2012
22nd International Lightning Detection Conference
2 – 3 April • Broomfield, Colorado, USA
4th International Lightning Meteorology Conference
4 – 5 April • Broomfield, Colorado, USA

EVALUATION OF AN IMPROVED STORM CELL IDENTIFICATION AND TRACKING (SCIT) ALGORITHM BASED ON DBSCAN CLUSTERING AND JPDA TRACKING METHODS

Jenny Reed[1] and John Trostel[1]
Georgia Tech Research Institute, Severe Storm Research Center

ABSTRACT

Storm cell identification and tracking (SCIT) is a highly challenging problem with many potential applications. The storm cell identification and tracking algorithm based on density based spatial clustering with applications in noise (DBSCAN) and joint probabilistic data association (JPDA) is an advanced algorithm that has been proposed as a solution to this problem. It is designed as a method to identify and track storms accurately enough to spatially associate lightning strikes and temporally correlate lightning trends to storm cells and associated meteorological phenomena, particularly tornado genesis.

A brief discussion of the DBSCAN and JPDA SCIT algorithm will be given. The main focus of the paper will be an evaluation of the performance of this algorithm with a comprehensive set of proposed metrics. The storm cell identification and tracking algorithm is applied to a diverse set of reflectivity data cases. The performance of the algorithm is then scored based on the proposed set of metrics.

1.0 INTRODUCTION

Identifying and tracking storm cells, utilizing reflectivity data from the current operational weather radars, is a current topic of concern in the meteorological community. Numerous algorithms have been proposed to address this challenging problem (Dixon and Weiner, 1993; Han et al., 2009; Johnson et al., 1998; Lakshmanan, 2008; Matthews and Trostel, 2010; Reed and Trostel, 2011). The SCIT algorithm proposed by Matthews and Trostel (2010) is one such novel algorithm designed to address the SCIT problem. An evaluation of the tracking portion of this algorithm is presented in this paper.

1.1 PROBLEM STATEMENT

An evaluation of the tracking portion of the SCIT algorithm proposed in Matthews and Trostel (2010) and Reed and Trostel (2011) is presented. Unfortunately, there is no current standard for evaluating storm cell tracking algorithms. For this reason, previously proposed methods (Johnson et al., 1998; Lakshmanan, 2011) for scoring storm cell tracking algorithms are scrutinized to select a fair set of scoring metrics. Furthermore, a description and justification of the proposed scoring metrics are given. A summary of these metrics shall be used to draw conclusions regarding the performance of the algorithm under test.

1.1 ALGORITHM DESCRIPTION

The algorithm under test is designed for the purpose of tracking storm cell properties, specifically lightning activity, over the lifecycle of storm cells. It is composed of two separate parts: an identification algorithm and tracking algorithm. While only tracking performance is evaluated in this paper, a brief discussion of the identification algorithm is given since this method is used to identify the storm cells that are utilized in the tracking evaluation.

[1] Severe Storms Research Center, Georgia Institute of Technology, Smyrna, GA 30080; email: {jenny.reed,john.trostel}@gtri.gatech.edu

1.2 STORM CELL IDENTIFICATION

The storm cell identification algorithm processes volumetric reflectivity data from the NEXRAD weather radar. Storm cells are distinguished from within lower reflectivity regions by using a multi-threshold, three-dimensional, grid-based DBSCAN algorithm. For complete details of the identification algorithm please see Matthews and Trostel (2010) and Reed and Trostel (2011).

1.3 STORM CELL TRACKING

The storm cell tracking algorithm under test is described in detail in Matthews and Trostel (2010) and Reed and Trostel (2011). It is a particle based tracking algorithm that solves a series of JPDA problems. First, a path coherence cost function is minimized with respect to particle locations. Each particle belongs to the storm cell in which it lies. These particle assignments are used to create a probability model for possible storm cell associations. This probability model is then used to minimize a probability cost.

There are two main intended advantages of this algorithm. The first advantage is that it is a non-centroid based tracking algorithm, unlike other storm cell tracking algorithms. Since storm cells are constantly evolving, nonlinearities in the storm cell centroid are likely. The tracking algorithm described here attempts to find linear paths for the storm cell particles rather than the storm cell centroids. The second advantage of the proposed tracking algorithm is its ability to associate a single storm cell at a given time with multiple storm cells at the next time, making it feasible to track two branches of a storm cell that were once part of a single track.

2.0 PREVIOUS WORK

A brief discussion of past storm cell identification and tracking algorithms is given. Further, an assessment of methods that have been previously suggested for scoring storm cell tracking algorithms is given.

2.1 PREVIOUS STORM CELL IDENTIFICATION AND TRACKING ALGORITHMS

A number of other storm tracking algorithms have been proposed. The most well-known of these algorithms include TITAN Dixon and Weiner, (1993), ETITAN (Han et al., 2009), and SCIT (Johnson et al., 1998). While these algorithms have significant differences, all use centroid based tracking techniques.

For an evaluation of the proposed tracking algorithm, the Johnson et al. SCIT algorithm is used as a benchmark for comparison. The Johnson et al. (1998) SCIT algorithm uses a least squares method for estimating storm cell velocity and a simple radial search method as a basis for storm cell associations. This algorithm is chosen as a benchmark for multiple reasons. First, it was included in the WSR-88D build 9.0 of Radar Products Generator Software (Johnson et al., 1998); also, it is extremely well-documented for implementation purposes.

2.2 PREVIOUSLY PROPOSED SCORING METHODS

There are two main scoring methodologies that have been used or suggested for the purpose of evaluating storm cell tracking algorithms. The "Percent Correct" method is used by Johnson et al. (1998) to evaluate the tracking performance of SCIT in a select set of reflectivity cases. The second technique, proposed by Lakshmanan (2011), computes a set of bulk statistics without any knowledge of what the correct tracks are.

Johnson et al. evaluates tracking performance utilizing a single metric, the "percent correct", which Johnson et al. describes as "simply the ratio of correct time associations divided by the total number of correct time associations. This metric has the advantage of having a number with a known significance; i.e., no interpretation is required to determine the implication of the metric with respect to the algorithm performance. This is, however, an incomplete metric. While conveying the number of correct versus incorrect associations made, the given metric does not address the number of missed associations. Furthermore, it does not completely communicate how the algorithm performance is affected, e.g., how long the algorithm is typically able to hold accurate tracks. Lastly, one major drawback to this methodology of scoring is that it requires a subjective, labor-intensive, manual labeling of what constitutes a correct versus incorrect association.

As a remedy to the subjective and labor intensive requirements of the "percent correct" method utilized by Johnson et al. (1998), Lakshmanan (2011) proposed a set of metrics based on bulk statistics that surpasses the requirement for knowing what a correct association is. The obvious advantages of this method are that it does not require labor-intensive human evaluation and it does not introduce the element of human error or subjectivity into the scoring process. For these reasons, Lakshmanan (2011) proposes a set of three metrics:

1. $\widetilde{Dur}$: This is the median duration, in seconds, of all tracks formed by a given tracking algorithm when applied to a given set of reflectivity data over a specified period of time.
2. $\overline{\sigma_{VIL}}$: This is the average standard deviation of the vertically integrated liquid (VIL) for all tracks with a duration greater than or equal to the median duration. In other words, a VIL standard deviation is computed for each track with a duration greater than $\widetilde{Dur}$. The mean of these values is then reported.
3. $\overline{E_{XY}}$: This is the average root mean square of the linearity error over all tracks with duration greater than or equal to the median duration. In other words, linearity error is computed for each track with a duration greater than or equal to $\widetilde{Dur}$. The mean of these values is reported.

The assumptions that are used as a basis for choosing these metrics, respectively, are:

1. In general, a better tracking algorithm should be able to maintain more tracks of longer duration.
2. In general, incorrect associations create a mismatch in VIL creating an increase in VIL standard deviation along each track.
3. In general, correct tracks shall be more linear.

There are, however, a number of problems with the assumptions that have been made to come up with the given set of metrics. First, the given metrics are correlated. This is a problem for two reasons. First, as track duration increases, $\overline{\sigma_{VIL}}$ and $\overline{E_{XY}}$ will have a natural tendency to increase, even if evaluating only valid tracks, i.e., tracks that do not have any erroneous associations. This makes it very difficult to determine, as longer tracks are created, if $\overline{\sigma_{VIL}}$ and $\overline{E_{XY}}$ are increasing because the algorithm is able to keep a longer valid track or is creating longer invalid tracks. Secondly, it is suggested that the given metrics should only be computed utilizing statistics from tracks greater than the median duration. Since these metrics are correlated, the only way to fairly compare $\overline{\sigma_{VIL}}$ and $\overline{E_{XY}}$ for two different algorithms are if they are over tracks of similar duration, meaning that the algorithms being compared must yield the same median duration and a similar number of tracks above this median duration. The use of $\overline{\sigma_{VIL}}$ as a metric is also troublesome due to the fact that VIL is not a stationary process. VIL is often used as an indication of storm cell severity since the two are correlated. This means that VIL is expected to increase as a storm cell grows in severity and decrease as a storm cell decays. As a result, a significant amount of deviation in VIL over the life cycle of a storm is expected. The increased deviation of VIL in a given track that would result from an improper storm cell association would be insignificant compared to the overall expected deviation of the VIL. Therefore, it is unlikely that any significant conclusions regarding the validity of storm cell tracks can be made using this metric. Lastly, centroid linearity error, $\overline{E_{XY}}$, is a poor metric as well because storm cell tracks can often curve over time creating nonlinear tracks. Furthermore, even if a storm cell does follow a linear track, in general, nonlinearities in the storm cell centroid location are still likely to occur due to the nature of storm cell evolution and identification. Lastly, the tracking algorithm under test (Matthews and Trostel, 2010; Reed and Trostel, 2011) is specifically designed to cope with the nonlinear motion of storm cell centroids, while the algorithm used for comparison is designed on the sole idea of minimizing linearity error of storm cell centroids over a track. This makes linearity error a very unfair metric in this particular comparison.

3.0 SCORING METHODOLOGY

The proposed scoring methodology utilizes several metrics to infer tracking performance of a storm cell tracking algorithm. It requires subjective human based decisions, similar to the "percent correct" method of Johnson et al. (1998), but is slightly less labor intensive. It also a more complete set of metrics, allowing interpretation of how missed associations and bad associations effect the expected duration over which a valid track may be kept. The proposed scoring metrics are as follows:

1. N_T: Total number of tracks. This is the total number of tracks (lasting for at least three volume scans of the radar) identified by a tracking algorithm in a given set of reflectivity data over a specified period of time.
2. N_{EE}: Number of tracks terminated early. This is the number of tracks that ended before it should have; e.g., the tracking algorithm missed an association for a track and terminated a track before a storm cell dissipated.
3. N_{SL}: Number of tracks that started late. This is the number of tracks that started after it should have; e.g., the tracking algorithm missed an association for a track and did not start tracking a storm cell when it first formed.
4. N_{IA}: Number of tracks containing an incorrect association. This is the number of tracks that have improperly connected two separate tracks via an incorrect association.
5. N_P: Number of perfect tracks. This is the number of tracks that did not start late, did not terminate early, and do not contain an incorrect association.
6. $\widetilde{Dur}$: Median duration of valid tracks. This is the median duration over all tracks that are considered to be valid, i.e., any track that contains all valid associations and no incorrect associations.
7. Dur_T: Total duration of valid tracks. This is the sum of the durations of all tracks that are considered valid.

All of the described metrics can be derived based on the following procedure:

1. Each identified track is plotted one at a time. All storm cells that exist in the volume scan prior to the start of the track are plotted as well as shown in Fig. 1a. Based on this plot, a human may make a subjective decision as to whether a storm cell identified prior to the start of a track should have been part of that track, thus determining if a track started late.
2. Each identified track is plotted one at a time. All storm cells that exist in the volume scan after the end of a track are plotted as well as shown in Fig. 1b. Based on this plot, a human may make a subjective decision as to whether a storm cell identified after a track terminated should have been part of that track, thus determining if a track terminated early.
3. Each identified track is plotted one at a time as shown in Fig. 1c. Based on the locations of those storm cells, a human may make a subjective decision as to whether that track contains any incorrect associations.
4. The number of perfect tracks may be derived based on the tracks that were determined to have started late,

terminated early, and contained an incorrect association.

5. The duration of each track is recorded. Any track that does not contain an incorrect association is used to determine the median duration of valid tracks.

While this methodology requires subjective and labor intensive human labeling, it has a number of advantages. It is less labor intensive than the “percent correct” method since it requires the human to make a set of decisions for each track rather than for each individual storm cell association. Furthermore, it conveys a more complete set of metrics than the other two previously proposed methods giving metrics directly related to correctly-made associations, incorrectly-made associations, and missed associations. Also, these metrics are not designed in a way that they would favor any particular algorithm.

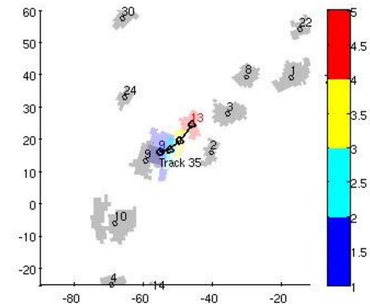
4.0 RESULTS

Results are given for three different reflectivity cases. Each case represents a different reflectivity scenario. The chosen reflectivity sets are chosen because they are used as benchmark reflectivity sets in one or both of the references for previous scoring metrics (Johnson et al., 1998; Lakshmanan, 2011).

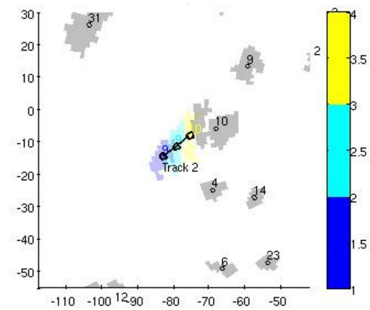
These cases include:

1. KMLB 3/25/1992 1600-2400 GMT (Mesoscale Convective System)
2. KLSX 6/8/1993 1600-2400 GMT (Convective Line)
3. KTLX 2/21/1994 1600-2400 GMT (Stratiform Event)

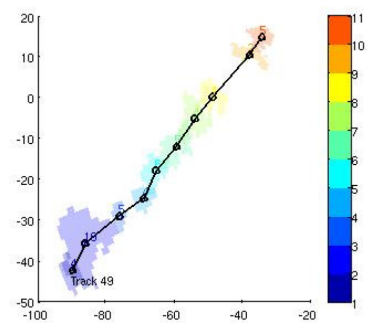
For each reflectivity set, the baseline Johnson et al. SCIT tracking algorithm is run over a sweep of values (30, 50, 70, 90 m/s) for the algorithm parameter “Correlation Speed”, i.e., the single variable parameter in the SCIT tracking algorithm that may affect tracking performance. The tracking algorithm under test is run using a single set of default parameters. The tracking metrics are computed for each of these cases and compared.



(a)



(b)



(c)

Figure 1. Example plots used to determine storm cell track statistics. The track of interest is plotted where color represents time. (a) Example plot of a track that starts late. Storm cells at the volume scan prior to the start of the track of interest are shown in black. (b) Example plot of a track that terminates early. Storm cells at the volume scan after the end of the track of interest are shown in black. (c) Example track containing an incorrect association.

The results for the mesoscale convective system, convective line, and stratiform event are shown in Tables 1, 2, and 3, respectively.

In the case of the algorithm under test, two values are reported for most metrics in the given tables because this particular tracking allows splitting of tracks. Where two numbers are given, the first number scores across all tracks. However, this number is not a very accurate representation because if a track splits into two separate tracks, then the statistics of that track are doubly counted. For example, suppose a track starts late and splits into two separate branches. This is a single case where a track started late, but it is counted as two late starts because each branch of the original track is counted as a separate track. The second number reported for each metric in the tables is computed by removing the “double count” that results from splitting. Therefore, the second metric is comparable to those reported for the Johnson et al. SCIT tracking algorithm.

5.0 CONCLUSIONS

The metrics chosen to evaluate and compare the algorithm under test with the baseline SCIT algorithm convey the capabilities of both algorithms to create and maintain tracks. From the given metrics, a significant performance comparison can be made. Like the “percent correct” method used in Johnson et al., (1998), it requires subjective and labor intensive human labeling; however, the scoring method proposed in this paper slightly decreases the

Algorithm / Parameter	SCIT 30	SCIT 50	SCIT 70	SCIT 90	Reed
# Tracks	85	126	126	130	137/117
# Early Term.	47	52	29	25	14/9
# Late Start	50	58	36	35	24/21
# Bad Assoc.	3	3	5	6	4/4
# Perfect	13	46	70	79	100/87
Median Duration	897	898	1196	1196	1198/1198
Mean Duration	1424	1674	1807	1770	2104/2268
Total Duration	109,682	205,963	218,615	219,513	279,823/278,932

Table 1 Tracking Results for Mesoscale Convective System (KMLB 3/25/1992)

Algorithm / Parameter	SCIT 30	SCIT 50	SCIT 70	SCIT 90	Reed
# Tracks	204	322	354	352	444/357
# Early Term.	82	95	74	48	44/30
# Late Start	86	91	70	50	44/33
# Bad Assoc.	0	1	2	13	5/1
# Perfect	76	163	22	249	354/296
Median Duration	703	1050	1052	1079	1077/1066
Mean Duration	1029	1228	1423	1609	1623/1570
Total Duration	209,942	394,231	501,048	546,756	712,680/555,802

Table 2 Tracking Results for Convective Line (KLSX 6/8/1993)

Algorithm / Parameter	SCIT 30	SCIT 50	SCIT 70	SCIT 90	Reed
# Tracks	57	85	92	99	103/81
# Early Term.	32	23	19	25	6/4
# Late Start	26	22	16	25	11/4
# Bad Assoc.	0	0	0	2	0/0
# Perfect	12	47	63	60	86/73
Median Duration	868	868	1048	1049	1745
Mean Duration	1134	1382	1602	1653	2078
Total Duration	64,611	117,456	147,391	160,314	168,347

Table 3 Tracking Results for Stratiform Event (KTLX 2/21/1994)

labor intensiveness of the human labeling process and results in a more complete set of metrics.

The metrics shown in the tables above suggest that the algorithm under test performs comparably, if not significantly better, than the baseline SCIT algorithm. Overall, the algorithm under test yielded longer valid tracks, fewer late starts, fewer early terminations, fewer incorrect associations, and more perfect tracks. As a result, it is concluded that non-centroid based tracking yields a significant improvement over other tracking algorithms that used centroid-based tracking. An example track, from the tested reflectivity cases, where non-centroid based tracking proved particularly advantageous is shown in Fig. 2.

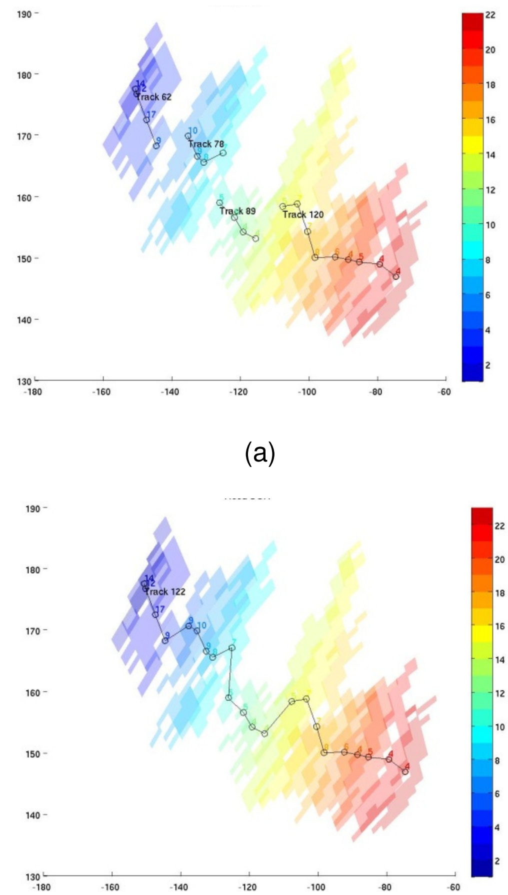
The improved tracking performance yielded by the non-centroid based tracking technique evaluated in this paper is exploited in other research to track lightning densities and other storm cell features over the life cycle of various storm cells that yield tornadoes or other severe weather. Lightning densities and other storm cell attributes may then be accurately correlated with severe weather phenomena.

ACKNOWLEDGEMENTS

We gratefully acknowledge useful discussions with Valliappia Lakshmanan regarding his previous work on scoring of storm cell tracking algorithms.

REFERENCES

Dixon, Michael, and Gerry Weiner, 1993: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-based Methodology. *Journal of Atmospheric and Oceanic Technology Vol. 10 No. 6,* 785-797.



(a)

(b)

Fig. 2 (a) Example where the baseline SCIT algorithm fails to maintain a consistent track due to nonlinearities in the storm cell centroid motion. (b) The algorithm under test is able to maintain the storm cell track during the storm cell life cycle despite the centroid nonlinearities. Note that storm cell color is used to represent time over the track duration.

Han, Lei, Shengxue Fu, Lifeng Zhao, Yongguang Zheng, Hongqing Wang, and Yinjing Lin, 2009: 3D Convective Storm Identification, Tracking, and Forecasting—An Enhanced TITAN Algorithm. *Journal of Atmospheric and Oceanic Technology, Vol. 26, No. 4*, 719-732.

Johnson, J.T., Pamela L. MacKeen, Arthur Witt, E. DeWayne Mitchell, Gregory J. Stumpf, Michael D. Eilts, and Kevin W. Thomas, 1998: The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm.

Weather and Forecasting Vol. 13, 263-276.

Lakshmanan, V., K. Hindl and R. Rabin, 2008: An Efficient, General-Purpose Technique for Identifying Storm Cells in GeoSpatial Images.

Lakshmanan, V., and T. Smith, 2011: Evaluating a Storm Tracking Algorithm. *Proceedings of the Fifth American Meteorological Society Conference on the Meteorological Applications of Lightning Data*, Seattle, WA 23-27 January 2011.

Matthews, J. and J. Trostel, "An Improved Storm Cell Identification and Tracking (SCIT) Algorithm based on DBSCAN and JPDA Tracking Methods," *Proceedings of the 21st International Lightning Meteorological Conference*, Orlando, FL, 19-22 April 2010.

Reed, J. and J. Trostel, "A 3-Dimensional DBSCAN Storm Tracking and Identification Algorithm – Description and Implementation," *Proceedings of the Fifth American Meteorological Society Conference on the Meteorological Applications of Lightning Data*, Seattle, WA 23-27 January 2011.