Evaluation of the Storm Cell Identification and Tracking Algorithm used by the WSR-88D

Honors Thesis
Presented to the College of Agriculture and Life Sciences, Physical Sciences
of Cornell University
in Partial Fulfillment of the Requirements for the Research Honors Program

by
Alexander Stuart Lanpher
May 2012
Arthur DeGeatano

ABSTRACT

The Storm Cell Identification and Tracking (SCIT) algorithm used by the WSR-88D radar identifies storm cells and projects their future movements. It can be extremely useful in making short-term predictions about the impact that a thunderstorm may have on a localized area, allowing people to prepare for an approaching storm. In order to assess the accuracy of the Storm Prediction Forecast algorithm within the SCIT, 48 storms in New York and 24 storms in Kansas were chosen and the error between their forecast tracks and their actual paths was determined. These regions were chosen to look for any significant differences in algorithm performance in two disparate geographical regions and to include a region of the U.S. for which the SCIT algorithm has not been previously evaluated. The storms were chosen based on their severity, all being tornadic at some point during their lifespan. The Storm Position Forecast algorithm acts to predict positions for an identified cell in 15-minute intervals, out to a maximum of 60 minutes. To verify that the algorithm improves with time with respect to a given storm cell, successive scans were also analyzed to verify an improvement in the forecast tracks. Once the errors for each position forecast were calculated, comparisons were done between several different parameters to search for correlations with the errors, including distance to the radar station, dBZ value, lead-time, year and time of day. The errors were broken down further based on direction - where a storm went versus its projection. The most outstanding result discovered was in the separation of error into individual directions. For both the New York and Kansas storms, there was a prominent trend for the storm's actual path to be to the right of the projected path. The reason for this preferential rightward error can only be hypothesized here and would be an area for further research.

1. Introduction

Severe thunderstorms impact areas of the United States quite frequently, causing numerous deaths and millions of dollars in destruction. Such storms bring large hail, gusty winds and the potential for dangerous lightning and tornadoes, the most ominous force spawned by thunderstorms. Thunderstorm motion can be difficult to predict, especially further than a few minutes out, but even short term predictions can help get people out of harm's way. These short term predictions are made by the Storm Cell Identification and Tracking (SCIT) algorithm, one of many algorithms incorporated into the WSR-88D Radar Product Generator (RPG).

The SCIT algorithm was first included in the RPG in late 1996, replacing the previous Storm Series Algorithm (Johnson et al. 1997). This algorithm performed significantly better than its predecessor, identifying 96% of storm with maximum reflectivity above 50 dBZ compared to 41% from the Storm Series algorithm according to Johnson et al. Because proper identification is essential to forecasting the motion of storm cells, performed by the Storm Position Forecast sub-algorithm, the SCIT algorithm clearly has the advantage in skill. This improvement, however, isn't sufficient in itself to have confidence in the projections produced by the algorithm. Its projections must be compared to true storm cell locations to have a sense of how accurate the algorithm is. In the case of a tornadic event, even a small error could make the difference between a storm hitting a moderately populated suburban region and a storm striking a densely populated downtown area. This study looks at the errors that can be expected from storms in New York State and Kansas.

a. The SCIT Algorithm

The SCIT Algorithm acts in four sub-functions: Storm Cell Segments, Storm Cell Centroids, Storm Cell Tracking, and Storm Position Forecast (U.S. Department of Commerce 2006). The first two sub-functions act together to identify the location of storm cells and calculate their attributes while the latter two sub-functions track the location and calculate the movement of identified cells (U.S. Department of Commerce 2006). The Storm Position Forecast algorithm is evaluated in this study and will be described in more detail. For more information on the SCIT algorithm, please refer to Part C of the Federal Meteorological Handbook No. 11: Doppler Radar Meteorological Observations.

The Storm Position Forecast sub-function projects movement for a particular storm cell based on the age of the cell. If the cell is newly identified (i.e., the first volume scan in which the cell is identified), the motion is predicted based on the average speed and direction of surrounding cells or, if no other cells are present, on a default speed and direction. If the cell has been present for more than a single scan, its future motion is predicted using its past positions using a linear least squares extrapolation (U.S. Department of Commerce 2006). Projected positions are made in 15-minute intervals, producing 15-, 30-, 45- and 60-minute forecasts. The number of forecasts produced is determined by the accuracy of the algorithm in previous scans – if the accuracy is poor, fewer projections will be made (U.S. Department of Commerce 2006).

b. Previous Studies

A similar study to the present one was done in late 1996 by Johnson et al. (1997) to analyze the performance of the newly implemented SCIT algorithm. In this study, the SCIT algorithm was evaluated on many fronts such as its ability to correctly identify storm cells and, more relevant here, its error in projecting storm cell movement. Analyses were performed on 898 storm cells that were detected by the radar for at least 2 volume scans and the difference between the forecast position and the actual position (as determined by the SCIT algorithm) was calculated (Johnson et al. 1997). Averages were calculated for each 15-minute interval forecast. A table of results is shown in Table 1.

Table 1 – Average forecast error for given lead times for SCIT algorithm. (Johnson et al. 1997)

Forecast time (min)	Number of cells in sample	Average forecast error (km)
5	898	2.0
15	498	5.0
30	227	9.9
45	109	15.2
60	55	22.8

It seems reasonable that, given 898 storms, this is fairly representative data for the performance of the algorithm. And these storms did come from various regions of the country, as shown on the following page. However, it appears that the storms analyzed only reach as far northeast as Virginia, leaving out a significant portion of the northeast in the

evaluation. One motivation for the present study is to incorporate the northeast in this analysis as there are many high-density population areas that could be impacted tremendously by strong thunderstorms, especially if not forecasted properly.

Table 2 – Locations of analyzed storms (Johnson et al. 1997)

Radar and city	Date	Total # human-verified cells	Storm type
KFDR, Frederick, OK	20 April 1992	368	Isolated severe
KOUN, Norman, OK	2 September 1992	473	Isolated severe
KMLB, Melborne, FL	25 March 1992	734	MCS
KMLB, Melborne, FL	9 June 1992	591	Isolated nonsevere
KMLB, Melborne, FL	12 June 1992	580	Isolated nonsevere
KLSX, St. Louis, MO	8 June 1993	512	MCS/Line
KLSX, St. Louis, MO	2 July 1993	325	Line
KTLX, Oklahoma City, OK	18 June 1992	147	Isolated severe
KTLX, Oklahoma City, OK	21 February 1994	116	Stratiform
KIWA, Phoenix, AZ	6 August 1993	867	Isolated severe
KIWA, Phoenix, AZ	20 August 1993	832	Isolated severe
KLWX, Sterling, VA	14 April 1993	282	Minisupercell
KLWX, Sterling, VA	1 May 1994	218	Minisupercell
KLWX, Sterling, VA	6 October 1995	237	Minisupercell
KCBX, Boise, ID	1 May 1995	129	Isolated severe
KBIS, Bismark, ND	21 May 1995	152	Isolated nonsevere
KBIS, Bismark, ND	7 June 1995	235	MCS/Line
Total		6561	

2. Data Sources

The primary resources used for this analysis are the NCDC Storm Event Database, archived NCDC NEXRAD Data and the NOAA Weather and Climate Toolkit. In order to obtain a list of thunderstorms to analyze, certain criteria were chosen to restrict the number of potential storms to choose from. Here, only storms that were tornadic in nature at some point in their lifespan were analyzed. The NCDC Storm Event Database was used to access a list of recorded tornadic storms in both New York and Kansas during the time period in which NEXRAD data was available (1994 to the time of the analysis). Once a list of storms and associated locations was obtained, archived NEXRAD data (specifically, the Storm Tracking Information (STI), Long Range Base Reflectivity (460 km) and Tornado Vortex Signature (TVS) products) were downloaded and visualized using the NOAA Weather and Climate Toolkit. The radar site chosen for a particular cell was based on the proximity of the cell to the site (the closest site to the town experiencing the tornadic event was chosen). Upper level wind data were obtained from the NCDC Service Records Retention System (SRRS) to compare thunderstorm cell motion to upper level winds.

3. Methodology

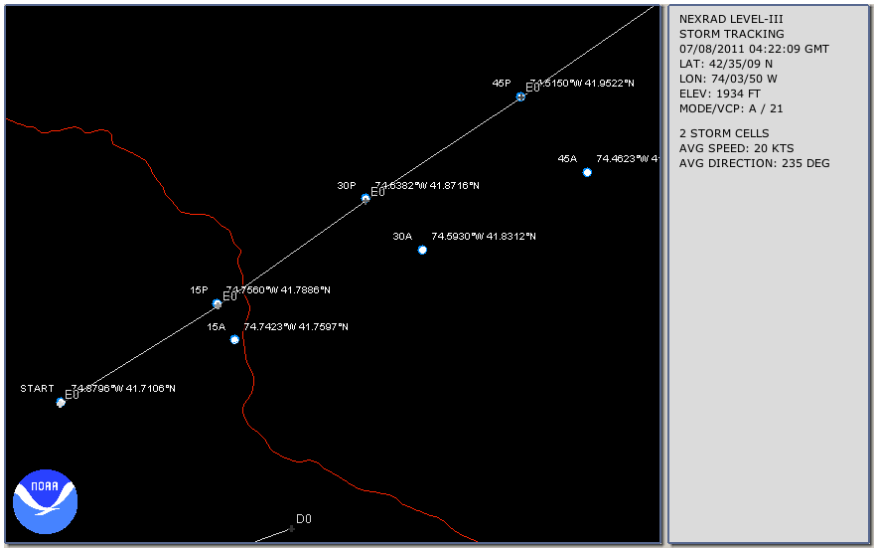
a. Data Collection

The primary goal of the study was to calculate the error for each forecast interval of a set of thunderstorm cells in New York as well as for a set in Kansas. This would provide an average error for the algorithm's 15-, 30-, 45- and 60-minute forecasts, very similar to Johnson's study. A secondary goal was to verify the algorithm's improvement with time with respect to a particular thunderstorm cell. This seems like a reasonable behavior to

expect given that the forecast track for a continuing cell is based on previous motion (U.S. Department of Commerce 2006). First, though, the storms to be analyzed had to be chosen from the Storm Event Database list.

Tornadic thunderstorm cells in New York from 1994 – 2011 were limited in number and some did not have sufficient data in the NEXRAD archive data. Forty-eight storms were chosen under the condition that the cell associated with the tornado could be identified with sufficient confidence by comparison of STI identified cells to radar images and tornadic signatures. Without clear overlap of at least 2 of the features (either overlap between the STI identified cell and a relatively isolated reflectivity-based storm cell or between the STI identified cell and the associated TVS), a tornadic event was excluded from the set. More generally, storms with longer tracks were chosen in order to maximize the data obtainable from each storm cell. Kansas storms were more numerous and easier to find, but the same criteria were applied. With these conditions satisfied, current and forecast storm positions were plotted using the toolkit as shown in figure 1.

Figure 1 – Recording methodology using NOAA Weather and Climate Toolkit



The letter-number combination is the cell ID, the plotted points give latitude and longitude, "P" represents a projected position, "A" represents an actual position, a number (15, 30, etc.) indicates the lead-time and white lines simply connect forecast positions. These latitude-longitude values were manually typed into a text document that was later read in by a FORTRAN program to calculate errors and other parameters.

Unfortunately, this process did introduce some error because the radar volume scans were not consistently performed in terms of frequency. In the ideal case, scans were done every 5 minutes, easily allowing collection of actual cell positions at 15-minute intervals. Due to the inconsistency, some estimation was involved in identifying actual cell locations. In the case that a volume scan only differed by 1 minute from the desired time of cell identification, the position given by the STI product at the time of the volume scan was used for the identification of the actual storm position for the desired interval. For example, if a storm were initially identified at 00Z with associated 15-minute interval forecasts, there would be (at best) a forecast position for the storm at 00:15Z, 00:30Z, 00:45Z and 01:00Z. To obtain the actual position of the cell at 00:15Z and evaluate the 15-minute

projection would require a volume scan at 00:15Z. In the case that the closest available volume scan was at 00:16Z, the actual cell location indicated on this volume scan was taken as the actual position of the cell for the 15-minute forecast. In the case that there was more than a one minute discrepancy between available and desired volume scans, the actual cell locations were estimated under the assumption that cells moved at a constant speed within a short (~5 minute) interval. So, following the aforementioned example, if the only available volume scans were at 00:13Z and 00:17Z, the positions of the cell at these times were marked and then a visual midpoint was labeled as the actual position for the 15 minute forecast.

Additionally, depending on the accuracy of previous forecast positions, a cell may not have all four 15-minute interval forecast positions available. For example, some cells had all four forecast positions produced by the algorithm, but others only had 15- and 30-minute projections if the Storm Position Forecast algorithm was having trouble with forecasting the cell's positions. In other cases, later volume scans did not produce projections when the cell was no longer identified by the SCIT algorithm. Specifically, in this data set, there are 218 15-minute, 186 30-minute, 150 45-minute and 117 60-minute forecast-actual location pairs.

In order to evaluate the change in forecast accuracy over time with respect to an identified cell, consecutive volume scans were analyzed. In particular, if the analysis in figure 1 represents the initial set of forecast positions for a particular cell, a volume scan 3 scans later was loaded using the NOAA Toolkit and marked in the same manner. Following this, a volume scan 3 scans after this one was loaded and once again marked and so on until 5 sets of forecasts were obtained for the cell (an initial scan and then 3, 6, 9 and 12 volume scans out). These scans will be referenced as the initial scan, scan 1, scan 2, scan 3 and scan 4 from here on out. The choice of 3 volume scans between sampling times was fairly arbitrary but gave sufficient time for the algorithm to accumulate a history of cell motion on which to base future projections and kept the time short enough to track a cell over numerous time intervals before it weakened/was no longer identified by the algorithm.

In summary, the values recorded for each storm cell and for each scan were the time of the volume scan, the dBZ value of the cell at the time of the volume scan and the coordinates of the starting location of the storm cell, 15-minute interval projections of the cell, actual positions of the cell at these intervals, and location of the radar site used.

b. Calculations

Once all of the data was recorded, a MATLAB program was written to calculate a number of values. These values were the 15-, 30-, 45- and 60-minute error for each storm cell tracked, the distance between the storm cell and the radar site used to track it, the directional error (i.e., where a cell ended up with respect to its projected location) and the projected and actual speeds of the cell over each 15-minute interval. To calculate the errors, the north-south and east-west distances between the coordinates of the projected and actual cell positions were found and then multiplied by a factor representing the number of kilometers per degree of latitude or longitude as appropriate. Then the Pythagorean theorem was simply applied to obtain a straight-line difference between the two locations. The distance to the radar site was calculated in the same manner. The directional error was calculated based on the sine of the angle between the north-south

error and the absolute error (the hypotenuse of the triangle formed by the error components). It describes where a cell ended up relative to its projected position, either N, NE, E, SE, S, SW, W or NW. The calculation of actual speed/projected speed was done by taking the distance the cell traveled/was projected to travel (again, found using the Pythagorean theorem) and dividing by the time of travel (15 minutes in each case) and converted into meters/second. All of these values were calculated with both the New York and Kansas storm sets, but only the raw errors and directional errors were analyzed in the Kansas data due to time restrictions and the significance of the findings calculated first in the New York set.

c. Correlations

In addition to the raw errors, many values were compared in order to look for correlations between variables. These comparisons were error versus lead-time, distance to radar site, year, dBZ value and time of day and storm motion versus upper level winds. The comparison of error to lead-time is one relationship that Johnson analyzed in his 1996 study so the current values will be an update as well as a look into values for the Northeast. Distance to the radar site was looked at to see if greater distances gave the SCIT algorithm more difficulty since the energy from a radar beam spreads out with longer distances. The year was looked at to see if the SCIT algorithm improved with updates, specifically the Radar Product Generator update that was released in September 2004 (U.S. Department of Commerce 2006). The storm motion versus upper level wind was analyzed in an attempt to see which level of the atmosphere had the greatest impact on the motion of thunderstorm cells. Though any correlations certainly do not guarantee causation, they can lead to important relationships in the behavior of severe thunderstorm cells.

4. Results and Discussion

a. Error versus Time

The results of the average calculated error values are shown in table 3 below. Not surprisingly, error values do increase with increased lead-time. Longer-term forecasts allow more time for factors affecting thunderstorm motion to change. This is unchanged from Johnson's study. The magnitudes, however, are significantly different from those in Johnson's study with reductions in error of up to 6km. The second important result seen from this table is the improvement in forecasts over time. The largest forecast error tends to occur earlier in the track (i.e., the earliest scan) when the SCIT algorithm has a shorter time period over which a cell has been tracked. This supports the idea that the algorithm does use accumulated history of a cell's motion in its projections successfully. Another result apparent from comparison of the tables is that the performance of the algorithm is, qualitatively speaking, overall worse in Kansas than in New York. To test for significance, a t test was conducted after necessarily checking that the data was sufficiently Gaussian (Wilks 2011, p. 143). A bootstrapping method was used to verify the presence of this requirement (Barreto 2005). The test indicates a significant difference in the 45- and 60-minute projections when combined across all scans (initial through 4) with p values of 0.03 and 0.02, respectively. The 15- and 30-minute error did not show statistically significant

differences. The relatively poor performance in Kansas is somewhat surprising given that the topography of Kansas is more uniform and flatter than that of New York. Seemingly, this would allow for more linear and less turbulent thunderstorm cell motion, making cells easier to track; in this sample, however, that was not the case. This surprising result may be attributable to the Kansas tornadoes being stronger, implying more severe storms and potentially more rightward motion.

Incorporating time of day into the analysis in table 4 shows a slight correlation between the time of day and the error associated with the algorithm. The time of day was divided into 3 parts, morning, afternoon and evening defined as 4AM–12PM (08Z–16Z), 12PM–8PM (16Z–00Z) and 8PM–4AM (00Z–08Z), respectively. The errors tended to be noticeably smaller in evening storms compared to morning and afternoon. The reason for this is not clear; however, since the 48 New York storm cells had to be divided into even smaller subgroups, any outliers in terms of error could have significantly skewed the average for its subset. T-tests were not able show statistically significant differences.

Table 3.1 – Lead-time versus average error (km), New York

LEAD TIME	INITIAL SCAN	SCAN 1	SCAN 2	SCAN 3	SCAN 4
15	4.4	4.2	4.1	4.0	3.6
30	8.9	7.0	6.1	6.2	5.6
45	13.3	8.3	7.6	7.6	7.1
60	16.3	11.4	8.1	9.0	8.6

Figure 2.1

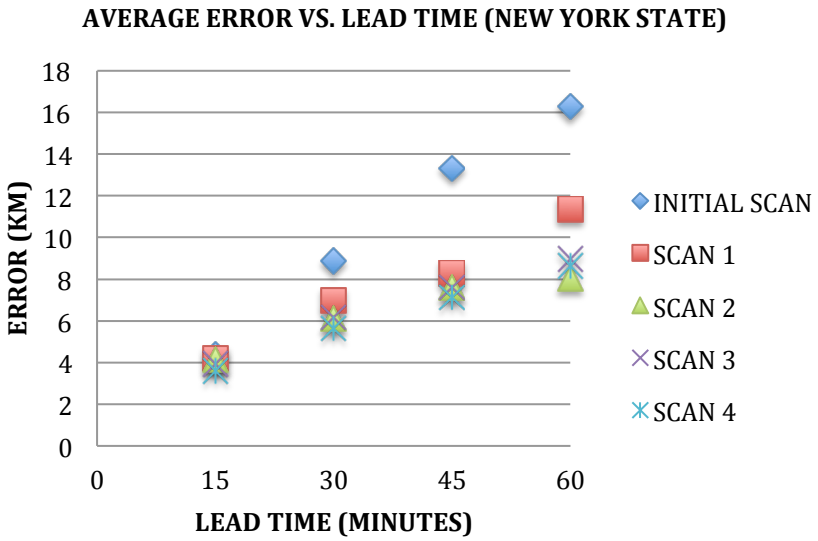


Table 3.2 Lead-time versus average error (km), Kansas

LEAD TIME	INITIAL SCAN	SCAN 1	SCAN 2	SCAN 3	SCAN 4
15	4.4	3.7	3.7	4.2	4.1
30	8.7	7.3	7.4	6.9	6.1
45	13.5	10.9	11.2	9.3	10.5
60	17.7	14.4	13.3	12.4	12.3

Figure 2.2

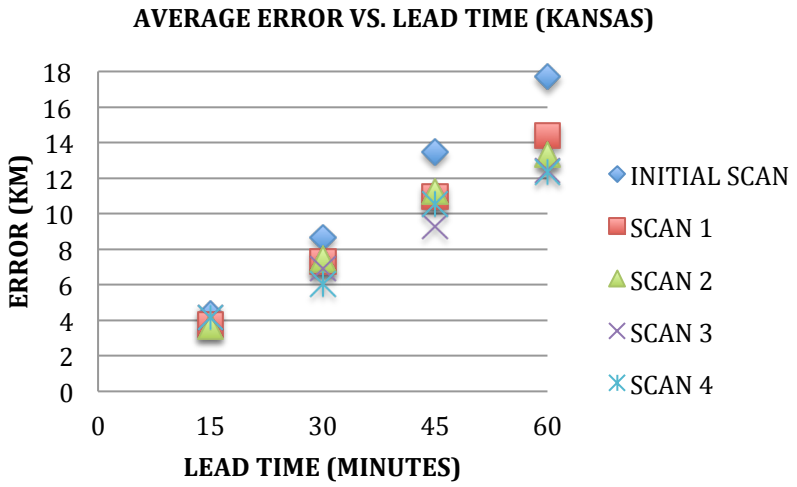


Table 4 – Average error (km) versus time of day

TIME OF DAY	15E	30E	45E	60E
MORNING	4.7	9.4	12.0	14.3
AFTERNOON	4.1	6.7	9.2	10.8
EVENING	3.3	5.6	7.6	11.4

If the average error is compared to the year of storm occurrence, there is a consistent downward trend with time across most lead-time errors as seen in table 5. This drop in error should be most significant around the time of the Radar Product Generator update released in September 2004 (U.S. Department 2006). A t-test does show significant differences between the 1997-2004 period and the 2005-2010 period. The p values were .01, .02, .03, .01 corresponding to the 15-, 30-, 45- and 60-minute errors, respectively, when combined across all scans (initial through 4). This supports the effect of the RPG update increasing the accuracy of the SCIT algorithm to a significant extent.

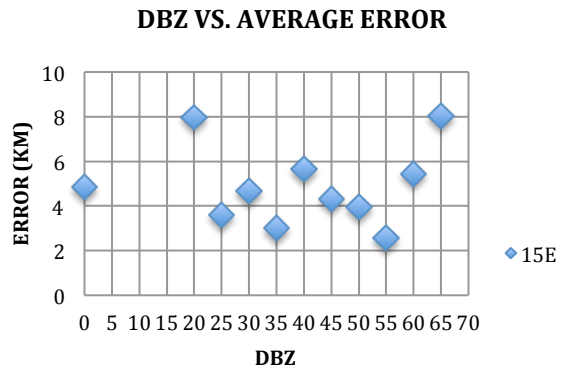
Table 5 – Average error (km) versus year, New York

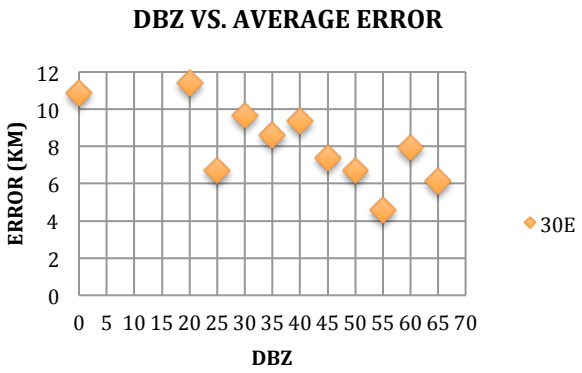
YEARS	15E	30E	45E	60E
1997-2000	4.9	8.2	10.4	12.3
2001-2004	4.0	7.1	10.8	14.7
2005-2007	3.9	6.3	8.5	9.2
2008-2010	3.2	6.0	8.1	10.5

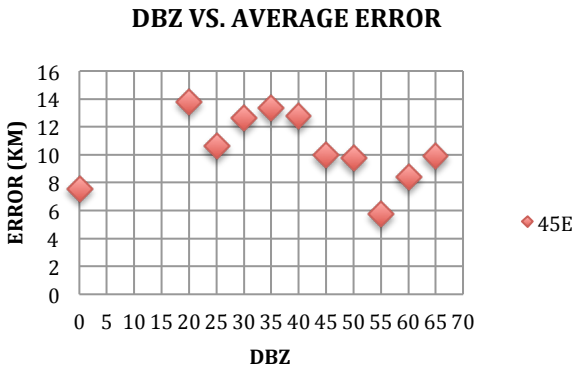
b. Error versus Reflectivity

An analysis of average error versus reflectivity value for New York cells presents an interesting relationship. When looking at average errors across all dBZ values, the correlation is weak to non-existent (see figure 3). If, however, only a certain range of dBZ values is analyzed, there is actually a strong linear relationship. Viewing the average errors for only those storms with a dBZ between 40 and 55, there is a strong negative correlation

Figure 3a – dBZ versus average error







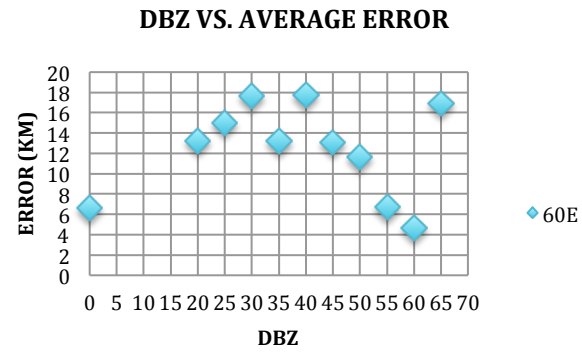
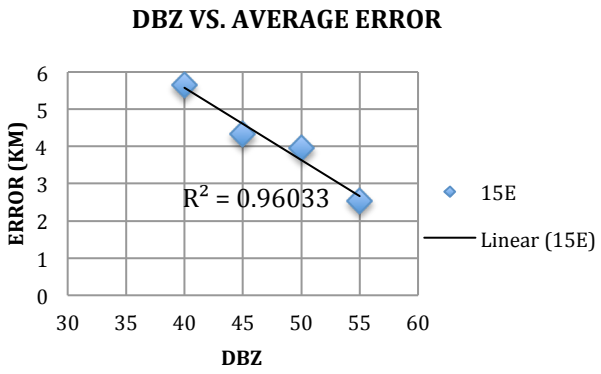
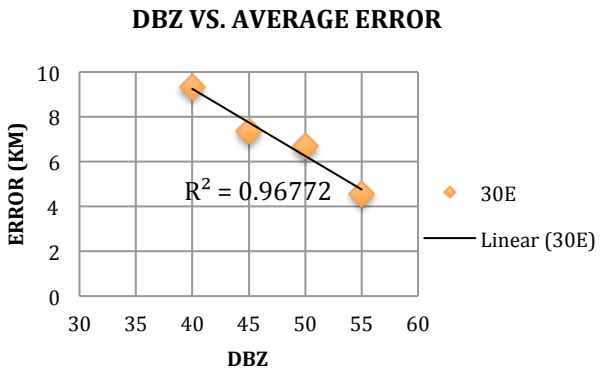
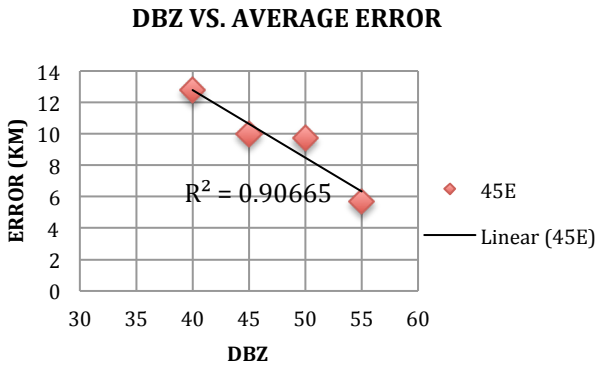
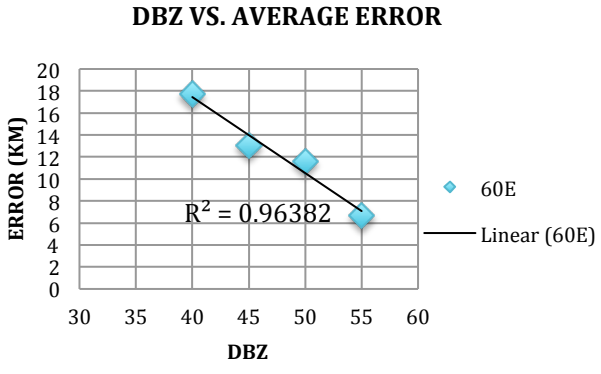


Figure 3b– dBZ versus average error 40-55 dBZ







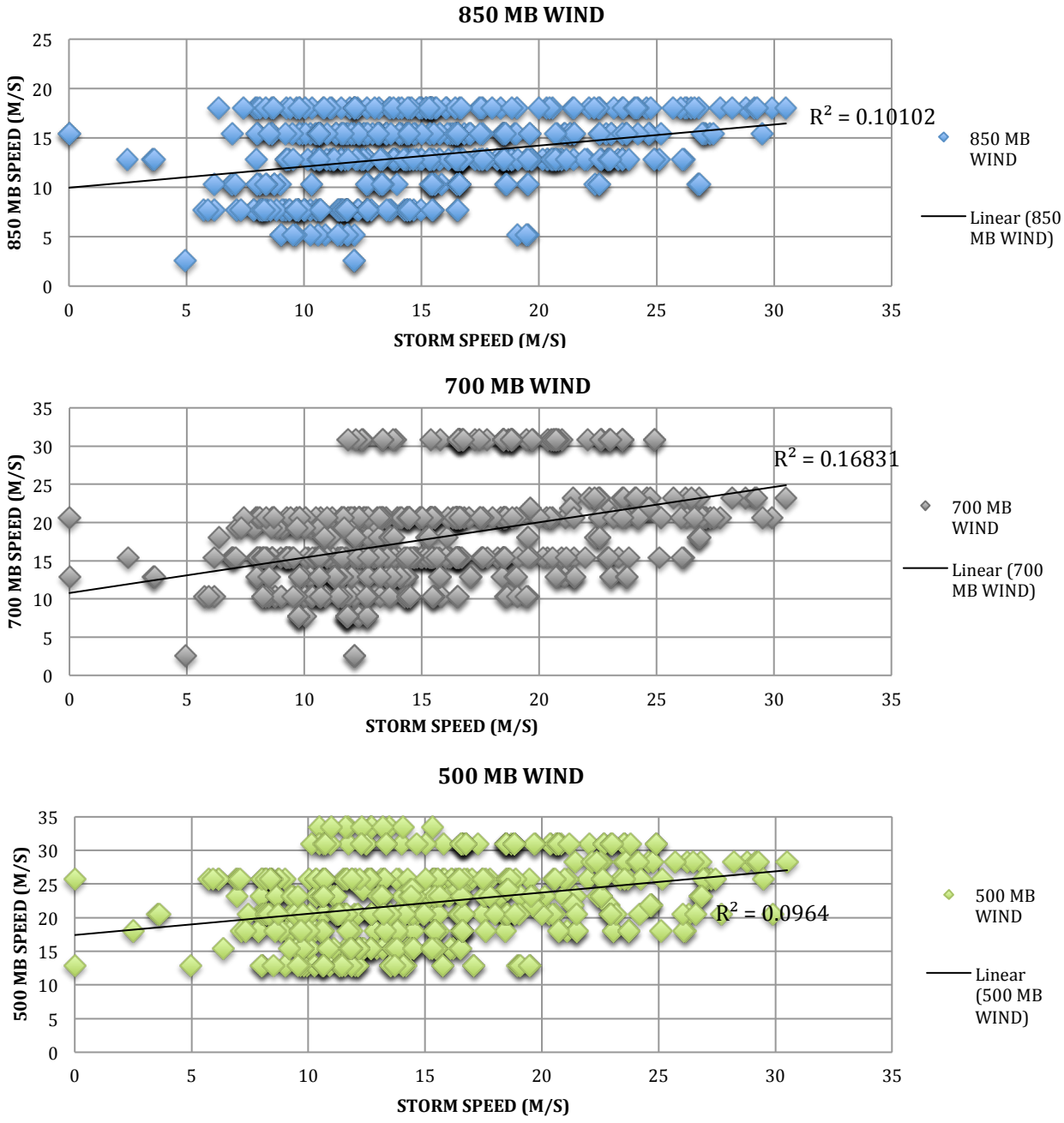


between error and reflectivity with the coefficient of determination, R^2, ranging from 0.91 to 0.97. While neglecting data is not ideal in any dataset, it is at least somewhat justifiable here based on the ability of the SCIT algorithm to correctly identify weak thunderstorm cells and due to differences in sample size. In Johnson's study, less than 30% of isolated severe and mesoscale convective system/line storms with a reflectivity less than 40 dBZ were detected (Johnson et al. 1997). Since the algorithm has trouble identifying weaker storms, it may have trouble identifying them properly when such cells are detected, which is the foundation of the Storm Position Forecast sub-function within the SCIT algorithm and of the accuracy of its projections (Johnson et al.). Furthermore, since the sample of storms obtained in the present study contains only 27 storms with a reflectivity less than 40 or greater than 55 out of 217 total storm cells (12.4%), the average errors for very weak and very strong storms certainly may not be representative. Since the number of storms in the range of 40 to 55 dBZ is far greater (87.6% of the total), there is much more confidence in these average errors being representative of the population of thunderstorm cells within these reflectivity limits. To claim that there is a definitive relationship between dBZ and error is beyond the scope of this sample size, but it motivates further investigation.

c. Motion versus Upper Level Wind

In order to explain some of the error and directional error (which will be described in the next section), an attempt was made to find a correlation between thunderstorm cell motion and the upper level winds at 850, 700 and 500 mb. There was not a strong correlation between storm cell speed and any of the upper level winds analyzed as seen in figure 4. This is likely due to three reasons: 1. Upper level wind data is fairly sparse in terms of the density of available data; for example, there are only 3 upper air stations in New York State, located in Buffalo, Albany and Upton, from which it was necessary to estimate an upper level wind speed for each thunderstorm cell, 2. Thunderstorm cell motion with respect to upper level winds is a fairly complex relationship involving an average value of wind speeds and directions; for example, a study conducted by Maddox in the 1970's on environmental conditions for tornado occurrence, a mean storm motion estimate involved an average of wind speeds at the surface, 850, 700, 500, 300 and 200 mb (Maddox 1975), and 3. Upper level wind measurements are only done twice per day at 00Z and 12Z, seriously restricting the availability of data at the time of storm occurrence. Accordingly, the resulting correlations were weak but positive with R^2 values ranging from 0.10 to 0.17. The strongest relative correlation was with the 700 mb wind speeds. A resulting analysis of the wind direction at 700 mb was performed to search for a link between the directional error and the 700 mb wind. Again, this had to be estimated using the data available. This analysis didn't yield particularly useful results, mostly because all of the upper level wind observations were dominated by a westerly component, making it difficult to see any distinctions storm by storm. If more significant correlations could be found, upper level wind data could potentially be incorporated into the SCIT algorithm to improve its accuracy given that it currently only uses radar data to make projections.

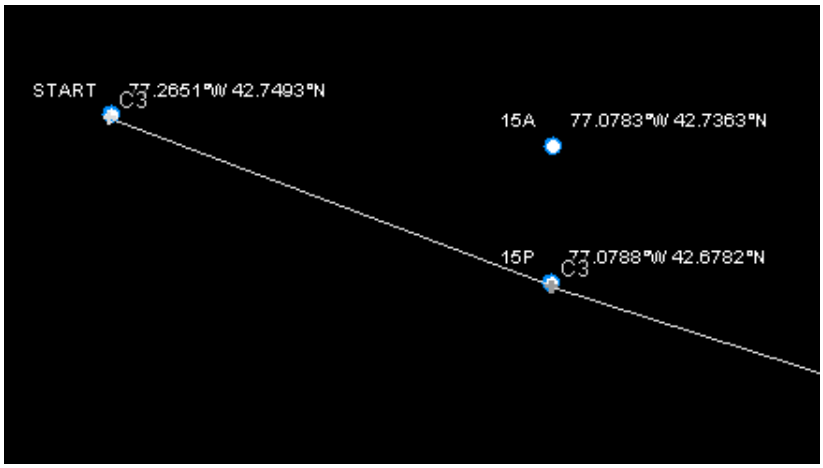
Figure 4 – Upper level winds versus error



d. Directional Error

As previously defined, directional error represents the actual location of the storm cell with respect to its projected location at the respective time. In figure 5, with 15P and 15A representing the 15-minute projected position and the 15-minute actual position, respectively, the directional error would be north since the cell ended up north of its projected location. The results of this analysis, shown in figure 6, were the most surprising

Figure 5 – Directional Error (NOAA Weather and Climate Toolkit)



and significant. In both the New York and Kansas sets, there was a distinct tendency for the directional error to be to the south, southwest or southeast. This distinction was broken down further into rightward and leftward error. Rightward error would represent a cell ending up to the right of the track proposed by the SCIT algorithm and vice versa for the leftward error. The error in figure 5 would be leftward error since the cell ended up going to the left of the forecast track. The directional error was broken down into the

Figure 6a - Directional error, New York

DIRECTION	FREQUENCY
N	56
NE	55
E	70
SE	109
S	148
SW	134
W	62
NW	37

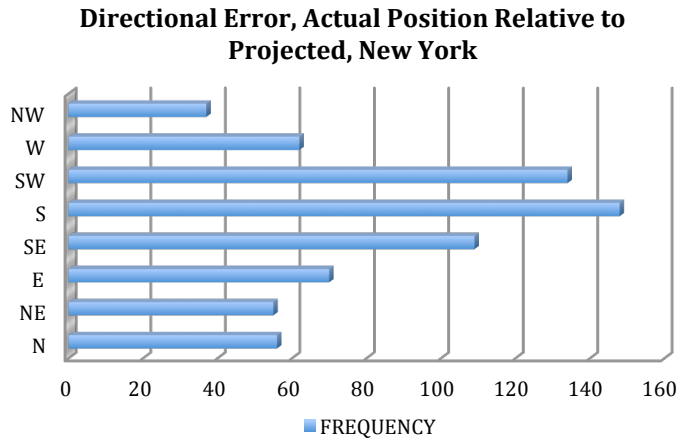
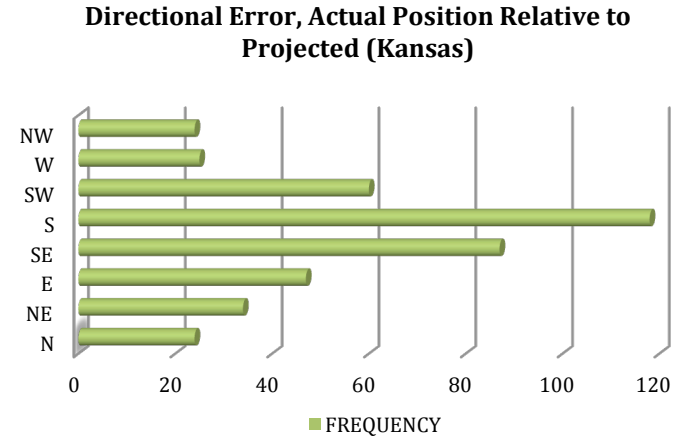


Figure 6b – Directional error, Kansas

DIRECTION	FREQUENCY
N	24
NE	34
E	47
SE	87
S	118
SW	60
W	25
NW	24



rightward/leftward error construct and into a speed error category for each scan (initial, 1, 2, 3 and 4). This was determined using the projected direction of motion, the actual direction of motion, the directional error and the error in speed. If the actual direction of storm travel was to the right of the projected path, rightward error was recorded. If the actual and projected directions were the same, the directional error was used to identify a rightward vs. leftward error. The result was a very strong tendency for storms to travel to the right of their forecast tracks. Storm motion was also categorized by the actual speed relative to the projected speed (i.e., did the storm move more quickly or more slowly than predicted). Figure 9 shows the results of the rightward/leftward error and speed analysis along with the ratio of cells with rightward error to cells with leftward error for each scan.

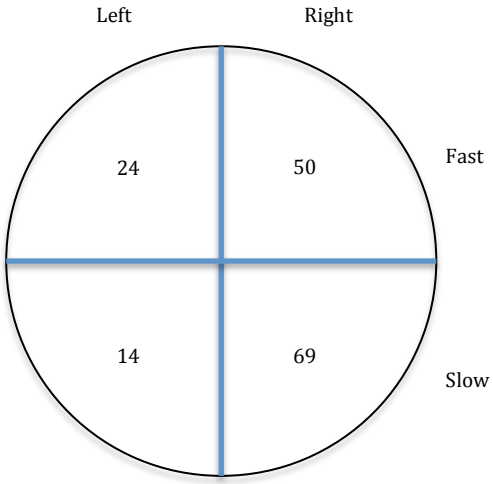
To obtain a sense of how statistically significant the difference in directional error is, a chi-squared test was performed. Given 8 categories for the directional error, the expected value in each category is the total number of storm divided by 8 – 84 for New York and 52 for Kansas. The sum of the squares of the difference between the observed and expected values divided by the expected values across the categories provided a chi-squared value of 140 for New York and of 158 for Kansas. Both of these values are significant at a probability of 0.9999. It seems as though this pattern in error is systemic given its prominence in both storm sets.

It is not clear based on the data collected why this trend in directional error is present. One hypothesis is that it has a relationship to the tendency for severe thunderstorms to move to the right of the area of convective activity in which it is located. This observation is affirmed by Browning, stating that many studies in the late-1950's and early 1960's show frequent rightward motion of severe thunderstorm cells (Browning 1964). This pattern strongly identifies with figure 7, implying some relationship to this observation since all thunderstorms cell in the present study produced a tornado and were severe as a result by the National Weather Service definition (U.S. Department/NWS . The rightward movement of severe thunderstorms seems a plausible explanation for the persistent rightward error in the data sets. This claim is supported by the fact that rightward error is, on average, greater in Kansas where storms could be expected to be more severe than those in New York. This reasoning, however, becomes complicated considering that the SCIT algorithm makes forecasts based on previous motion. Therefore, if a cell developed a consistent track to the right of its earlier tracks, the algorithm should adjust and reduce the error by producing a track that bears more to the right. Accordingly, the rightward errors should be maintained in number if not decreased across successive scans. This is almost the case for the New York set of storms as the rightward to leftward ratio drops continuously with time except for scan 3. The ratio for Kansas storms is much more erratic, potentially due to a smaller sample size. Reason behind the persistent rightward error is an area for future research.

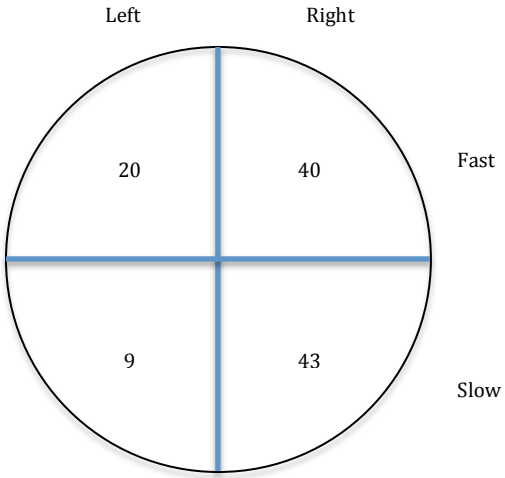
e. Error versus Distance to Radar

The comparison of storm error to distance from the radar beam yielded a very poor correlation. If anything, there is a slight negative correlation, but it is not significant enough to make any conclusions. A comparison of error to distance to radar for the 15-minute

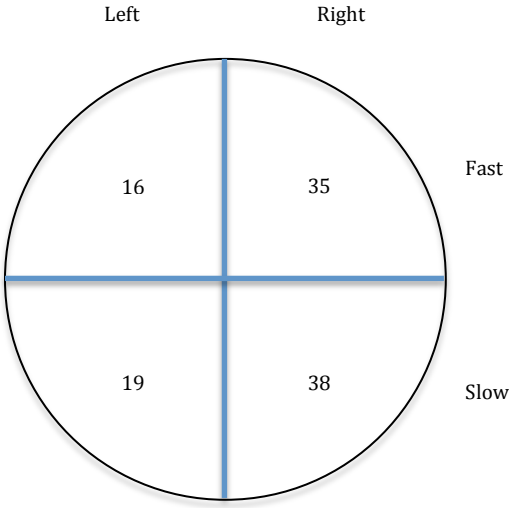
Figure 7a – Rightward vs. leftward error with right:left ratio and speed relative to projection, New York



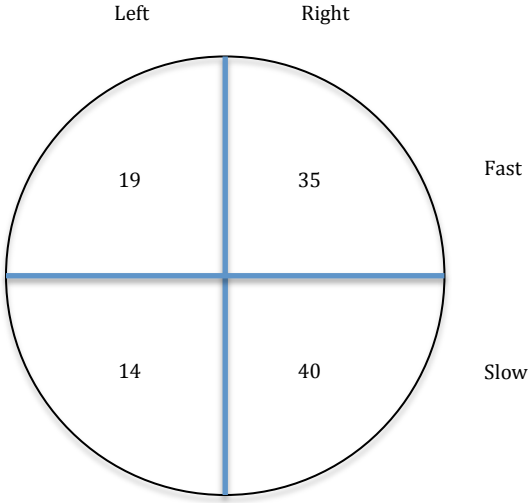
Initial Scan, 3.13:1



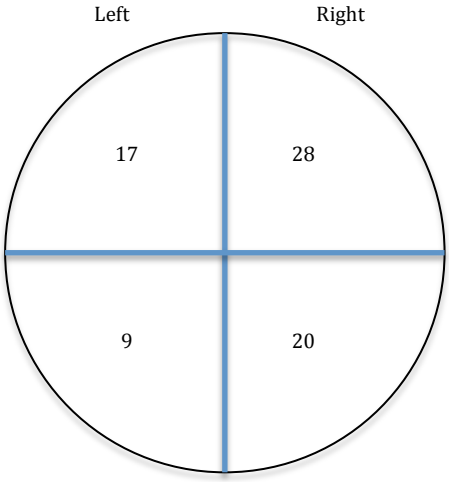
Scan 1, 2.86:1



Scan 2, 2.09:1

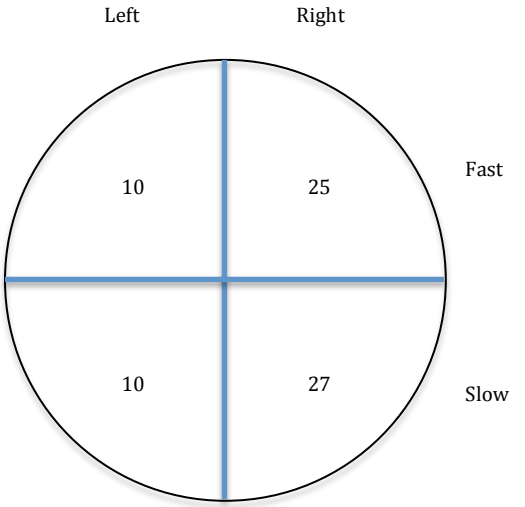


Scan 3, 2.27:1

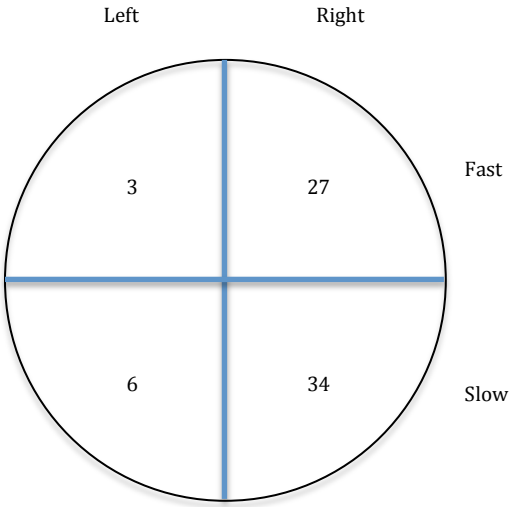


Scan 4, 1.85:1

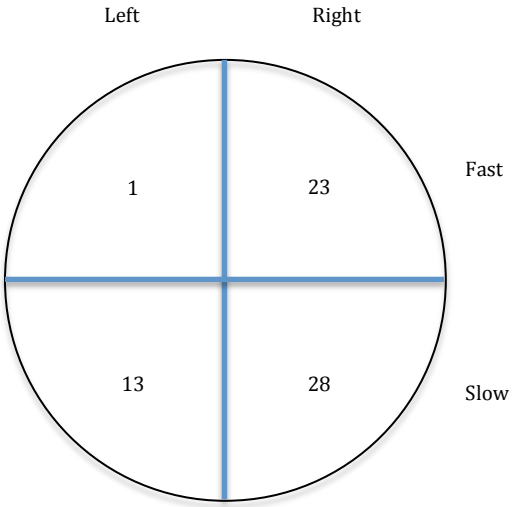
Figure 7b – Rightward vs. leftward error with right:left ratio and speed relative to projection, Kansas



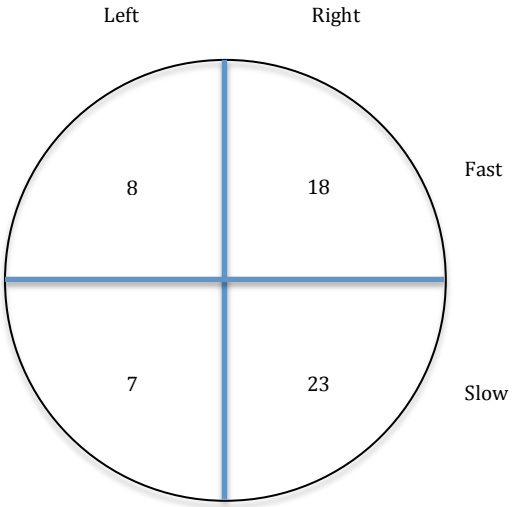
Initial Scan, 2.6:1



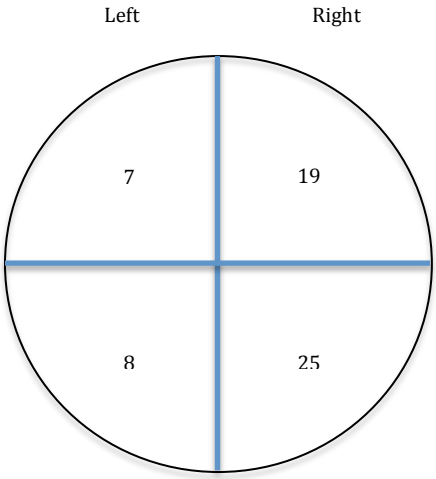
Scan 1, 6.78:1



Scan 2, 3.64:1



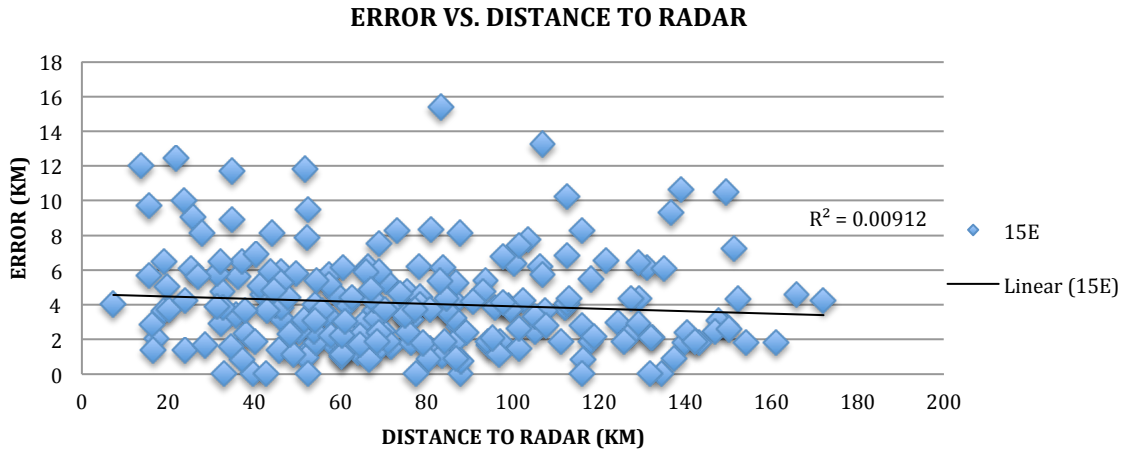
Scan 3, 2.73:1



Scan 4, 2.93:1

error for all scans combined (initial through 4) is shown in figure 8 below to represent the general correlation. This would imply that the SCIT algorithm performs equally well regardless of the distance between the storm cell and the radar beam.

Figure 8 – 15-minute error across all scans versus distance to radar



5. Summary

Though the SCIT algorithm does do a relatively good job making storm cell projections, it does produce some errors with its current forecasting method. The majority of comparisons between error and other variables produced weak correlations; some other correlations, however, specifically those related to directional error, appear significant and to be systemic. If the cause of these consistent errors could be discovered, the algorithm could be improved further to make more accurate storm cell location forecasts. This could increase meteorologists' ability to correctly forecast thunderstorm motion and warn the right area of an approaching storm.

6. Future Research

The most significant finding in this study is the rightward error of thunderstorm cells. The reason for this error could not be confidently identified, but a hypothesis was proposed. To verify or disprove this hypothesis, rightward error and raw error would have to be analyzed on a storm-by-storm basis and, to provide a control, a set of non-severe thunderstorms would have to be analyzed to check for the rightward error. Additionally, the dBZ versus error showed some correlation, but the data would have to be broadened to include cells of more reflectivity values to make a more generalized correlation.

Acknowledgements. Special thanks to research advisor, Arthur DeGaetano, for support as well as Mark Wysocki for input and research guidance and Brian Belcher for technological assistance.

Works Cited

Barreto, H. and F. Howland, 2005: Bootstrap Excel Add-in. Wabash College.

Browning, K. A., 1964: Airflow and Precipitation Trajectories Within Severe Local Storms Which Travel to the Right of the Winds. *Journal of Atmospheric Science*, **21**, 634–639.

Johnson, J. T., P. L. MacKeen, A. Witt, E. D. Mitchell, G. J. Stumpp, M. D. Eilts, and K. W. Thomas, 1998: The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm. *Weather and Forecasting,* **13**, 263-276.

Maddox, R. A., 1976: An evaluation of tornado proximity wind and stability data. *Monthly Weather Review*, **104**, 133–142.

U.S. Department of Commerce/National Oceanic and Atmospheric Administration, 2006: *Doppler Radar Meteorological Observations*. Federal Meteorological Handbook, **11**, 3-23–3-29.

U.S. Department of Commerce/National Oceanic and Atmospheric Administration, cited 2012: National Weather Service Glossary. Available online at [http://www.weather.gov/glossary.]

U.S. Department of Commerce/National Climactic Data Center, cited 2012: Storm Events Database. Available online at [http://www.ncdc.noaa.gov/stormevents.]

---, cited 2012: NCDC NEXRAD Data Inventory Search. Available online at [http://www.ncdc.noaa.gov/nexradinv.]

---, cited 2012: SRSS Analysis and Forecast Charts. Available online at [http://nomads.ncdc.noaa.gov/ncep/charts.]

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press, 676 pp.