKeen *et al.* (1999) that using reflectivity data alone does not provide the necessary information for predicting storm longevity and evolution. The tree is created using information on reflectivity-derived products and shape information. Reflectivity-derived products include: maximum reflectivity, mean reflectivity, the standard deviation of reflectivity, vertically integrated liquid water (VIL), VIL density, probability of severe hail (POSH), maximum estimated hail size (MESH). The shape parameters include the echo top height, the aspect ratio of the storm, and the storm size. The aspect ratio is found simply by dividing the major axis of the storm by the minor axis determined by rotating the storm around its geographic centroid. The test cases outlined in Table 8.1.1 were used to create the initial tree. Equal populations, 16 of each storm type, were used in the creation process. Figure 9.2.1 shows the classification tree using reflectivity data only. Figure 9.2.2 shows the misclassification costs which corresponds the number of terminal nodes created by the tree. For the reflectivity only tree it is noted that the best level to prune the tree corresponds to where the lowest cross-validation error occurs, in this case at level 5 where the cross-validation error is at 0.4833. It is interesting to note that using linear classification yields a cost of 1.0333. Clearly, using reflectivity only to create a decision tree fails when using linear classification methods.

Upon tree creation, estimated probabilities of the given storm type can be viewed at each node. From the variable ranges given, a series of “if then”

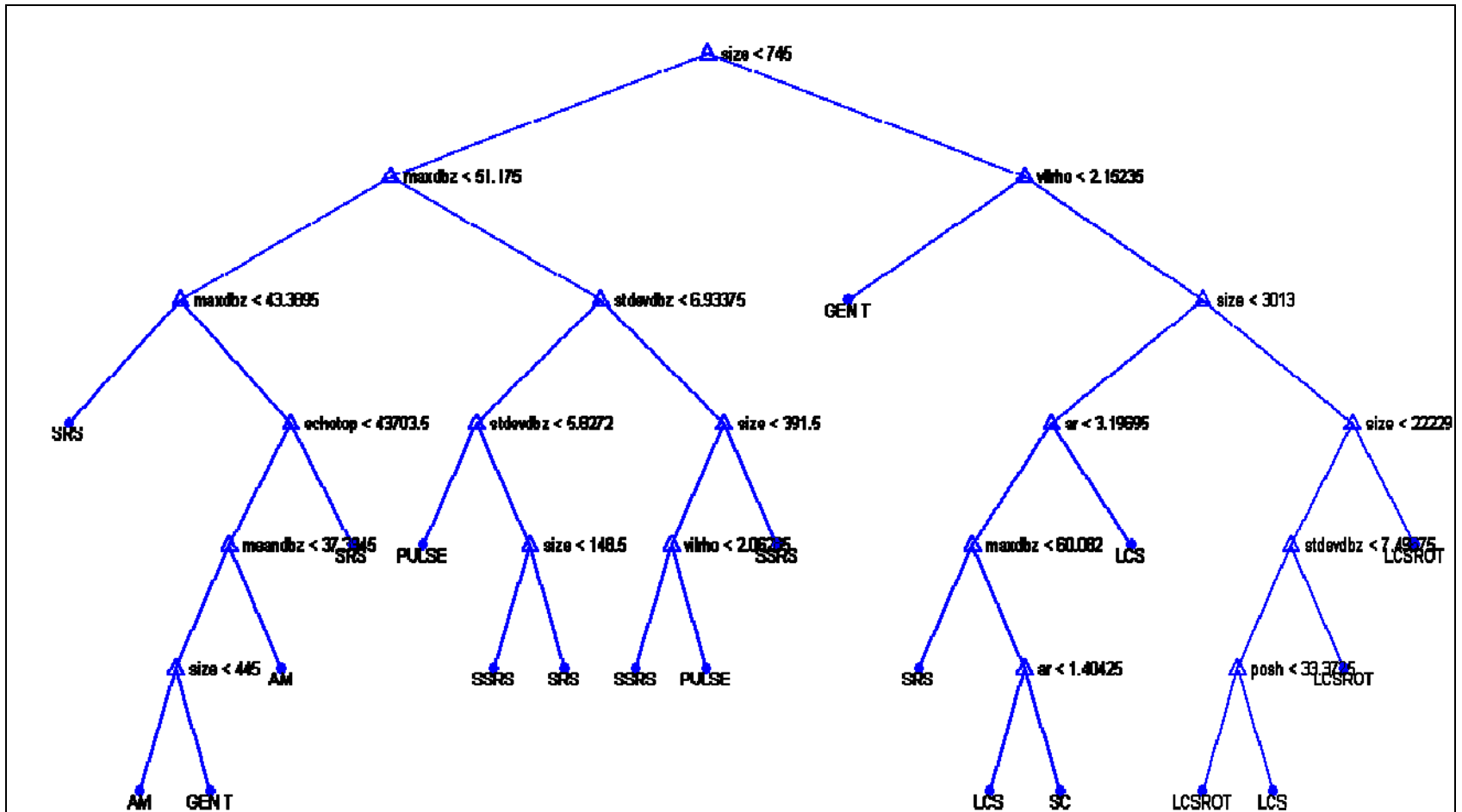


Figure 9.2.1. The cell classification tree using reflectivity only.

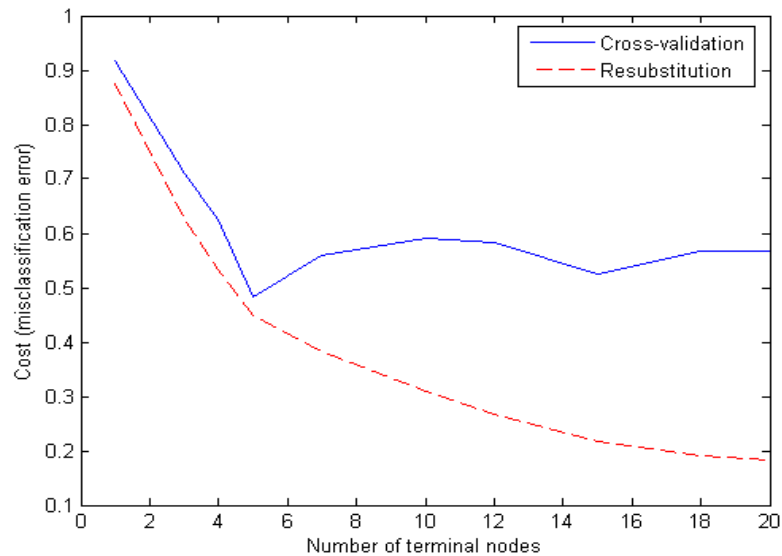


Figure 9.2.2. The misclassification error (cost) for the classification tree using reflectivity-derived products only. The resubstitution error is an idealized cost where the cross-validation curve shows the actual cost for the different terminal nodes.

statements can be used to yield a probability of a certain storm type given reflectivity data only. The user may prune the tree as desired to reach a level at which the probabilities give a general solution. The user may also use the end-nodes as they indicate a deterministic solution without estimated probabilities. For the end-node that results in the supercell classification, the size range was 745 to 3013 km², the mean VIL density was greater than 2.15 g m⁻³, the maximum reflectivity was greater than 60.08 dBZ, and the aspect ratio was less than 3.2. The aspect ratio being less than 3.2 reveals that the storm should appear more circular in nature which agrees with the supercell structure. The intense reflectivity and size range also coincide with most supercell observations. From the supercell end-node, estimated probabilities of the eight storm types are given and can be incorporated into an automated procedure (neural network) for classification. An

example of the estimated probabilities for the supercell case given reflectivity only can be seen in Table 9.2.1.

Upon further examination of the classification tree using reflectivity-derived products only, it is apparent that there are a few problems with the logic used to create a few of the branches in the testing phase. The main issue stems from the frequency of using maximum reflectivity for splitting, an example is found at the first two nodes on the left side of the tree in Figure 9.2.1. The first split using a maximum reflectivity of 51.175 dBZ should result in a separation of severe and non-severe events; however, an examination of the final nodes shows that non-severe storms are still classified with size becoming the main discriminating factor. Distinguishing between small storms using storm size does not make intuitive sense as the smaller scale storms should have approximately the same size. In addition to the overuse of maximum reflectivity there seems to be improper splitting at the end of the tree in the discrimination between linear convective systems and supercells. The aspect ratio is used in this case and the classification results in a supercell if the ratio is greater than 1.40425. Clearly a linear convective system should have a greater aspect ratio and some considerable misclassification is occurring. Finally, the tree does not seem to have an ideal pruning level; the eight classification types do not appear uniquely at one pruning level as is shown when combining RUC-20 data and radar-derived fields.

Once the training phase is complete using equal populations and reflectivity-derived fields only, additional cases are put into the scheme and the created tree classifies the new and complete data set using all storms. At this point the quality of the tree can be assessed. Information from other identified cells from various dates and geographic locations may be used with the same reflectivity-derived data set used as input to the classifier. Once the new data set is run through the

Supercell Node Estimations Using Reflectivity Only	
Storm Type	Prob
Supercell	0.8
Linear Convective System	0
Linear Convective System with Rotation	0.0667
Small Rotating Storm	0
Severe Small Rotating Storm	0
Pulse	0.1333
Air Mass	0
General Thunderstorm	0

Table 9.2.1. Estimated probabilities at the unique supercell end-node from the classification tree (Figure 9.2.1) using reflectivity-derived parameters only.

Standard Skill Scores (Exact Match)			
Storm Type	POD	FAR	CSI
Supercell	0.317	0.412	0.260
Linear Convective System	0.370	0.706	0.196
Linear Convective System with Rotation	0.321	0.591	0.220
Small Rotating Storm	0.240	0.765	0.135
Severe Small Rotating Storm	0.389	0.548	0.264
Pulse	0.385	0.753	0.177
Air Mass	0.600	0.580	0.328
General Thunderstorm	0.550	0.333	0.431

Table 9.2.2. Skill scores for exact matches for the classification tree using reflectivity-derived parameters only.

scheme, the deterministic classification from the test tree is compared with the expert classification. The first comparison is an exact-match approach where the deterministic classification from the tree must match the expert classification exactly. The second comparison made allows for adjustments to the match as long as the match is deemed to be not a total miss in classification. For example, for an event to be classified by the tree as a supercell but the expert classification reveals

Standard Skill Scores (Near Match)			
Storm Type	POD	FAR	CSI
Supercell	0.571	0.163	0.514
Linear Convective System	0.556	0.516	0.349
Linear Convective System with Rotation	0.821	0.042	0.793
Small Rotating Storm	0.300	0.674	0.185
Severe Small Rotating Storm	0.542	0.339	0.424
Pulse	0.385	0.726	0.190
Air Mass	0.857	0.483	0.476
General Thunderstorm	0.550	0.333	0.431

Table 9.2.3. Skill scores for near matches for the classification tree using reflectivity-derived parameters only.

a severe, small-scale, rotating storm, the case is flagged as a match instead of a non-match as the classification is close to the truth and would not prove to be a problem operationally. However, for the case of a linear convective system with rotation being matched to a pulse thunderstorm, this would still be considered a mismatch. In the case of using reflectivity data only to create the classification tree, the exact classification rate from 365 cells including the training cells was 0.417, while the marginal classification rate of those same cells was 0.618. Table 9.2.2 shows a breakdown of skill scores including POD, FAR, and CSI for the eight classification types for exact matches, while Table 9.2.3 shows skill scores for near matches. As a general conclusion, the classification rates of success are not as high as desired when using reflectivity-derived products only; this coincides with the observations made by MacKeen *et al.* (1999).

A contingency table for each of the eight storm classifications for exact matches using reflectivity-derived products only is shown in Table 9.2.4. The table was created using 365 expert classification observations and the resulting classification from the decision tree. For optimal performance values should be

	SC	SSRS	SRS	LCS	LCSROT	AM	PULSE	GEN T	Total Obs
SC	20	10	2	14	8	0	5	3	62
SSRS	3	28	8	0	0	1	29	2	71
SRS	0	3	12	0	0	26	9	0	50
LCS	4	0	4	10	5	0	1	3	27
LCSROT	5	3	1	8	9	0	1	1	28
AM	0	2	3	0	0	21	9	0	35
PULSE	2	16	10	1	0	2	20	1	52
GEN T	0	0	11	1	0	0	6	22	40
Total Class	34	62	51	34	22	50	80	32	365

Table 9.2.4. Contingency table for the expert classified observation cell types (x-axis) and the output from the reflectivity-derived products only decision tree. The table was created for 365 total cells. Refer to Table 7.3.1 for abbreviations.

maximized along the diagonal of the table. Outliers are readily apparent when examining the table, especially when examining the severe small-scale rotating storm (SSRS) and the non-severe small-scale rotating storm (SRS). These outliers are mirrored in the pulse and air mass storm classifications as well. From the table it is also clear which storm types are being over- or under-classified. For example, there are a total of 62 observed supercells and only 34 are accurately classified for the reflectivity-derived product only tree. Overall, it is evident that distinguishing between larger-scale storms using reflectivity products alone is a weakness.

9.3 Classification Using Radar and RUC-20 Data

The advantage of the tree classification scheme is that large datasets may be used to construct the ideal-sized (pruned) tree. There is not an immediate

necessity to select certain RUC-20 fields for use in the classification tree formulation. The scheme itself will select the best splits at each node given the large data array by the methods described in the previous sections. This allows for easy comparison between previous studies on near-storm environmental fields for certain convective storms and the fields identified for use in the classification tree scheme. The classification tree created using fields created by the WDSS-II NSE algorithm using RUC-20 data should provide better results than using reflectivity only. The tree created when using the combination of radar reflectivity-derived fields along with RUC-20 NSE fields is shown pruned to the eight classification types in Figure 9.3.1. The misclassification errors (costs) at each terminal node are shown in Figure 9.3.2. The cross-validation cost of the tree at the pruned level shown in Figure 9.3.1 is at a minimum (0.275). If doing a linear classification the cost drops to around 0.125; however, more work would need to be done to examine the full effectiveness of a linear classification tool in this seemingly nonlinear application. The supercell classification end-node shown in Figure 9.3.1 reveals that the supercell classification relies on a mean VGP to be greater than 0.299 ms^{-2} and a size value between 745 and 3734 km^2 . The VGP value is consistent with the findings by Thompson *et al.* (2003) for supercells utilizing RUC proximity soundings. It is interesting to note that for the classification tree using RUC data and reflectivity-derived products, the main split utilizes the VGP and separates the tree into the four non-rotating storm types and the four rotating storm types. Physically this is a logical first split. The addition of some kind of model-derived rotation quantity alone improves the formulation of the classification tree(s). Once again, the supercell estimated probabilities at the end-node given in Figure 9.3.1 are shown in Table 9.3.1.

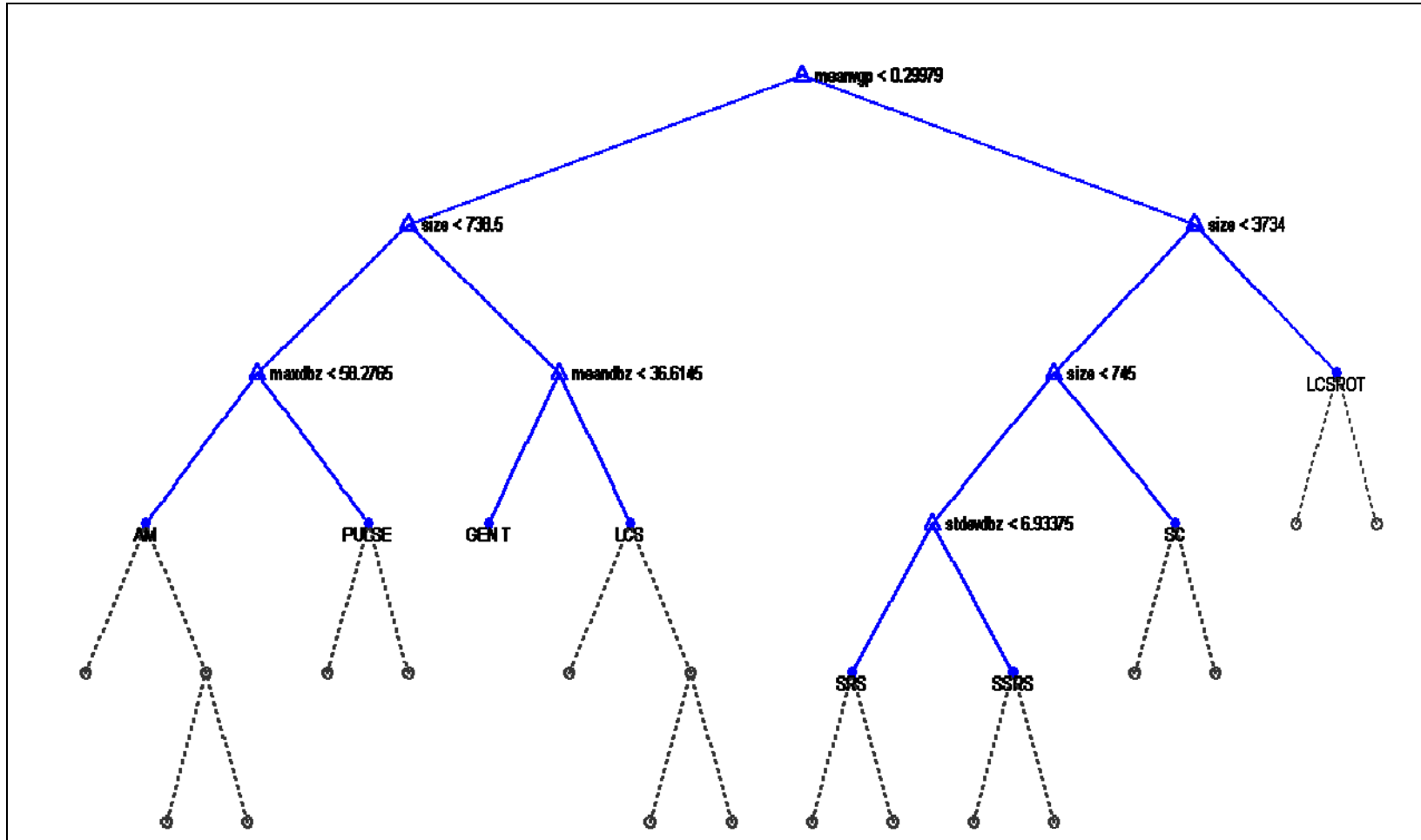


Figure 9.3.1. Classification tree for the RUC-20 fields combined with reflectivity data calculated from the WDSS-II system. The tree is pruned at level 2 of 8 to capture the eight classification types at the end-nodes.

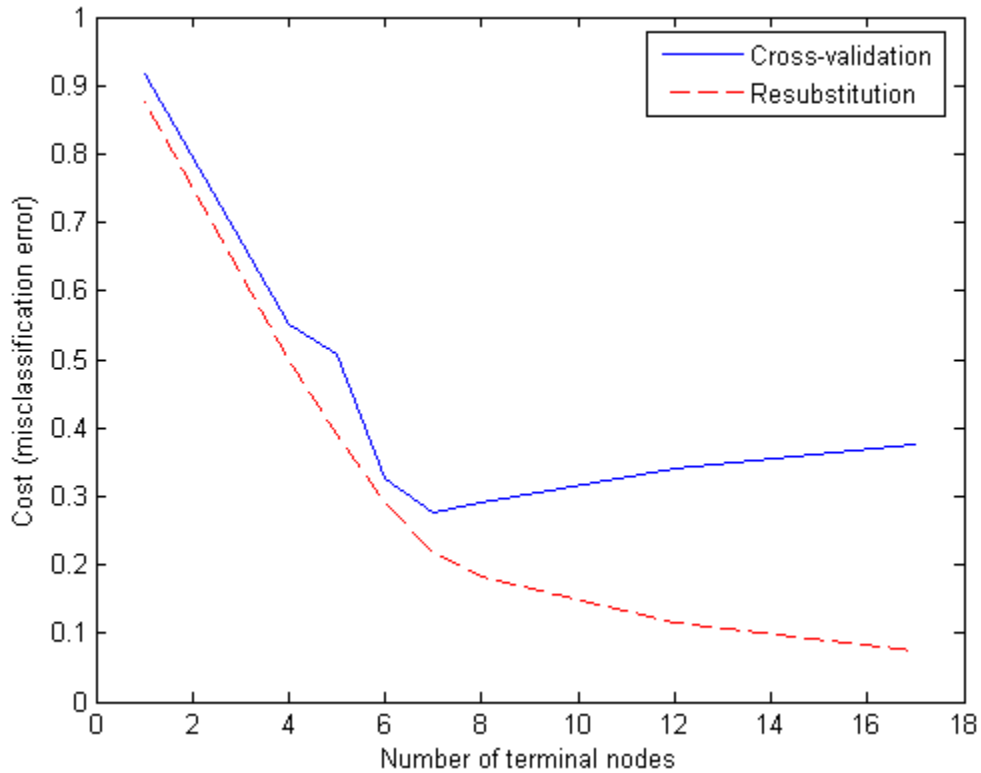


Figure 9.3.2. The misclassification error (cost) for the classification tree using an even population with reflectivity-derived products combined with RUC-20 NSE fields. The resubstitution error is an idealized cost where the cross-validation curve shows the actual cost for the different terminal nodes.

Supercell Node Estimations Using RUC and Reflectivity	
Storm Type	Prob
Supercell	0.765
Linear Convective System	0
Linear Convective System with Rotation	0
Small Rotating Storm	0.0588
Severe Small Rotating Storm	0
Pulse	0
Air Mass	0
General Thunderstorm	0.176

Table 9.3.1. Estimated probabilities at the pruned supercell end-node from the classification tree using reflectivity-derived parameters combined with RUC-20 NSE fields.

Standard Skill Scores (Exact Match)			
Storm Type	POD	FAR	CSI
Supercell	0.556	0.300	0.449
Linear Convective System	0.778	0.447	0.477
Linear Convective System with Rotation	0.071	0.600	0.065
Small Rotating Storm	0.720	0.478	0.434
Severe Small Rotating Storm	0.333	0.467	0.258
Pulse	0.269	0.600	0.192
Air Mass	0.400	0.754	0.179
General Thunderstorm	0.750	0.565	0.380

Table 9.3.2. Skill scores for exact matches for the classification tree using and even population of reflectivity-derived parameters combined with near storm environmental data from the RUC-20 model.

Standard Skill Scores (Near Match)			
Storm Type	POD	FAR	CSI
Supercell	0.841	0.000	0.841
Linear Convective System	0.852	0.233	0.676
Linear Convective System with Rotation	0.821	0.000	0.821
Small Rotating Storm	0.740	0.260	0.587
Severe Small Rotating Storm	0.653	0.060	0.627
Pulse	0.635	0.340	0.478
Air Mass	0.886	0.380	0.574
General Thunderstorm	0.850	0.433	0.515

Table 9.3.3. Skill scores for near matches for the classification tree using and even population of reflectivity-derived parameters combined with near storm environmental data from the RUC-20 model.

From the classification tree using reflectivity and RUC-20 model data fields, the success of classification can be examined. From the exact comparison, the tree using RUC-20 fields with the radar-derived fields yields a classification success rate of 0.50. When accounting for near matches this figure jumps to 0.78. When

examining standard skill scores for the eight different types of storms for exact matches (Table 9.3.2) and for near matches (Table 9.3.3) an improvement is evident in comparison with the reflectivity-derived parameters only classification tree. It is important to keep in mind that in the success rate comparison, deterministic classifications are used. If the estimated probabilities, or misclassification rates, were used instead of the deterministic solution more insight into the success of the classification scheme could be attained. A Brier score may be useful in determining the skill in classification when estimated probabilities are considered. The problem with using the estimated probabilities is that they are locked once the tree from the training set is created. This may limit the effectiveness of picking up on hybrid cases as they are often avoided during the tree creation process for purity.

When examining where problems exist within the classification scheme, it becomes evident that the small-scale convective cells and some of the linear type of storms are frequently misclassified. Although upon examination of the small-scale systems the node splits are consistent with observations. For example, the standard deviation of reflectivity is used to discriminate between severe and non-severe storm types. The standard deviation is a way of examining maximum reflectivity gradients in storms. In the case of small-scale severe storms this gradient should be large when viewing composite reflectivity as this would represent a nearly vertical updraft supporting large hail. The problem in the small-scale storms may fall into the expert classification of these systems. Comparisons of storms with hail reports and storms of similar structure with no hail reports may need to be made for an accurate subjective classification. Objects that are similar in overall structure but do not produce observable severe weather still need to be grouped in the same category. A contingency table for each of the eight storm classifications for exact matches is shown in Table 9.3.4, highlighting

	SC	SSRS	SRS	LCS	LCSROT	AM	PULSE	GEN T	Total Obs
SC	34	17	2	3	1	0	0	5	62
SSRS	3	24	19	0	0	7	14	4	71
SRS	0	1	36	0	0	11	1	1	50
LCS	0	0	0	21	2	1	2	1	27
LCSROT	11	0	0	10	2	0	0	5	28
AM	0	1	3	0	0	14	4	13	35
PULSE	0	2	4	2	0	20	14	10	52
GEN T	0	0	4	2	0	4	0	30	40
Total Class	48	45	68	38	5	57	35	69	365

Table 9.3.4. Contingency table for the expert classified observation cell types (x-axis) and the output from the even population RUC-20 model fields combined with reflectivity-derived products decision tree. The table was created for 365 total cells. Refer to Table 7.3.1 for abbreviations.

some of the strengths and weaknesses of using the combination of RUC-20 data with the radar reflectivity-derived products. Overall, there is vast improvement in discriminating between rotating and non-rotating types with the addition of RUC-20 data which is of primary importance; however, work still needs to be done on the smaller-scale storms. This issue will be addressed in the following sections.

9.4 Classification Using Uneven Populations

A final classification tree was created to test the issue of using uneven populations in the training set. For this analysis, the clearest example storm types for each of the eight classifications were used. This resulted in a much smaller training set which may limit the effectiveness of the classification. For this training set, all reflectivity-derived products and RUC-20 fields were used. The resulting classification tree for this study is shown in Figure 9.4.1. The tree is notably different from the one created using equal populations in the training set

as was done in section 9.3. The primary differences are found in the number of branches created by the uneven population tree, as well as rotation not being used at the primary split. The uneven population tree has four pruning levels which lead to a more efficient tree; however, some detail may be lost.

The resulting tree in Figure 9.4.1, although different from the one shown in 9.3.1, has a better correct classification rate than just using reflectivity-derived products alone and is close in skill to using model- and radar-derived products. Skill scores from the eight classification types can be seen for exact matches (Table 9.4.1) and near matches (Table 9.4.2). It is readily apparent that the uneven population tree outperforms the others in terms of matching the small-scale convective storm classification types to the expert classification. A contingency table (Table 9.4.3) for the uneven population tree further reinforces the skill score tables. The contingency table highlights the lack of confidence in the prediction of linear convective systems with rotation (LCSROT).

Upon examining the supercell classification node, it is shown that the VGP was not used in making this classification. Instead, values of downdraft CAPE (DCAPE) and surface-based CAPE along with size are used as supercell discriminators. Although CAPE values are good in determining potential updraft strengths or cold pool strengths (DCAPE) it says nothing about the orientation or possible rotation that could be occurring. Supercells and linear convective systems may both form in similar CAPE environments but the discriminating factor is often the type and amount of shear present. The large-scale storm classifications found on the right-hand side of the tree do not follow physical logic for rotating versus non-rotating storms. A shear measure is not used to distinguish between rotating and non-rotating storms and this is evident in the lower skill scores for the supercell and linear convective system categories when

Standard Skill Scores (Exact Match)			
Storm Type	POD	FAR	CSI
Supercell	0.333	0.625	0.214
Linear Convective System	0.481	0.658	0.250
Linear Convective System with Rotation	0.143	0.765	0.098
Small Rotating Storm	0.600	0.318	0.469
Severe Small Rotating Storm	0.556	0.535	0.339
Pulse	0.750	0.235	0.609
Air Mass	0.571	0.429	0.400
General Thunderstorm	0.550	0.450	0.379

Table 9.4.1. Skill scores for near matches for the classification tree using an uneven population of reflectivity-derived parameters combined with near storm environmental data from the RUC-20 model.

compared against the resulting tree using even populations and RUC-20 model data.

The tree created by uneven population seems to classify the smaller-scale storms correctly with more frequency than the one using even populations. Most likely, this stems from the size criteria being used to make the first split instead of rotation. Once the tree determines that it is a small-scale storm, mean VGP is used to separate the rotating and non-rotating storm with a value of greater than 0.299 ms^{-2} , which is again consistent with other studies. Also, the standard deviation in reflectivity and maximum reflectivity are logically used to distinguish between severe and non-severe storms. It may be noted that the maximum reflectivity of 51.4715 dBZ is comparable to 53 dBZ which has been used for distinguishing hail for estimations of rainfall from radar reflectivity in previous research.

As the tree only has four pruning levels and noticeably fewer training cells are used in its creation, it may be deemed that a larger population and vastly different population numbers of each storm type should be used to accurately determine the differences from even to uneven population classification trees in

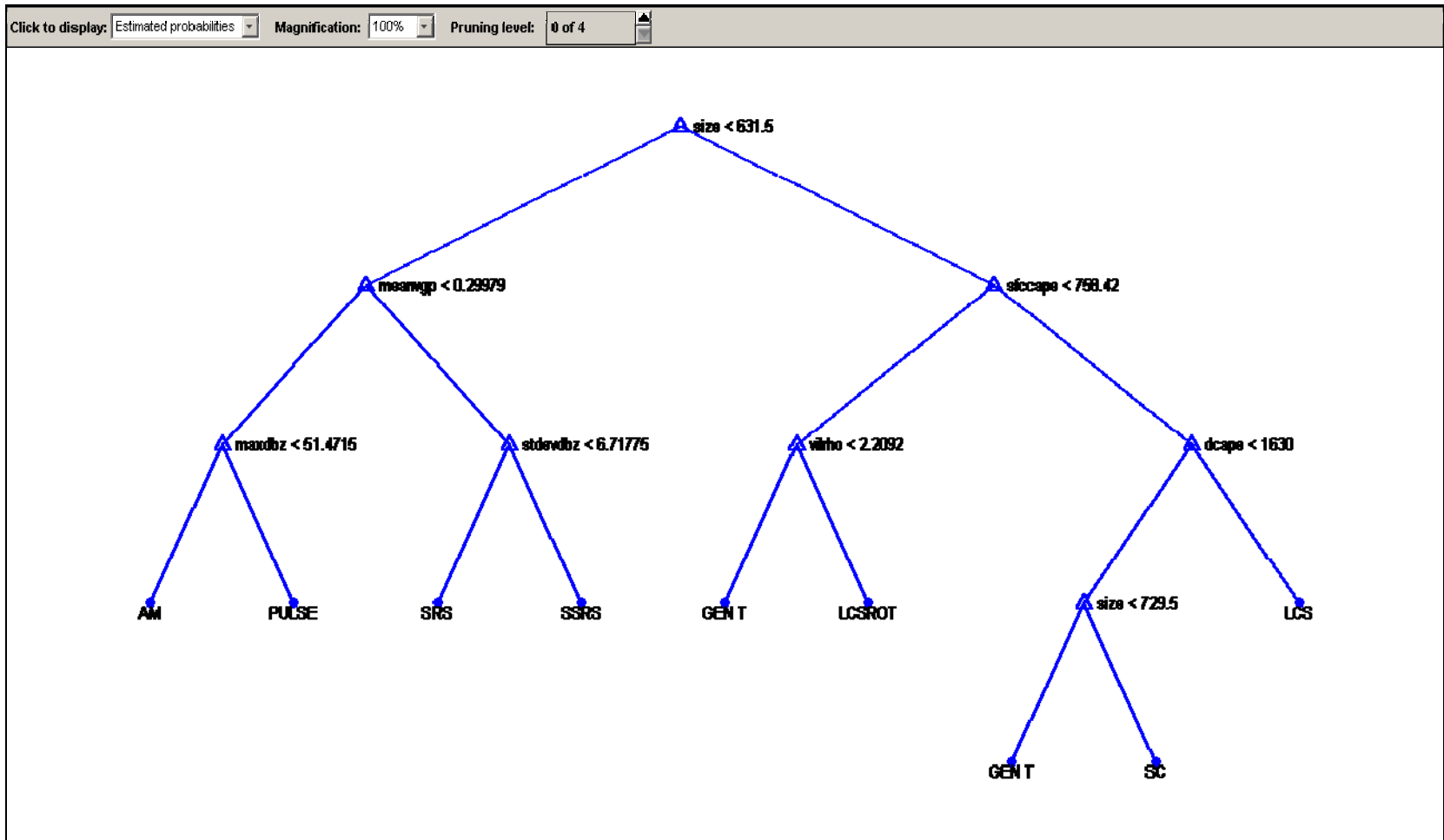


Figure 9.4.1. Classification tree for the uneven populations of storm types using RUC-20 fields calculated from the WDSS-II NSE algorithm and reflectivity-derived fields.

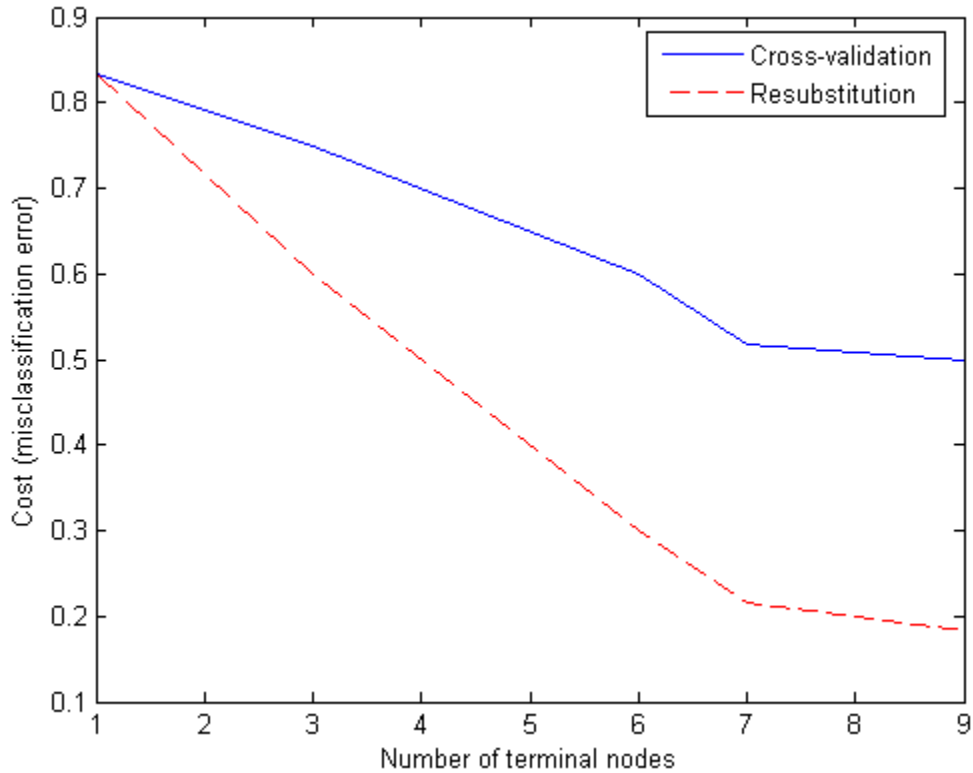


Figure 9.4.2. The misclassification error (cost) for the classification tree using an uneven population with reflectivity-derived products combined with RUC-20 NSE fields. The resubstitution error is an idealized cost where the cross-validation curve shows the actual cost for the different terminal nodes.

Standard Skill Scores (Near Match)				
Storm Type	POD	FAR	CSI	
Supercell	0.556	0.146	0.507	
Linear Convective System	0.815	0.421	0.512	
Linear Convective System with Rotation	0.857	0.040	0.828	
Small Rotating Storm	0.800	0.245	0.635	
Severe Small Rotating Storm	0.903	0.286	0.663	
Pulse	0.788	0.226	0.641	
Air Mass	0.886	0.139	0.775	
General Thunderstorm	0.550	0.267	0.458	

Table 9.4.2. Skill scores for near matches for the classification tree using an even population of reflectivity-derived parameters combined with near storm environmental data from the RUC-20 model.

	SC	SSRS	SRS	LCS	LCSROT	AM	PULSE	GEN T	Total Obs
SC	28	3	0	11	11	0	2	0	55
SSRS	17	26	1	0	0	1	2	0	47
SRS	1	16	36	0	0	4	4	4	65
LCS	1	1	0	2	1	0	0	0	5
LCSROT	11	4	0	9	15	0	2	1	42
AM	0	1	10	0	0	20	1	10	42
PULSE	0	19	3	1	0	10	40	3	76
GEN T	4	1	0	4	1	0	1	22	33
Total Class	62	71	50	27	28	35	52	40	365

Table 9.4.3. Contingency table for the expert classified observation cell types (x-axis) and the output from the uneven population RUC-20 model fields combined with reflectivity-derived products decision tree. The table was created for 365 total cells. Refer to Table 7.3.1 for abbreviations.

future work. The cost of the tree at the different terminal nodes shows a relatively high misclassification error even out to the higher number of terminal nodes (Figure 9.4.2). The minimum cross-validation error of 0.5167 illustrates great uncertainty when using a decrease in the number of training cells with uneven populations. A linear classification scheme yields a cost of 0.1167, which is actually the best out of the three experiments herein. Again, more work would need to be done to examine the appropriateness of a linear classification.

Chapter 10 Cell Classification Conclusions

10.1 Classification Conclusions

The use of classification trees to the storm cell identification problem is illustrated in the above sections. The primary goal of cell identification and classification is to be able to use the information as input into a reflectivity-based nowcast model for growth, decay, and morphology issues. Classification may also be used for real-time monitoring of significant storms by forecasters. It may also be used for verification purposes in creating databases for storm types or for the matching problem in object-oriented schemes.

Overall, identifying storm types gives insight into the underlying physical processes which can be utilized in a short-term forecast. For example, the continuous propagation aspect of a supercell, or rotating storm, allows for the maintenance of the system over time given the atmospheric properties along the path of the storm do not change much with time. Combining a short-term model (RUC-20) with reflectivity data allows for the monitoring of the atmospheric conditions as related to the current storms as visualized by radar. This field can be updated in real-time to capture storm morphology. For instance, a supercell entering a region in which support for continuous development at low levels is cut off would result in the system morphing into more of a linear convective system.

For the nowcasting problem, the information needed to capture growth, decay, and morphology is encapsulated in the eight classifications described in previous sections. This does not include the many sub-types of supercells and linear convective systems (squall lines) that may appear in the literature. For the research herein, the focus is primarily on the type of weather produced by the different storm types. It can be assumed that the different types of linear convective systems produce high winds on many occasions and that the sub-types of supercells all have the potential to produce hail and damaging winds possibly tornadoes. This research attempts to categorize smaller types of cells such as the pulse, air mass, and small rotating storms, and this can be understood from the stand point of lifecycles. For example, an air mass storm is more likely to have a shorter residence time than a small rotating storm which may evolve into a supercell or at least persist for a longer time. Furthermore, the air-mass to pulse discrimination results in the prediction of severe weather reports. Pulse thunderstorms tend to have strong updrafts capable of producing hail or strong downbursts. These smaller types of storms have been shown to be difficult to classify due to the nature of their short residence times and unavailable information of the three-dimensional structure of the updraft or suspended higher reflectivity cores. It may therefore be concluded that with the lack of three-dimensional information (updraft orientation and height of maximum reflectivity), it may only be useful to separate the rotating small-scale storms from the non-rotating small-scale storms instead of including an attempt to distinguish between severe or non-severe (pulse versus air mass).

Potential uses of the classification include informing an object-oriented nowcast scheme on how to evolve the shapes in space and how to grow and decay based on trends found in the model and trends found in the reflectivity fields. For example, if the objects in the nowcaster are based on ellipses,

background information obtained by the classifier that identifies a supercell thunderstorm at an initial time step and a linear system at a future time step may indicate that the nowcast solution allows for the separation of the focus points found within the ellipse. This can be accomplished by following the trends of a classifier or using forecast model fields. In addition, at a lead time of less than 60 minutes the nowcaster may use information of storm motion to determine a downstream environment that the identified cell will enter. This information may be used to forecast storm type at some lead time less than 60 minutes. For nowcasts made further than 60 minutes, a RUC-20 forecast solution may be used to predict future classification of the identified cell. The incorporation of RUC-20 data at forecast times, instead of the initial time, may introduce more error to the nowcast system but allows the classification to still have some of its roots in the general reflectivity structure of the given identifiable storm.

Trends in the classification over time can also be used to provide insight into the storm's life cycle. This requires the need for some consistency in the observation; however, storm splits and mergers may be able to be properly handled with the increase or decrease in observed cells and their corresponding properties used in the classification process. These trends may also be used to pick out time steps that may have been misclassified due to poor model initialization or gaps in radar data.

MCS longevity studies (i.e. Coniglio *et al.*, 2006) indicate the possibility of using model fields for the nowcast problem which is being used in operation at the SPC. The research herein takes this a step further by breaking down longevity based on individual cells instead of entire regions. It may also be noted that the occurrence of the storm cells in the domain may alter the environment, thus making it important to evaluate storm cells on a case-by-case basis. Finally, it may be useful for future derivations of the storm classifier to use

only the model fields utilized for MCS maintenance outlined by Coniglio *et al.* (2006). This may give better results and physically more realistic results than allowing the tree classification creation scheme to select from a large data set.

The results of the preliminary research in applying classification tree techniques to the nowcasting problem are summarized by the following bullet points.

- Cell classification on the convective scale is more insightful than linear vs. cellular storm classification especially for the nowcast problem.
- Using radar reflectivity-derived products alone does not provide enough information to discriminate the rotational nature of storms.
- The use of RUC-20 NSE fields enhances the capability of detecting rotating storms; however, there are some false alarms associated with using EHI and VGP.
- The results of the VGP to discriminate between rotating and non-rotating storms are consistent with research found examining proximity soundings near supercells.
- The decision tree branches tend to favor more physically meaningful meteorological variables such as VGP and CAPE rather than empirically derived parameters such as EHI. It would be interesting to include other empirically derived variables such as lifted index (LI) and Total Totals.
- The use of storm cells that fit nicely into the given eight categories should only be used; hybrid cases may increase the misclassification rate.
- A weakness of the cell classifier is that it depends on the identification of objects which is based on both verification thresholds and a minimum size criterion. There is a need for a more robust threshold technique for storm identification.

- Additional fields may be incorporated in the future including moisture convergence and reflectivity core area fields.

10.2 Classification Future Work

Future work with the cell classification problem includes a large-scale effort with developing an expert classification database that would require continuous updating. For a given date, cases must be selected that exhibit cells that have a pure classification into one of the eight categories. Further research may also be conducted on decreasing the number of cell types. It may be useful to reduce the number of classification types to six. This would result in removing the severe versus non-severe type of small-scale convection. Combining the air-mass thunderstorm with the pulse thunderstorm along with combining the two smaller-scale rotating thunderstorms may be essential to improving classification until more three-dimensional fields can be used in a near-real time setting. Alternatively, there is the possibility of an increase in storm categories if it is determined there are more than eight life-cycle models that should be applied in the nowcasting problem. Overall, the first steps in future work of the classifier should be to create a large data base of ideal classes of storms to be used in the problem.

The use of velocity products from Level II radar data may give some insight for the storm classification problem. Currently, there are products that can be displayed in the WDSS-II software package which include azimuthal shear and divergence derived from Doppler velocities. The fields appear in regions of significant reflectivity, including ground-clutter. For an example of how these velocity fields may be used, the 12 March 2006 example from 2200Z is shown.

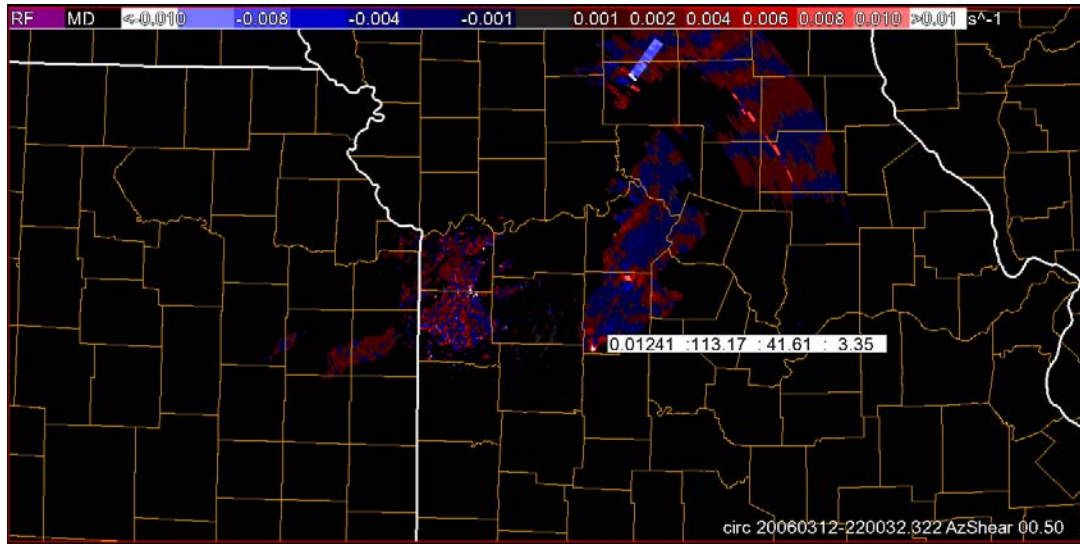


Figure 10.2.1. The azimuthal shear product for 12 March 2006 at 2200Z centered over KEAX at the lowest elevation scan. The southern most supercell has a peak in azimuthal shear of 0.01241 s^{-1} .

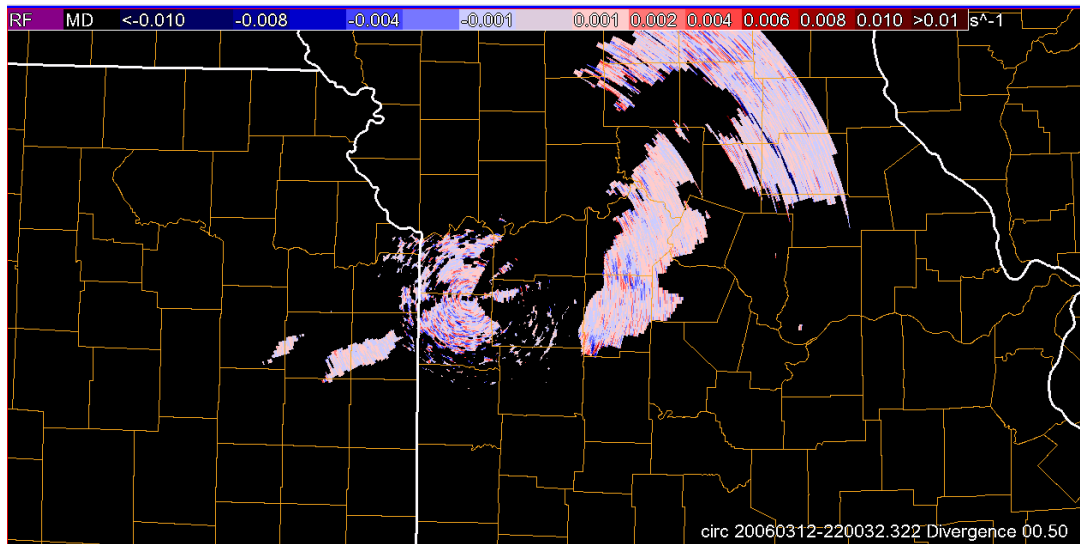


Figure 10.2.2. The divergence product for 12 March 2006 at 2200Z centered over KEAX at the lowest elevation scan. There are significant couplets of convergence and divergence near the main updraft in the supercells present.

Figure 10.2.1 shows the azimuthal shear product for the lowest elevation angle (0.5°). From this product it can be shown that for areas of significant rotation, there is a peak in azimuthal shear values. Figure 10.2.2 shows the divergence field from the lowest elevation scan. At this level, it appears that significant rotation may be found near a couplet of strong divergence and convergence. The utility of these fields in the classifier may be assessed in the near future.

Dual-polarization radar techniques are also a possibility for future cell classification. Dual-polarization observations include hydrometeor classification. This allows for a better estimate of hail probability and therefore better estimations of updraft strength. A vertical cross section using dual-polarization output may also result in observing the orientation of the main updraft for storm longevity considerations. Overall, the classification of hydrometeors from dual-polarization radar can enhance classification of individual storms.

Enhancements of storm identification for the classification experiment may also be needed. Storm cells that are merging or splitting may cause a problem for identification. This can be avoided by using a multi-Doppler radar analysis to calculate updraft motions in the storms (Stalker and Knupp, 2002). From the identification of significant updrafts, storms may be readily identified as being separate entities in the splitting/merging stages of development or morphology. This process will remain a problem in the immediate future as multi-Doppler analysis is only available during research and in a few select regions that may have significant radar coverage overlap.

Finally, there are a number of other methodologies to use in place of the classification tree used in this research. The use of substituting a Bayesian Additive Regressive Tree (BART: Chipman *et al.* (2005)) may give better estimations on the probabilities of proper classification of the expert storm types. In addition, a comparison between the linear Fisher discriminant (LFD) and the

non-linear kernel Fisher discriminant (KFD) with the classification tree methodology might also show interesting results in classifying convective storm types.

Appendix A- Using the Verification Tool

Requirements

Working version of WDSS-II software installed

MATLAB with the Image Processing Toolbox add-on and at least 1GB of RAM

(OR) Octave with similar toolboxes loaded.

I. Collect Data

Level-II NEXRAD files can be obtained from the National Climatic Data Center (NCDC) Hierarchical Data Storage System (HDSS) Access System (HAS) at:

<http://hurricane.ncdc.noaa.gov/pls/plhas/has.dsselect>

Any United States or U.S. territory NEXRAD may be selected; however, when selecting the data only a few days can be selected at any one time. Enter a valid .edu or .gov email address for use of the system. As a side note, if Rapid Update Cycle (RUC) data is desired for specific WDSS-II algorithms only radar data after April 2002 should be selected (RUC-252 only). Once data is selected in from the HAS system the selected files get a HAS identification number. It is important to record this number as many times an alert email indicating the data is ready for download sometimes is not received. The file request normally processes in about 10-15 minutes although when the system is extremely busy it may take up to 48 hours.

To retrieve the data log on to the machine where you wish to store the data, logging in remotely will do the trick. Once logged in create a directory for the case in your home directory for wdss cases (i.e. /home/lack/wdss_cases/) using

the format DDMonthYYradarsite (i.e. 12mar06eax). Change directory into the newly created case directory (i.e. /home/lack/wdss_cases/12mar06eax) and create a new directory with the radar site (i.e. /home/lack/wdss_cases/12mar06eax/eax). Change directory into the radar site directory. Run the following commands:

```
ftp ftp1.ncdc.noaa.gov
login=anonymous
pass=email address used in HAS selection process
type binary <cr>
type prompt <cr>
cd /pub/has/HASXXXXXXXXXX (where X's are case ID #)
mget *.tar
once files are received exit by typing quit
```

Once data is received, run the following commands:

```
tar -xvf (each individual tar file in directory)
gzip -d *.Z
```

II. Setting Up Basic WDSS-II Scripts

The data is now ready to be processed through the WDSS-II system. However, we wish to run cell attributes and possibly other data so we must make a one time download. Change directory into the wdss_cases directory (i.e. /home/lack/wdss_cases). Then use the following commands (NOTE: This is only necessary the first time you wish to set up directories and run the first case).

```
mkdir netssapdat
```

```
wget http://www.cimms.ou.edu/~lakshman/WDSSII/ssapdat.tgz  
tar xzvf ssapdat.tgz
```

You should now have a directory `netssapdat` with many files in it (i.e. `/home/lack/wdss_cases/ssapdat` should exist). This is the directory you must run the WDSS-II commands with in order to run cell attributes algorithms (SCIT, etc.).

Once the directory above it created we are ready to setup a shell script in the `netssapdat` directory to run some of the initial WDSS-II algorithms. Before the script is setup we need some additional information about the radar so that we can get a Cartesian grid of radar data. First of all, we need to know the latitude/longitude coordinates of the radar. Due to security concerns, this information can be tricky to find.

Go to the following webpage in order to calculate the needed NW and SE corners latitude and longitude coordinates for the Cartesian domain:

<http://www.waypoint.org/gps1-calc.html>

Use the second set of form input boxes on the page titled "Compute lat/lon given radial and distance from a known point" and input the coordinates of the radar in the first two boxes. If obtaining a 256x256 domain use `course=0` and `distance=128` making sure distance units is changed to km (instead of nm). The output will be in degrees and minutes so you must convert the minutes into decimal degrees by dividing the minutes by 60. Record the N latitude in decimal degrees for half of the NW coordinate and change course to 270. Repeat procedure recording the W coordinate to get the complete NW set of coordinates in decimal degrees. Repeat the procedure for 180 and 90 (S and E respectively)

until you have NW and SE coordinates. WDSS differentiates between E and W longitude by using a negative sign for W coordinates so make sure you record the longitude with a negative sign. Finally, subtract the N and S components and divide by 256 to get the latitude spacing and repeat for the E and W components to get the longitude spacing. This gives you all the information for a specific radar site to obtain a Cartesian grid of radar elements on a 256x256 grid.

With the geographic data and the radar data acquired from HAS, a shell script can now be written and stored in /netssapdat directory for processing. The following steps should be taken:

cd into the netssapdat directory

vi *radarsite_cases.sh* (i.e. *eax_cases.sh*)

Within the .sh file the following commands should be used with adjustments made to the current radar coordinates (making sure negative sign appears for W coordinate) and the home directory paths are corrected with the proper information (dates, coordinates, radar name, etc.) In the w2merged portion of the script -t refers to the NW corner and -b refers to the SE portion the 3rd set of values refers to the vertical coordinate and you may not want to change these, the -s refers to the spacing first being lat spacing, then lon spacing, and finally vertical spacing for CAPPIs. A sample sh script is found at:

<http://weather.missouri.edu/People/lack/verification>

When this file is created use the command ESC :x to exit and save the .sh file. If you are not running model data make sure to not include the commands that

deal with model files in the sample script, namely anything below `gribToNetcdf` and including `nse`.

To run this file and make sure it is an executable file by running `chmod a+x {filename}.sh` (this may be necessary for all `sh` and `cs` files you download remotely). Then you may run the following command:

```
nohup sh radarsite_cases.sh &
```

This enables the system to be logged out and the `sh` file to be still executed in the background so that others may login to the system and do work as the processing may take up to 8 hours, especially if model data is being processed.

Upon completion you should have a number of directories including `netcdf`, which is the regular radial velocity and reflectivity base directory, merged, for the Cartesian projection, and `ssap`, for storm attributes. You may have a few others if you ran RUC-252 data through the system.

The merged data directory is large directory as it creates files at ~1 minute intervals. For this project we are interested in a Cartesian grid of reflectivity at 10 minute intervals. Files have been set up to deal with this problem. In the `DDmonYYsit` directory create the following directories:

```
mkdir 10min_merged  
mkdir 10min_merged_echotop  
mkdir 10min_merged_mesh  
mkdir 10min_merged_posh
```

```
mkdir 10min_merged_vil  
mkdir 10min_merged_vildensity
```

Once the directories are created, copy `cp10min_add.csh` from <http://weather.missouri.edu/People/lack/verification>. Using emacs (since it has an easy find and replace function), open `cp10min_add.csh` and correct the directory paths and the current date stamps in the file. Run the file using `./cp10min_add.csh` (it should be very quick). If your radar data spanned multiple dates you need to change the current date in the file and rerun the `./cp10min_add.csh` to make sure all the `netcdf.gz` files are in the merged data directories.

From this point, unless you want to keep the many different levels in the merged data directory, it is safe to `rm -rf merged/`. You'll be left with the data at 10min intervals. The reflectivity images are maximum composite images, meaning it takes the maximum return in the vertical to make the composite.

III. Setting up the WDSS-II Nowcast Scripts

We are now ready to run nowcasts off the composite images created by the steps above. It is necessary to get the `segm_run.sh` file from <http://weather.missouri.edu/People/lack/verification>. This file also can be edited with emacs and its find and replace tool to make sure the paths are correct for the case being run. It may be useful to run the `nohup sh segm_run.sh &` command as it will take a bit to process the data. This will create files with nowcast lead times every 10 minutes out to 60 minutes at 6 different threshold ranges, which can be changed at will in the `-d` option within the `w2segmotion` part of the script.

IV. Converting NetCDF to TXT files

Finally, we are ready to make some txt files of all this data to be readily available to the verification scripts in MATLAB. It is now easier to copy these files into a verification folder in the home directory (i.e. mkdir /home/lack/verification). Copy all .m files from /procrustes, the /prog folder, and all other sh files from <http://weather.missouri.edu/People/lack/verification>. For ease of calling files later, make sure all .m files are in the /verification folder and not in a separate /procrustes folder.

The files that now need to be edited using find and replace in emacs include the netcdf_to_txt_info2.sh and netcdf_to_txt_info_segm.sh. If running radar products only (no RUC NSE fields) it may be helpful to copy netcdf_to_txt_info2.sh to another file and remove all the NSE fields in the file. Again all the time and date stamps need to be replaced. The script can then be run using ./netcdf_to_txt_info2.sh or whatever you copied the file to. Repeat the procedure for the netcdf_to_txt_info_segm.sh and run that file to create txt versions of the nowcast fields for the different thresholds and the different lead times.

This concludes the WDSS processing and conversion to txt data for radar and nowcast products only. This does not include model processing or near storm environment (NSE) processing for the cell classification experiment. That will be explained in Appendix B.

V. Calculating Mean Ensemble and Basic Skill Scores

M-files are available at <http://weather.missouri.edu/People/lack/verification> for the creation of ensemble means for the WDSS-II K-means products created with different threshold ranges as well as for standard skill scores including CSI, POD, FAR, and a few others. After downloading the m-files update the paths and the file names for each instance using find and replace in the MATLAB editor window. You may wish to rename and move the ensemble members you created with makeensemblemean.m file to run through the Procrustes Shape Analysis routine in the next step. It may also be useful to compare standard skill scores using podfarcsi.m over lead times with similar forecast verifications from the Procrustes output.

VI. Running the Procrustes Verification Scheme

We are now ready to run the Procrustes Shape Analysis Verification Tool which is run in MATLAB.

The main file to run is procrustesverify.m and the paths must be updated for the txt files you just made in the verification directory. The threshold can also be adjusted at the beginning of the code (variable=tol). There are two sample txt files that were located in the /procrustes directory on the web and with tol=1 you should receive some sample output.

The output should include many images and output statistics to the screen. The penalty function at the end of the code may also be altered as needed.

Appendix B- Using the Classification Tool

Requirements: WDSS-II full install, MATLAB with the Image Processing Toolbox and the Statistics Toolbox and 1GB of RAM

Note: Please refer to the verify_wdss_instruct.doc for some tips on how to process or obtain files to run WDSS-II algorithms.

I. Process RUC-252 Data through WDSS-II

In order to process RUC-252 data through the WDSS-II system, first appropriate RUC initialization data must be obtained. This data can be found or requested at two different places and is available after April 2002. At the present date, no other model formats are able to be processed through WDSS-II. The most popular place to obtain grib files for RUC-252 (RUC20) is from the NOAA National Operational Model Archive and Distribution System (NOMADS). A short cut to the RUC20 data is found at the link below:

<http://nomads.ncdc.noaa.gov/data/ruc/>

Again, you may process both initialization fields and forecast fields from the RUC20, but you must make sure the grib file is around 28-29MB in size or you may receive errors during the WDSS-II processing. A secondary source to request RUC20 files if the date and time you are interested in is not available through NOMADS is to request the data from the Department of Energy (DOE)/Atmospheric Radiation Measurement (ARM) Program. This is done by sending a request for RUC data to the following email address:

armarchive@ornl.gov

In the email specify what products you would like (initializations or both initializations and forecasts), what time frame (dates and range of times), and what data format you need. In the case for WDSS-II processing we need, RUC20 in isobaric coordinates which is identified in the email by specifying the format as 'sgpallruc20isobX1.00'. Once the email is received it may take a day or two for the data to be processed, upon that time you will receive an email with instructions on how to retrieve the data.

II. Setting up WDSS-II to process the data

Upon requesting the data or finding the data you want from the NOMADS site, directories need to be created to process the grib data. Within the case directory (i.e. /home/lack/wdss_cases/12mar06eax) create a directory named ruc20. Change directory so you are in the ruc20 directory and put all the grib files within this directory. Note: sometimes downloading NOMADS data may take some time so it may be useful to write a shell script to download the data using the wget command and return later.

Running algorithms in WDSS-II to process model data is actually quite simple once you have previously had experience running radar algorithms. The sample script in <http://weather.missouri.edu/People/lack/verification> named eaxcases.sh will have three or four lines toward the bottom of the file dedicated to processing RUC20 data. These are tagged with gribToNetcdf and nse with a replaceIndex command found after each. It is important to note that the model processing will

take place on the entire CONUS at 20-km unless bounds are specified in the gribToNetcdf as illustrated in the sample script. The bounds and grid spacing are setup to match the radar Cartesian grid specified. This allows model data to readily be overlaid on the Cartesian grid of reflectivity created earlier. The RUC20 data will be interpolated to a 1-km grid using an optimal interpolation technique.

If you have previously run radar data through WDSS-II, you may comment out the radar processing commands in the sample sh file using '#' in front of each line and comment in (remove the #'s) the gribToNetcdf and nse command lines and their corresponding replaceIndex command lines. This time you should want to use the nohup sh eaxcases.sh & as it will take a few hours to process even 6 hours worth of grib data through gribToNetcdf and the near storm environment (nse) scripts.

Once the data is completed you should have two additional directories named nse and grib and you may look at these in the wg display program.

III. Converting RUC Netcdf to Txt Data

This is done using a similar technique used for converting the radar derived products. However, the file netcdf_to_txt_info_wRUC.sh needs to be downloaded from <http://weather.missouri.edu/People/lack/verification>. This file included both radar derived products to txt, which may be deleted if you have already processed radar files to txt, along with select near storm environmental variables derived from the RUC20 model. To run these select variables through

the system just run the script similar to the manner of the radar files processed earlier. That means this must be done in the verification directory.

Important Note: If you desire to look at other near storm environment fields or some of the RUC20 raw fields you will need to do a little trial and error. Duplicate some of the lines of code in the `netcdf_to_txt_info_wRUC.sh` and replace them with directories of the model parameter you wish to look at by updating the path the netcdf files. Rerun the script and then look at the new directories of the new selected parameter and vi one of the newly created `.txt` files. If there is more than 256 lines in the new `.txt` file you will need to make a correction in the `netcdf_to_txt_info_wRUC.sh`. The correction comes in the form of which `netcdf_to_txt_info` used in the `/prog` directory. An example is below, where the correction needs to be made is in bold and italics.

```
##V_WIND20C

cp /home/lack/wdss_cases/14jul07epz/nse/MeanVWindAt-20C/nseanalysis/*.gz .

mkdir vwind20C

find ./ -type f -exec netcdf_gz_to_txt2.sh {}
/home/lack/verification/14jul07epz/vwind20C/{} \;

rm *.gz

cd vwind20C

#replace T%s.txt with whatever variable you want preceding the data time stamp
on the txt file

ls *.gz | awk -F. '{printf("mv %s.netcdf.gz VWIND20C%s.txt\n",$1,$1)}' | awk -F-
'{printf("%s-%s_%s\n",$1,$2,$3)}' > massrename
```

```

source massrename

rm massrename
cd ..

##MUCAPE

cp /home/lack/wdss_cases/14jul07epz/nse/MUCAPE/nseanalysis/*.gz .

mkdir mucape

find ./ -type f -exec netcdf_gz_to_txt.sh {}
/home/lack/verification/14jul07epz/mucape/{} \;

rm *.gz

cd mucape

#replace T%s.txt with whatever variable you want preceding the data time stamp
on the txt file

ls *.gz | awk -F. '{printf("mv %s.netcdf.gz MUCAPE%s.txt\n",$1,$1)}' | awk -F-
'{printf("%s-%s_%s\n",$1,$2,$3)}' > massrename

source massrename

rm massrename

cd ..

```

Notice that the VWIND20C variable which is the v-component of the mean wind at -20 degrees C uses the *netcdf_gz_to_txt2.sh* while the MUCAPE uses the *netcdf_gz_to_txt2.sh*.

These files need to be changed via trial and error to yield a .txt file which is exactly 256 lines long (or 512 depending on the specifications given for the domain selection). There should now be many folders in the

/verification/DDmonYYsit folder for both radar files and model files all in txt format.

IV. Cell Identification

We can now run a cell identification script to obtain near storm environmental variables within areas of reflectivity in the previously defined domain. The files needed for this are found at:

<http://weather.missouri.edu/People/lack/verification/cellIDclassify>

You need to copy these .m files into the /verification directory. The name of the file to obtain radar parameters and model parameters given a certain identified cell in the domain is numberID_simp.m. This file needs to be updated for paths and model time and radar time for many files. This can be done using find and replace in the MATLAB editor window. If you created a new model field of interest you will need to edit the script where necessary, this may involve adding numerous lines as it is a long and involved m file. For the truly adventurous type, you may find running numberID.m enthralling. This file includes additional lines of code to examine the downstream model fields that the identified storm may push into an hour later by using an estimated steering wind for advection and keeping the original size of the reflectivity object. Once the numberID_simp.m file is run the variable of interest is finalarray. This lists all of the cells and their various attributes that meet the size criteria specified in identifycells.m and threshold criteria set in numberID_simp.m. For what the corresponding columns represent you may find the variable finalarrayheaders extremely useful. This data may be easy cut and paste into excel for further analysis. The script also produces an image for each identified cell that may be useful in referencing the actual data values in the final array variable.

VI. Running the Classification Tree Script

The classification script is found at the above website and is named `treecreation.m` or `treecreationreflonly.m` for classification using reflectivity products only. The m files are very straight forward and the only thing that needs to be done in order to run these files are some postprocessing of the `finalarray` output in the first step. The decision tree inputs must match up to the variables listed in the measurement and data portion. In other words you must determine the type of cell you have list them accordingly to type and then match up the measurements used for classification from the final array. See the MATLAB documentation on `treedisp.m` for further information.

References

- Baldwin, M.E. and S. Lakshmivarahan, 2003: Development of an events-oriented verification system using data mining and image processing algorithms. 3rd Conference on Artificial Intelligence, AMS Annual meeting, Long Beach, CA.
- Baldwin, M.E., J.S. Kain, and S. Lakshmivarahan, 2005: Development of an Automated Classification Procedure for Rainfall Systems. *Mon. Wea. Rev.*, **133**, 844-862.
- Benjamin, S. G., G. Grell, S. Weygandt, T. Smith, T. Smirnova, B. Schwartz, D. Kim, D. Devenyi, K. Brundage, J. Brown, G. Manikin, R. Peterson, 2002: RUC20—The 20-km version of the Rapid Update Cycle. NOAA Tech. Memo. OAR FSL twentyeight. Forecast Systems Laboratory, Boulder, CO, 9 pp.
- Biggerstaff, M.I. and S.A. Listemaa, 2000: An improved scheme for convective/stratiform echo classification using radar reflectivity. *J. Appl. Meteor.*, **39**, 2129–2150.
- Bluestein, H.B., and M.H. Jain, 1985: Formation of Mesoscale Lines of Precipitation: Severe Squall Lines in Oklahoma during the Spring. *J. Atmos. Sci.*, **42**, 1711–1732.
- Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, 1984: Classification and Regression Trees. Chapman and Hall, pp. 358.
- Burrows, W.R., 2007: Dynamical-statistical models for lightning prediction to 48-hr over Canada and the United States. *Preprints: 5th Conference on Artificial Intelligence Applications to Environmental Science, San Antonio, TX*.
- Casati B., G. Ross, and D.B. Stephenson, 2004. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl.* **11**, 141-154.
- Chipman, H.A., E.I. George, and R.E. McCulloch, 2005: BART: Bayesian Additive Regression Trees. Technical Report.

Cohen, A.E., M.C. Coniglio, S.F. Corfidi, and S.J. Corfidi, 2006: Discrimination of mesoscale convective system environments using sounding observations. *Preprints: 23rd Conference on Severe Local Storms, St. Louis, MO.*

Coniglio, M.C., M. Bardon, K. Virts, and S.J. Weiss, 2006; Forecasting the maintenance of mesoscale convective systems. *Preprints: 23rd Conference on Severe Local Storms, St. Louis, MO.*

Davis, C., B. Brown, and R. Bullock, 2006: Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Mon. Wea. Rev.*, **134**, 1772–1784.

Dryden, I.L., and K.V. Mardia, 1998: Statistical Shape Analysis. John Wiley & Sons.

Ebert, E.E., 2007. Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteorol. Appl.* (submitted)

Ebert, E.E. and J.L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. *J. Hydrology*, **239**, 179-202.

Fox, N.I., and C.K. Wikle, 2005: A Bayesian Quantitative Precipitation Nowcast Scheme. *Wea. Forecasting*, **20**, 264–275.

Grams, J. S., W. A. Gallus, Jr., L. S. Wharton, S. E. Koch, A. Loughe, and E. E. Ebert, 2006: The use of a modified Ebert-McBride technique to evaluate mesoscale model QPF as a function of convective system morphology during IHOP 2002. *Wea. Forecasting*, **21**, 288-306

Hondl, K.: 2002, Current and planned activities for the warning decision support system-integrated information (WDSS-II). *21st Conference on Severe Local Storms, Amer. Meteor. Soc., San Antonio, TX.*

Johnson, J. T., P. L. MacKeen, A. E. Witt, E. D. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, 1998: The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Wea. Forecasting*, **13**, 263–276.

Lackshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *Atmos. Res.*, **67-68**, 367-380.

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

Lakshmanan, V., T. Smith, G. J. Stumpf, and K. Hondl, 2007: The warning decision support system - integrated information (WDSS-II). *Wea. Forecasting*, **22**, 596-612.

MacKeen, P.L., H.E. Brooks, and K.L. Elmore, 1999: Radar reflectivity-derived thunderstorm parameters applied to storm longevity forecasting. *Wea. Forecasting*, **14**, 289-295.

Mesinger, F, 2007: Bias Adjusted Precipitation Threat Scores. *Wea. Forecasting* (submitted).

Micheas, A., N.I. Fox, S.A. Lack, and C.K. Wikle, 2007: Cell identification and verification of QPF ensembles using shape analysis techniques. *J. Hydrology*, **343**, 105-116 .

NCAR ICP, 2007: Spatial Forecast Verification Methods Intercomparison Project: <http://www.rap.ucar.edu/projects/icp/index.html>

Parker, M.D., and R.H. Johnson, 2000: Organizational Modes of Midlatitude Mesoscale Convective Systems. *Mon. Wea. Rev.*, **128**, 3413–3436.

Parker, M.D., and R.H. Johnson, 2004: Structures and Dynamics of Quasi-2D Mesoscale Convective Systems. *J. Atmos. Sci.*, **61**, 545–567.

Rasmussen, E.N., and D.O. Blanchard, 1998: A Baseline Climatology of Sounding-Derived Supercell and Tornado Forecast Parameters. *Wea. Forecasting*, **13**, 1148–1164.

Roberts N.M. and H.W. Lean, 2007. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev* (accepted).

Rogers, E., T.L. Black, D.G. Deaven, G.J. DiMego, Q. Zhao, M. Baldwin, N.W. Junker, and Y. Lin, 1996: Changes to the Operational “Early” Eta Analysis/Forecast System at the National Centers for Environmental Prediction. *Wea. Forecasting*, **11**, 391–413.

Stalker, J.R. and K.R. Knupp, 2002: A method to identify convective cells within multicell thunderstorms from multiple Doppler radar data, *Mon. Wea. Rev.*, **120**, 188-195.

Stumpf, G.J., A.E. Witt, E. D. Mitchell, P. L. Spencer, J. T. Johnson, M. D. Eilts, K. W. Thomas, and D. W. Burgess, 1998: The National Severe Storms Laboratory Mesocyclone Detection Algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 304-326.

Thompson, R.L., R. Edwards, J.A. Hart, K. Elmore and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the rapid update cycle. *Wea. Forecasting*, **18**, 1243-1261.

Venugopal, V., Basu, S., and Foufoula-Georgiou, E., 2005: A new metric for comparing precipitation patterns with an application to ensemble forecasts, *J. Geophys. Res.*, **110**, D08111, doi:10.1029/2004JD005395.

Witt, A.E., M.D. Eilts, G. J. Stumpf, J.T. Johnson, E.D. Mitchell, and K.W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303.

Xu, K., C.K. Winkle, and N.I. Fox, 2005: A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities, *J. Am. Stat. Assoc.*, **100**, 1133–1144.

Vita Auctoris

Steven Andrew Lack was born on 25 February 1979 at Southwest General Hospital in Middleburg Heights, Ohio to Jerry and Jody Lack. Along with an older sister, Nicole (now of Roseville, California), the family lived in North Ridgeville, Ohio where the author's parents still reside.

Unlike most meteorologists, no single moment sparked the idea of studying weather as a career for Steven. It was the collective experiences of watching thunderstorms throughout the summer in the neighborhood or going up to Lake Erie to watch storms move onshore. Following graduation from North Ridgeville High School in June of 1997, the author enrolled in The Pennsylvania State University to study meteorology. In December of 2001, Steven graduated and left Happy Valley with a Bachelor of Science degree.

After taking, what amounted to, a semester off after graduation, a graduate research position at the University of Missouri-Columbia opened up in the summer of 2002. The author completed work on a Master of Science degree within the Radar-Hydrology Group in the Atmospheric Science Program in May 2004. He then began work on his Ph.D. in conjunction with the University of Missouri-Columbia Department of Statistics and the RAINS Group in the Atmospheric Science Program on an National Science Foundation (NSF) grant involving statistical radar-based nowcasting.

Upon completion of the Ph.D., Steven will reside in the Boulder, Colorado area with his fiancée, Lindsay, and dog, Dessie Belle, while working for CIRES at the University of Colorado in conjunction with the Aviation Branch of NOAA ESRL's Global Systems Division.