P2.2 PERFORMANCE OF A STATISTICAL FRAMEWORK FOR ANALYZING LARGE LIGHTNING AND RADAR DATASETS

Amanda R. S. Anderson*, Timothy J. Lang, and Steven A. Rutledge
Colorado State University, Fort Collins, Colorado

1. INTRODUCTION

The study of convective storms involving both polarimetric radar and lightning data has, in the past, been approached in a case study manner due to the large datasets involved. Examples of such case studies include Tessendorf et al. (2005, hereafter T05), Wiens et al. (2005, hereafter W05), Tessendorf et al. (2007, hereafter T07), Goodman et al. (2005), and Deierling and Peterson (2008), to name a few.

A supercell from the Severe Thunderstorm Electrification and Precipitation Study (STEPS; Lang et al. 2004) on 29 June 2000 was examined by T05 and W05. The T05 study concentrated on the formation of large hail within the storm, finding that hail embryos could come from a larger region than the embryo curtain identified by Browning and Foote (1976). The second part of the study, W05, focused on the lightning activity in the supercell. Findings included that the total flash rate was well correlated with volumes of updraft and inferred graupel echo, charge structure of the storm varied in the horizontal and with time, and the positive cloud-to-ground (+CG) flashes originated around 5-9 km altitude while the negative cloud-to-ground (-CG) flashes originated higher, consistent with an inverted tripole charge structure. Another storm from STEPS, the 3 June 2000 case, was examined by T07. This storm exhibited an inverted dipole, with positive charge situated beneath upper negative charge. No CG lightning was observed. The study found that the lack of a lower negative charge layer may have been a key factor in suppressing CG lightning production.

Studies using the lightning mapping array (LMA) observations (Rison et al. 1999) have also been performed outside of the STEPS experiment. Goodman et al. (2005) studied severe weather with the North Alabama LMA and radar data. They found total flash rates could provide an indication of possible severe weather to aid in the warning decision process. They also found that total flash rates tended to have large increases prior to tornadogenesis, although not all large increases in flash rate were followed by tornadogenesis. The Deierling and Peterson (2008) study showed that updraft volume > 10 ms^{-1} above the -5° C level was well correlated to the total lightning activity. This relationship did not vary between the two different climate regimes of the southeastern United States and the High Plains. They also found the relationship between updraft volume of the entire storm and total lightning activity varied as a function of temperature and, at certain temperatures, a function of region. Overall, they found that the total lightning activity of deep convection could provide a measurement proxy for updraft mass fluxes and associated ice-phase precipitation, useful in mesoscale numerical models and severe storm nowcasting systems.

Although this case study approach has proven valuable, as shown from these results, the case studies generally lack statistical significance, which makes it difficult to generalize findings. For this reason, a framework has been developed to study large volumes of data in a statistical sense. For the details of this framework, please refer to Lang and Rutledge (2008; 2009). It is important to test the performance of this framework before undertaking new studies. To test the framework, two test cases from STEPS, 29 June 2000 and 3 June 2000, were run through the framework and compared against previously published studies, specifically those by T05, W05, and T07. The following sections describe the methodology and results of these tests and draw conclusions about the overall performance of the framework.

2. DATA AND METHODOLOGY

The statistical framework combines gridded radar data and lightning data into one accessible file, which can then be run for statistics of the user's choice. To test the STEPS cases, radar data previously gridded by

* *Corresponding author address:* Amanda R. S. Anderson, Colorado State University, Department of Atmospheric Science, 1371 Campus Delivery, Fort Collins, CO 80523-1371; e-mail: aanderson@atmos.colostate.edu

T05. W05, and T07 were used. These gridded volumes included fuzzy logic hydrometeor identification. The gridded volumes were then run through an in-house cell tracking program. This tracking algorithm is a hybrid of the Storm Cell Identification and Tracking (SCIT) and the Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN) algorithms, and consists of tracking an ellipse drawn around a cell identified by 35 dBZ and 45 dBZ composite reflectivity thresholds. Information on SCIT and TITAN can be found in Johnson et al. (1998) and Dixon and Wiener (1993), respectively. The sensitivity of cell tracking performance with assumed thresholds is discussed below in the Results section.

Lightning data from the LMA was then analyzed using New Mexico Tech's XLMA program. Individual flashes were sorted by using the included “sort flashes ignoring altitude” feature. Positive and negative charge was assigned automatically using a program developed by Dr. Kyle Wiens. The goal of the flash sorting was to make the process as automated as possible, which would assist in processing the large amounts of data needed for a statistical analysis.

The lightning data were appended to the gridded radar volumes using the statistical framework, which assigned flashes to a given cell in each volume. The particular flashes used in the comparison were separated from those within other cells based on the results of the cell tracking algorithm.

To make the most complete comparison possible, dual-Doppler information from T05 and W05 were also added to these files. A framework module was created to append this dual-Doppler information to the files so they could be matched to the objectively defined cells.

Once the data were collected into one file, the data were used to emulate plots produced in past case studies of STEPS data. In addition to these figures, correlation calculations were used to examine the overall trends found in both the published studies and the framework runs. The goal of this process was to see if the results show the previous case study analyses are comparable to the results obtained from from the statistical framework.

3. RESULTS

3.1 *Tracking Code*

Overall, the tracking code performed well for both the 29 June case and the 3 June case. In both cases, the best tracking results were with the main cell of the study, though secondary cells were also tracked well over time.

A major issue with the tracking in the 3 June case was the identification of two cells within the one main cell. This resulted in “skips” in the main track, where the identification jumped from following one of the inner cells to following the other (Fig. 1). The track follows the main cell the entire time, but the particular cell within the main cell changes. This required following along the track and adding in the extra cell in each volume by hand in order to ensure the most complete comparison between the statistical framework and the literature. For this case, it would be better to have tracking that would follow the main cell in its entirety, so that a manual manipulation of the track was not needed.

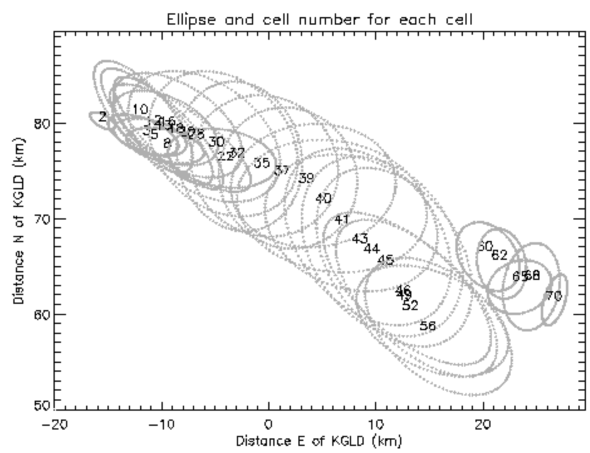

Fig. 1. Track of 3 June main cell. Jumps between cells occur between cells 8 and 10 and again between cells 56 and 60.

Another issue with the 3 June case was that, due to 35 dBZ being the lowest reflectivity value for cell identification, the main cell was no longer identified toward the beginning or end of its lifecycle. There were no lightning flashes during this time, so there was no problem in this case. This could present a problem for certain applications, but the framework allows for different tracking algorithms to be used on the front end, so this could be tweaked to suit the user.

For both cases, spurious cells were identified. These were sometimes identified as part of the main cell and added to the track for tests of the 29 June and 3 June storms. Some of these cells were also found to be correctly identified as separate from the main storm.

Again, a different tracking algorithm may be better for certain applications that do not need individual cells, such as feeder cells, identified as separate from the cell of interest. For example, a multicell storm that is being studied as one storm system, as opposed to a collection of separate cells.

3.2 *Lightning Flash and Charge Sorting*

Both the flashes used in this study and the flashes used in the T05 and W05 studies used flash sorting performed by the XLMA program. The number of LMA sources available to be counted depends, in part, on filtering done by the number of LMA stations that detect a source and the χ^2 value of the source. Initially, a 7 station filter was used and a χ^2 value of 1 assigned. After running the flash sorting program, a reasonable number of flashes per volume scan was found compared to the published studies. However, a much smaller source density was also found. Changing to a 6 station filter and χ^2 value of 2, the source density more closely approached that of W05 (not shown), but the number of flashes decreased. This may have to do with the way XLMA breaks up large groups of sources for flash sorting, but the code must be examined more closely for a more definitive answer.

The charge sorting program produced results similar to those achieved by T05 and W05, particularly with the positive charges. However, issues were found, mostly related to the negative charge. Many very high frequency (VHF) source points were being misclassified as negative charges. The charge sorting program looks for negative leaders traveling toward positive charge from the initiation of the discharge. Flashes that do not have this clear motion at the beginning of their lifetimes, or which have have poorly resolved structure, tend to be misclassified (Fig. 2).

The above findings of differences in XLMA sorting and the charge sorting misclassifications bring the question of whether it may be necessary to sort flashes by hand, as was done in the published studies on the 29 June and 3 June storms. The general trends of flashes are well correlated, as described in the following section. Also described below is an issue with an important finding from W05 unable to be replicated. Hand analysis of flashes would be detrimental to the goal of

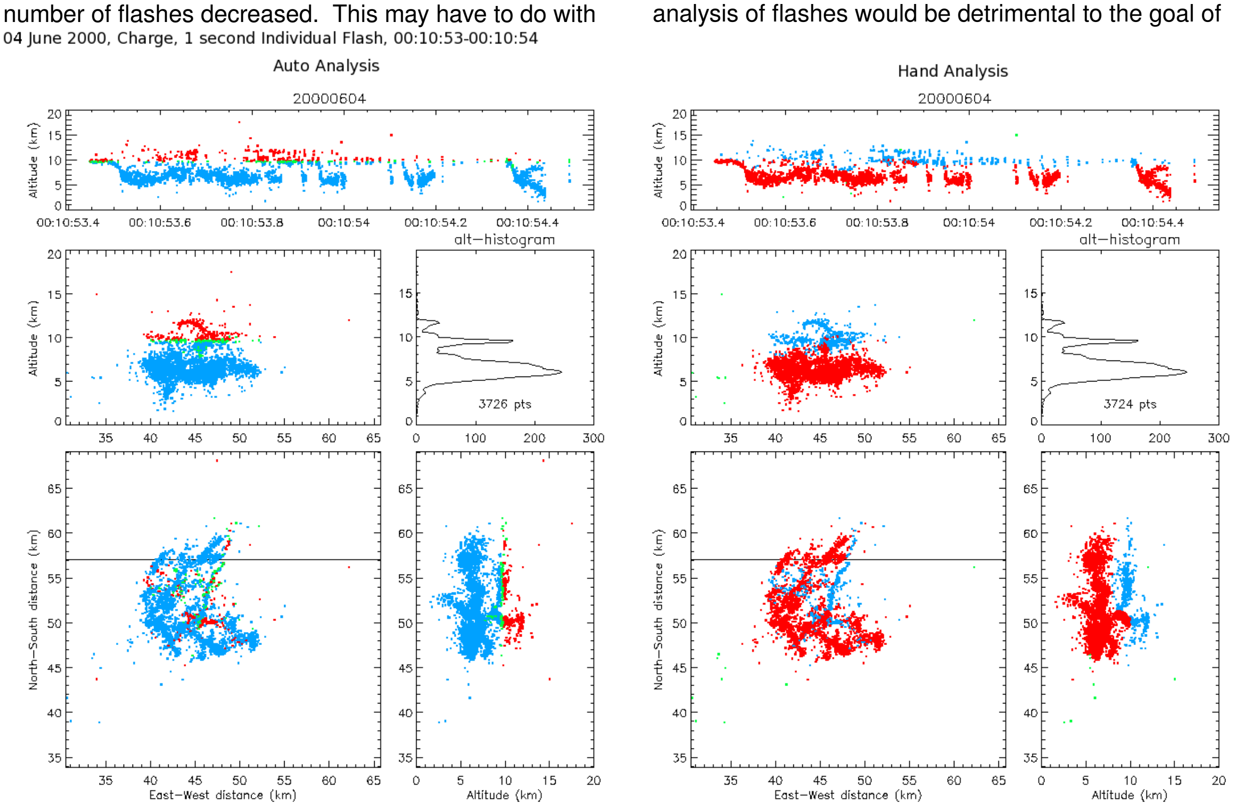

Fig. 2. Example of charge sorting results for 3 June storm. For this flash, the automated algorithm misidentified the charge structure of this flash compared to the results of the hand analysis, possibly due to the lack of clear motion at the start of the flash.

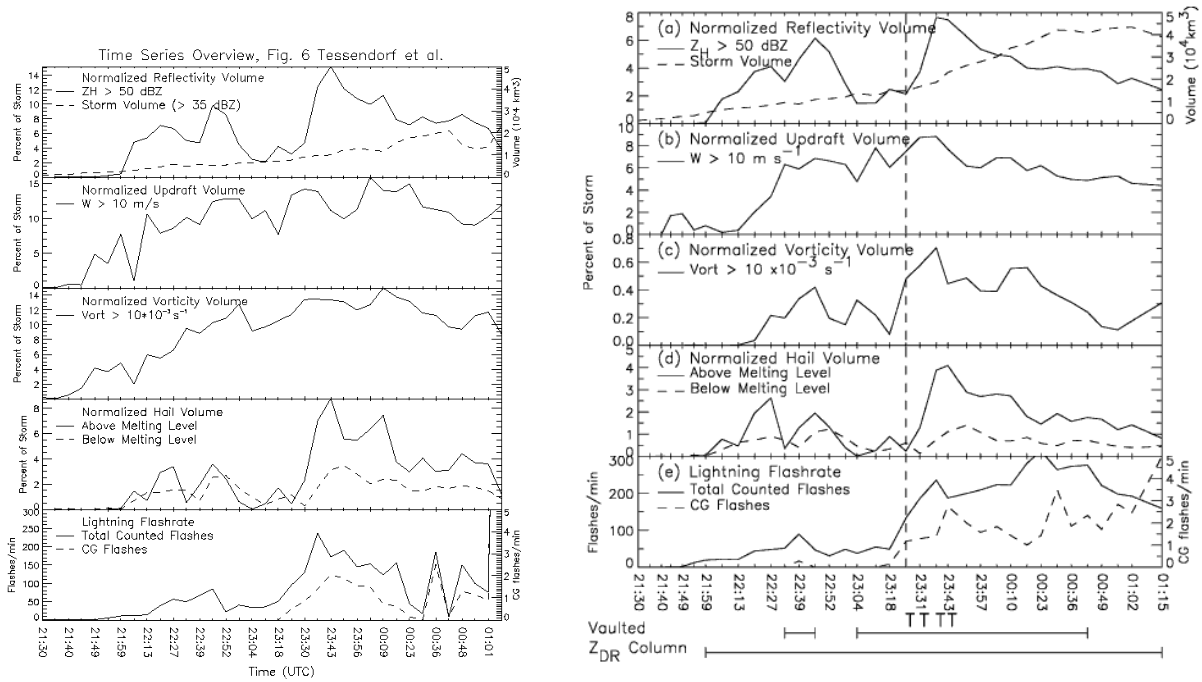

Fig. 3. Comparison of statistical framework reproduction of 29 June results (left) and T05's hand analysis results (right). Note that the general trends of Z > 50 dBZ and hail volume are quite similar. Issues exist with the updraft volume, vorticity volume, and lightning flash rate.

an automated system to analyze large amounts of data. More work must be done examining lightning data to see if the desired automated system is achievable.

3.3 *Usage of Final Data*

Once the data were run through the statistical framework, the resulting files were used to replicate figures from previous studies. Additionally, correlations were run to compare trends in lightning flashes between the statistical framework runs and the published studies.

Plots of HID information were made to compare to the 29 June and 3 June cases. The data from the published studies was not available for objective comparisons, but the plots provided a visual comparison of the data. For both cases, the HID data (graupel volume, hail volume) compared very favorably with the published studies (for example, Fig. 3). This was expected, as this data was taken from the same files used previously. The only difference was the way the data were processed into plots through the IDL program. The biggest difference was in plots of percent of cell volume. In the published studies, cell volume was defined as the volume of >0 dBZ echo (Fig. 3). The tracking program used on the front end of the statistical framework identified cells by the echo >35 dBZ in the composite reflectivity, or the maximum reflectivity in each column of bins. Consequently, plots based on this tracking cover a smaller volume of storm. The volumes of graupel and hail were not affected by this due to the fact that these hydrometeors typically exist in areas of higher reflectivity. However, the smaller storm volume affects both the percentage and general trends of the curve. Depending on the application, a different tracking algorithm on the front end could be more beneficial to the study. These findings are also apparent in Fig. 4. Here, graupel and hail echo volume are quite well matched with the hand analysis. Updraft volume is also well matched in this example.

The most significant differences were found in the lightning flash data (Fig. 3, Fig. 4). As described in the previous section, the thresholds of station detection and χ^2 value impacted the magnitude of flashes sorted. Additionally, these thresholds affected the density of LMA sources through the 29 June storm's lifetime (Fig. 5). The source density appears to trends similarly, but

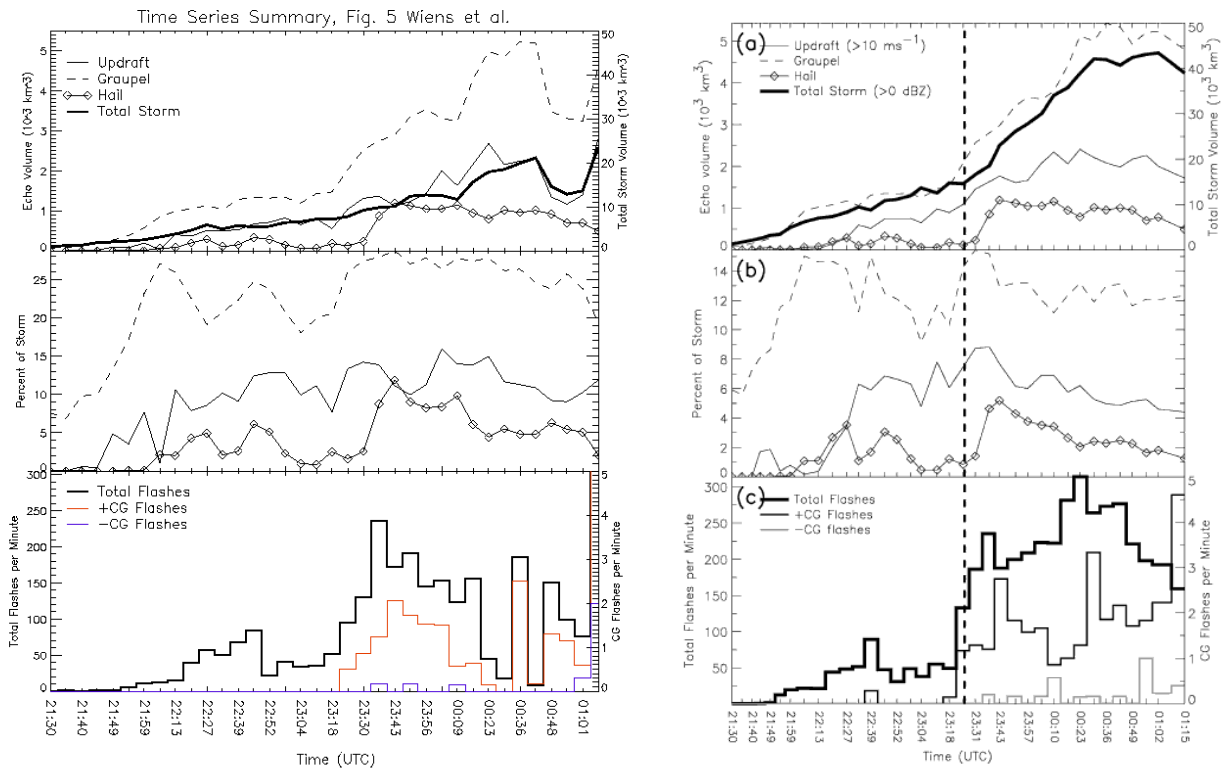

Fig. 4. Comparison of statistical framework reproduction of 29 June results (left) and W05's hand analysis results (right). Graupel and hail echo volumes show good agreement. Cell volume is much lower but trends similarly. Updraft trends similarly as well. Lightning flash rates are somewhat different, but correlate well as discussed in text.

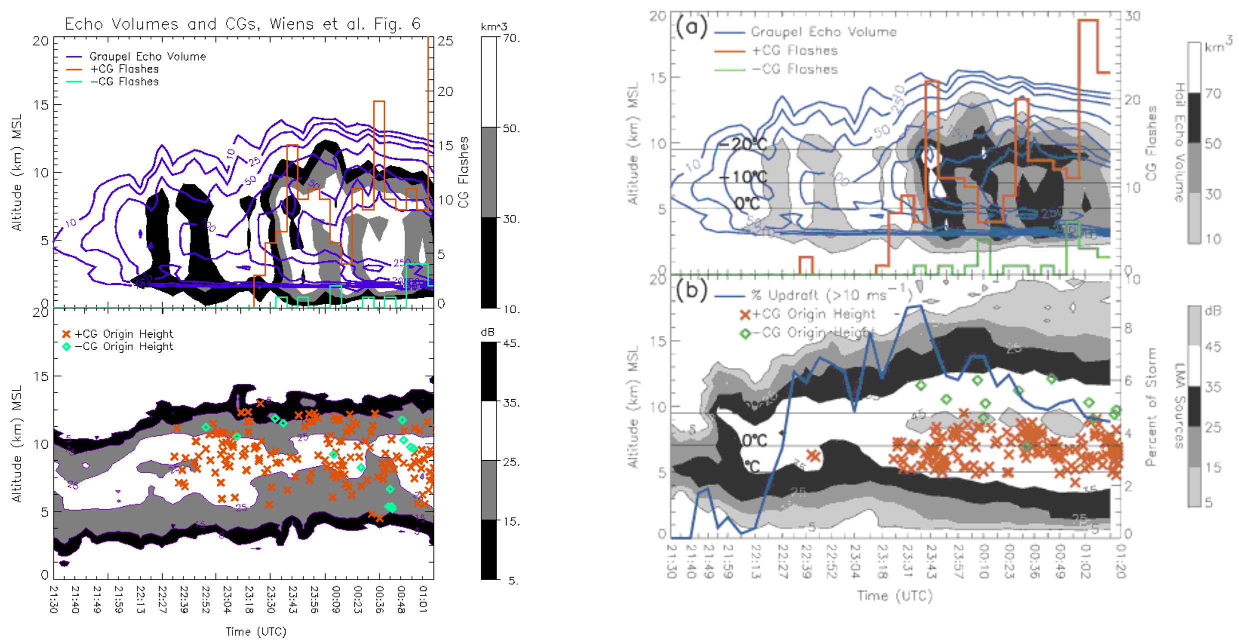

Fig. 5. Comparison of statistical framework reproduction of 29 June results (left) and W05's hand analysis results (right). Graupel and hail echo volume show very good agreement between the two methods. Issues exist with source density and especially with CG origin height.

more analysis is needed to make a more definitive claim.

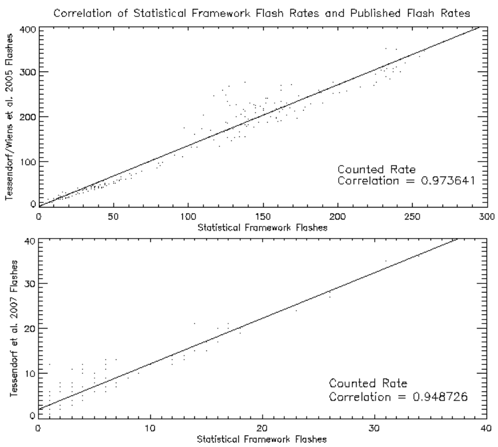

Fig. 6. Correlation of total flash rates from the 29 June (top) and 3 June (bottom) storms. Plots show statistical framework flashes versus hand analysis flashes, with a fit line drawn to show departure from one-to-one fit.

Flash rates from previous studies were available, so correlations between these flash rates and flash rates from the statistical framework were calculated (Fig. 6). For the 29 June case, the correlation of total lightning flash rate over the storm's lifetime, from flashes per minute, was 0.974. For the 3 June storm, the value was 0.949. These high correlations indicate good agreement between the flash rate trends found in the hand analyses versus those produced by the statistical framework. For CG flashes during the 29 June event, +CG correlation was 0.942 and -CG correlation was 0.756. The lower value of -CG correlation may be the result of two factors. First, few -CG flashes occurred over the storm's lifetime, so one missed flash could result in a much lower correlation than for +CG flashes. Second, T05 and W05 both included the storm to the northeast that existed in the beginning of the supercell's lifetime into their flash counts. For the statistical framework, it was simple to include only those CG flashes that occurred in the supercell itself, which was done. The timing of the extra -CG flashes could account for the lower correlation value. Overall, though, especially for +CG, these values indicate again the framework's ability to detect trends found in hand analyses.

The largest issue with the lightning data was with the initiation point of CG flashes (Fig. 5). In W05, it was found that there was a clear separation between the origin of +CG flashes and -CG flashes. This conclusion could not be drawn from the runs through the statistical framework. A more detailed examination involved plotting the sources of each CG with the radar cross section from that time. It was discovered that although there were some outliers that could influence the location of initiation, overall it appeared that the issue came from which sources XLMA's charge sorting program assigned to each flash. Replotting using various averaging of the first X sources (1,5,10,20,30 first sources) further confirmed that outliers were not having a large effect on the position of the flash's initiation (not shown).

Finally, dual-Doppler information was added into the file (Fig. 3, Fig. 4). Volumes of updraft and vorticity were plotted and visually compared to those in the aforementioned case studies. The updraft echo volume appeared to agree with that from the case studies. However, it was not as good a match as the HID information had been. Further, the vorticity volume as percent of storm was particularly off. Much of this could be due to the smaller cell volume produced by the tracking algorithm, as mentioned above.

4. CONCLUSIONS

Overall, the statistical framework shows the ability to produce the same trends and conclusions as those in the hand analyses, but some issues still remain, particularly with the lightning data. In particular, hydrometeor information compared very favorably with previous case studies. More issues existed with lightning data. The flash rates through the framework correlated well with those from the hand analyzed cases. The automated charge analysis misidentified some positive layers as negative, however, the automated charge analysis would still be useful in some applications, particularly for identifying the layers of positive charge, which contain more points and therefore the misidentification is less of an issue. One major problem was the lack of separation between the heights of -CG initiation and +CG initiation. Work to find the source of this issue is still ongoing.

As it continues to improve, the framework should prove very useful in processing large volumes of data to reach statistical conclusions. The large volumes of data analyzed by the statistical approach would be prohibitively time consuming following the case study approach. In the immediate future, problems presented in this paper will be further examined and corrections attempted. Beyond this, testing will continue with data from regions outside of STEPS, specifically from Norman, OK and Huntsville, AL, where both polarimetric radar observations and LMA observations are routinely made. Additionally, the SCIT and TITAN cell tracking algorithms are being prepared to examine differences in cell tracking on the front end. Environmental parameters, such as CAPE, and aerosol capabilities have also been added to the framework recently, and testing with these has been performed as well.

5. ACKNOWLEDGMENTS

This research was funded by a grant from the National Science Foundation, ATM-0649034. The statistical framework was developed by Timothy Lang. In house cell tracking algorithm was developed by Lee Nelson. The XLMA program was produced by New Mexico Tech.

6. REFERENCES

Browning, V. N., and G. B. Foote, 1976: Airflow and hail growth in supercell storms and some implications for hail suppression. *Quart. J. Roy. Meteor. Soc.*, **102**, 499-533.

Deierling, W., and W. A. Petersen (2008), Total lightning activity as an indicator of updraft characteristics, *J. Geophys. Res.*, 113, D16210, doi: 10.1029/2007JD009598.

Dixon, M. and G. Wiener, 1993: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting – A Radar-based Methodology. *J. Atmos. Oceanic Technol.*, **10**, 785-797.

Goodman, and Coauthors, 2005: The North Alabama Lightning Mapping Array: Recent severe storms observations and future prospects. *Atmos. Res.*, **76**, 423-437.

Johnson, J. T., P. L. MacKeen, A. 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. Forecast.*, **13**, 263-276.

Lang, T.J., and Coauthors, 2004: The Severe Thunderstorm Electrification and Precipitation Study. *Bull. Amer. Meteor. Soc.*, **85**, 1107–1125.

Lang, T. J., and S. A. Rutledge, 2008: A statistical framework for the analysis of large lightning and radar datasets. *3rd Conference on the Meteorological Applications of Lightning Data*, American Meteorological Society, New Orleans, LA.

Lang, T. J., and S. A. Rutledge, 2009: Further progress on a framework for the statistical analysis of large radar and lightning datasets. *4th Conference on the Meteorological Applications of Lightning Data*, American Meteorological Society, Phoenix, AZ.

Rison, W., R. J. Thomas, P. R. Krehbiel, T. Hamlin, and J. Harlin (1999), A GPS-Based Three-Dimensional Lightning Mapping System: Initial Observations in Central New Mexico, *Geophys. Res. Lett.*, 26(23), 3573–3576.

Tessendorf, S.A., L.J. Miller, K.C. Wiens, and S.A. Rutledge, 2005: The 29 June 2000 Supercell Observed during STEPS. Part I: Kinematics and Microphysics. *J. Atmos. Sci.*, **62**, 4127–4150.

Tessendorf, S. A., K.C. Wiens, and S.A. Rutledge, 2007: Radar and Lightning Observations of the 3 June 2000 Electrically Inverted Storm from STEPS. *Mon. Wea. Rev.*, **135**, 3665–3681.

Wiens, K.C., S.A. Rutledge, and S.A. Tessendorf, 2005: The 29 June 2000 Supercell Observed during STEPS. Part II: Lightning and Charge Structure. *J. Atmos. Sci.*, **62**, 4151–4177.