TRACKING THUNDERSTORM CELLS USING LIGHTNING DENSITY INFORMATION

746

V. Bourscheidt*, O Pinto Jr., K. P. Naccarato

Atmospheric Electricity Group

Brazilian National Institute for Space Research

1. INTRODUCTION

Severe weather has been noticed as an increasing environmental hazard in the last years. In the same way, pronounced improvements have been made in the forecast and nowcast of extreme weather by integrating the meteorological information with lightning data.

Several systems to track thunderstorms are available these days, including useful tools to the government and civil defense agencies. They use algorithms based on radar and satellite information and are able to assimilate data from Lightning Location Systems (LLSs) to improve the resulting information. Among the most commons tracking systems are (Dixon and Wiener, 1993; Lakshmanan et al, 2007):

- TITAN: integrates basically satellite/ radar information to track the convective activity. Lightning data integration is also available.
- WDSS II: Worldwide used tracking system. Lots of information can be used, including lightning data.

Only few tracking systems are based uniquely on lightning data, mainly due the limited number of events in a timestamp. Some commercial systems (e.g. Vaisala Inc and Earth Networks) as well as other researches focusing on lightning data are available. Examples can be found in the USA (Lojou and Cummins, 2006), Europe (Betz et al, 2008), and more recently in South Africa (Grant, 2010), and Colombia (Lopes et al, 2011). With the advance LLSs in the detection of total lightning (intracloud and cloud-to-ground), more information became available, allowing the use of point pattern analysis (PPA) tools to create a suitable spatial/temporal window to track the thunderstorm development.

* *Corresponding author address:* V. Bourscheidt, Brazilian Nat. Inst. For Space Res., Atm. Electricity Group ELAT, Sao Jose dos Campos, 1758, Postal Code: 12227-010; e-mail: vandoir@gmail.com.

The proposed system is based on a kernel density (somewhat similar to the method used by Betz et al (2008)), using contours (polygons and centroids) and some spatial (Boolean) criteria - overlapping areas, merge and splitting conditions. The use of density information regardless to other cluster methods (DBSCAN, K-means, mean shift etc.) to aggregate lightning belonging to the same thunder cell reduces the noise and allows the use of different statistical information about the density distribution inside each polygon.

The use of lightning information become essential for countries like Brazil were the radar/satellite spatial/temporal coverage is limited. Since the total lightning network is being expanded over the country, the algorithm might be applied for wide areas, and could improve the civil defense warning systems.

2. THE TRACKING SYSTEM

2.1 Contours of Lightning Activity

The first step to obtain a suitable tracking algorithm based on lighting data is to define the method used to separate individual storm cells. The use of clustering methods like k-means, mean shift and DBSCAN (acronym for Density-Based Spatial Clustering of Applications with Noise) is very common in data mining tools (Ester et al, 1996). The two first methods require the user to define the number of clusters in a window (Comaniciu and Meer, 2002, Hartigan and Wong, 1979), which is not suitable for stochastic events like lightning.

The last mentioned method is based on the separation (distance) of the events and could be applied to track lightning activity. However, this method needs a good separation of the data (dots) to provide realistic clustering.

In this way, density information might be a better alternative, especially because the number of thunder cells is based on density contours, which are independent of any clustering method and more likely precisely separated. Reinforcing the proposed method, recent works suggest that total lightning activity could be a proxy for precipitation estimation when analyzed using density information in a log-scale, providing results that are similar to radar data (Liu and Heckman, 2012). Also, density information was already used in other tracking systems (e.g. Betz et al, 2008). Thus, the use of lightning density information to track thunderstorm activity seems fairly reasonable.

In our case, polygons resulting from lightning density contours within each time step are used to track thunderstorm activity. Figure 1 is divided in three steps showing how contours (polygons) are obtained from the lightning data. The polygons are created based on lightning data grouped in 10 min intervals, which are then evaluated over a grid. The gridded data (2D histogram) is weighted through a quasi-log scale and, afterwards, smoothed using a Gaussian kernel. The resulting polygons are then evaluated for a minimum area and only polygons larger than the threshold are used in the tracking algorithm.

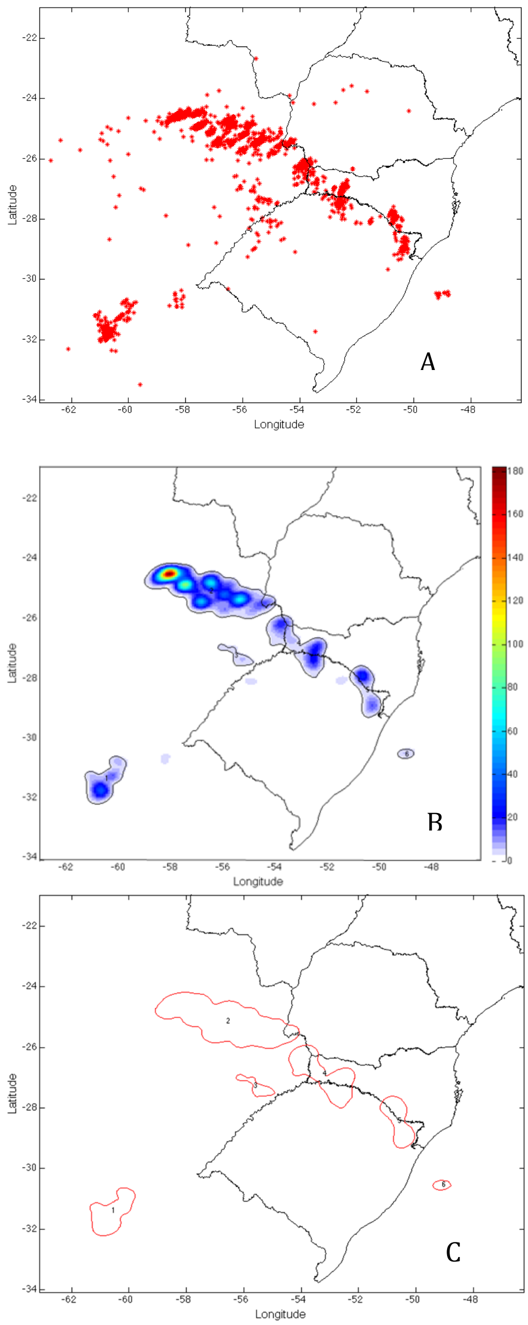

Figure 1 – Steps to generate polygons from lightning data (A to C). Process include a quasi-log weighting function and a Gaussian kernel smoothing. Resulting polygons have a minimum area and an associated ID.

2.1 Tracking thunderstorms

The literature (as mentioned in the Introduction) shows some different techniques to track thunderstorm using lightning data, i.e., based on density contours, cluster centroids, etc. that are within a time step and pass some predefined spatial and temporal criteria. Some techniques are even based exclusively on the individual discharge using overlapping time windows (Strauss and Stephany, 2011).

Our algorithm mixes with previous lightning-based tracking and others algorithm used with radar/satellite images. Most of the criteria used here are grounded on TITAN (and other algorithms that use similar idea), which is a sophisticated tracking system based on overlapping polygons. The idea is to use overlapping areas together with some information exclusive of lightning data, as well as to include some new features to improve nowcasting of hazards events.

2.1.1 Identifying thunderstorm cells

Once the contours within each time step are defined, the main idea of the tracking algorithm takes place: define consecutive regions of the same storm cell and track the evolution of this cell. Like for TITAN, different criteria will exist for different state/conditions of the cells being tracked. In the basic case of cells belonging to the same convective system, the criterion is based on the overlap of two consecutive polygons (different time interval). Figure 2 summarizes two cases: same cell and new cell.

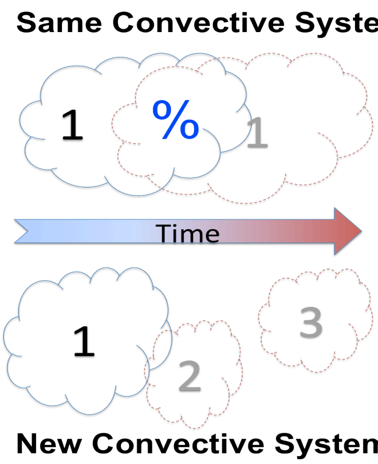

Figure 2 – Conditions to determine whether a polygon is part of a preceding storm or a new cell. In the first case, the previous ID is used. Otherwise, a new ID is created.

If the overlap percentage is over a defined threshold, the new polygon is considered as a part of the previous one (get the same ID). On the other hand, if there is no overlap among the consecutive polygons, the last polygon is considered as a new convective system with a new ID associated.

2.1.2 Merge and Split conditions

In many cases, however, several polygons overlap within the same timestamp, and in other cases the new polygons are inside the previous one. To deal with such scenarios, other conditions must be considered, including split and merge criteria. In our case they are treated in the following way (see Figure 3):

- Splitting Systems: A previous polygon results in more than one new polygon. New polygons get the ID of previous polygon.
- Merging Systems: More than one previous polygons result in one new polygon. New polygon gets the ID of the larger polygon or cell.

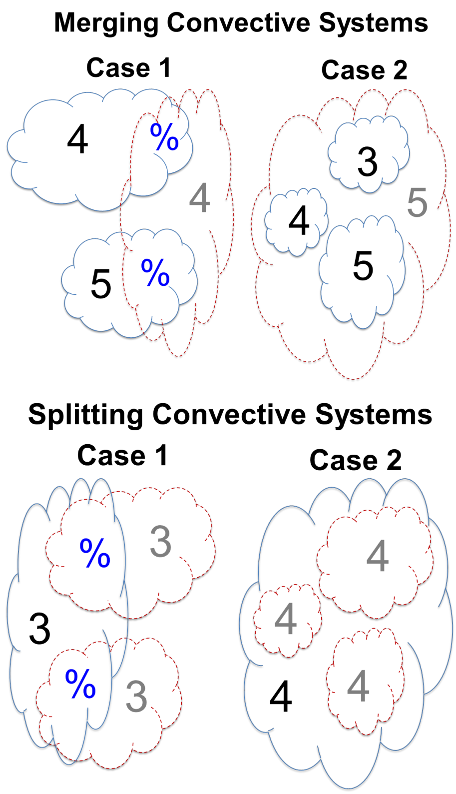

Figure 3 – Conditions to determine whether a polygon is splitting or merging. The ID of the largest previous polygon is kept for the next time step.

Few more criteria can be included in those cases to make sure that the convective systems are splitting/merging, like in other existing tracking systems. The discussed methodology is wide used in the literature, with some small differences in each system.

Together with this approach, some innovative ideas are included in our tracking, to take advantage of the most recent GIS tools available on the internet, as well as on the lightning data itself. These ideas are discussed in the following section.

3. NEW IDEAS

3.1 Multi-level Tracking

Usually thunderstorms behave differently depending on the scale of analysis: individual cells generally move in several directions while the whole convective system (CS) moves in one preferred direction. Thus, based on fractal analysis inside some public web apps (e.g. Google Maps) and on the ideas of hierarchical clustering, the tracking of convective systems may be carried out on different levels. Figure 4 shows a hierarchical clustering tree used to show how different cells (each polygon) aggregate according to the different (Euclidian) distances.

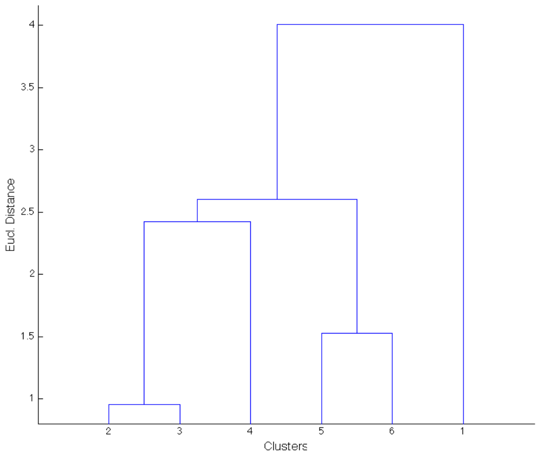

Figure 4 – Example of the hierarchical clustering tree of one time step including six polygons (clusters) in the base level.

Using the resulting clusters, the storms might be presented in several levels. In the example of Figure 5, two levels are shown: the base level, including the polygons of the actual track, and the fourth level, showing two arrows. The eastward arrow aggregates average location and azimuth from polygons 2 to 6.

3.1 Direction of Displacement

One interesting aspect of lightning data is that its time scale is independent of the instrumentation. Thus, contrary to radar and satellite images, which are limited by the temporal resolution, lightning data can be used almost continuously, giving short time information about the convective systems displacement.

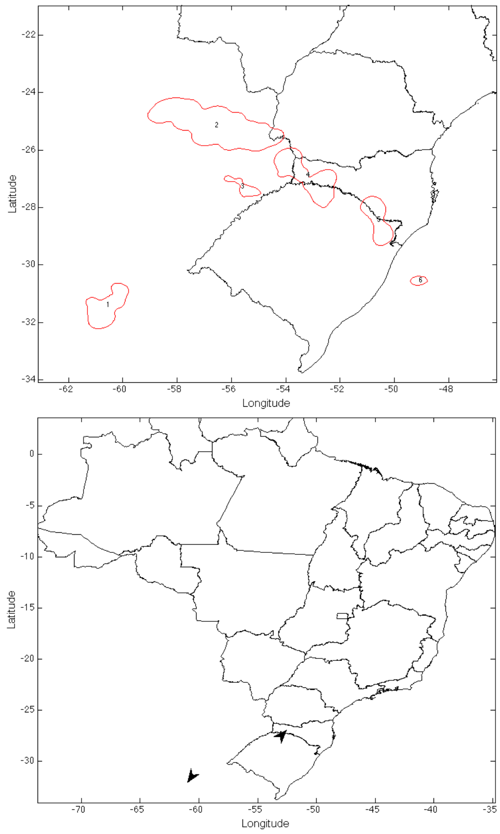

Figure 5 – Example of the hierarchical clustering showing the base level, with all the polygons, and the fourth level, with two arrows and the eastward arrow aggregating polygons 2 to 6.

On the proposed tracking system, data of each interval (time step) is separated into two time groups. The average geographic location (latitude/longitude) of the discharges of each time group is then used to create a displacement vector.

The displacement vector is used to define the weighted buffer. To produce this buffer, the initial polygon is shifted by a constant factor. Then, the vector is used to weight the buffer so that no buffering is created in the direction where the vector starts and the maximum buffer size is produced in the direction where the vector is pointing to.

The resulting expected displacement direction is given by the (weighted) polygon produced. The factor that defines the shift of the buffer, despite constant so far, it might be defined based on the distance between two time steps of data. In the example of Figure 6, a constant distance of 0.1° is used.

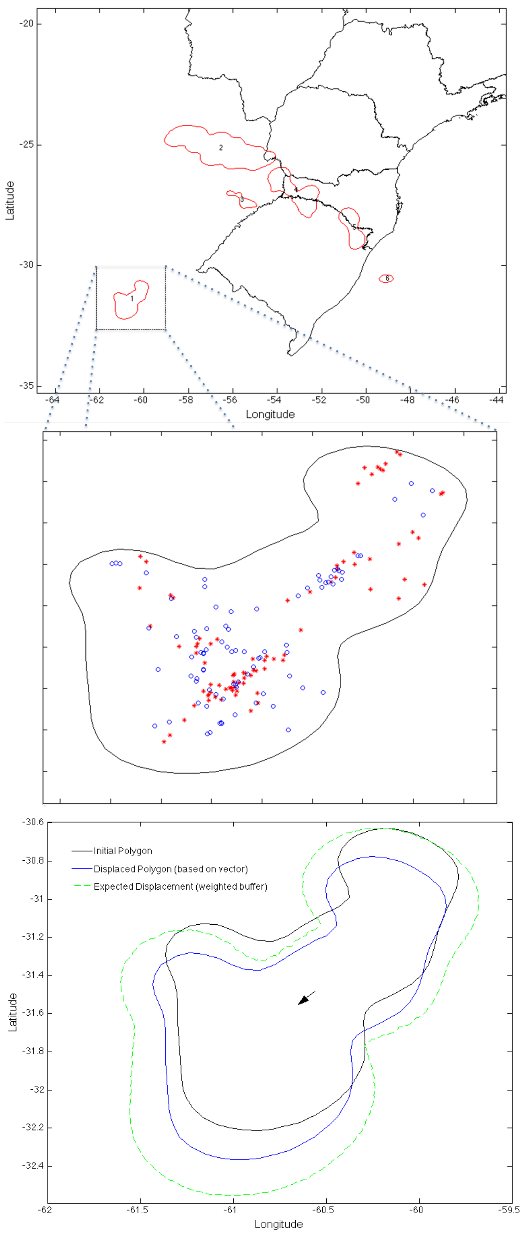

Figure 6 – Example of the convective system displacement showing the vector obtained with the discharges (black arrow). The vector results from the average location of red dots (time t) and blue dots (time t + Δt). The expected propagation of the system is indicated by the dashed green polygon.

4. INITIAL RESULTS

The panel on Figure 7 shows a short track example for October 25, 2011 based on data from the Brazilian total lightning network BrasilDAT. Five time steps (15 minutes, starting at 00:00 UTC) are used to show the evolution of the convective systems.

Satellite images for the same period are shown in the right. The images are also divided in 15 min intervals. However, lightning the lightning contours and satellite images are not time-aligned (lightning data was not centered, i.e., -7.5min to +7.5min, to coincide with satellite images).

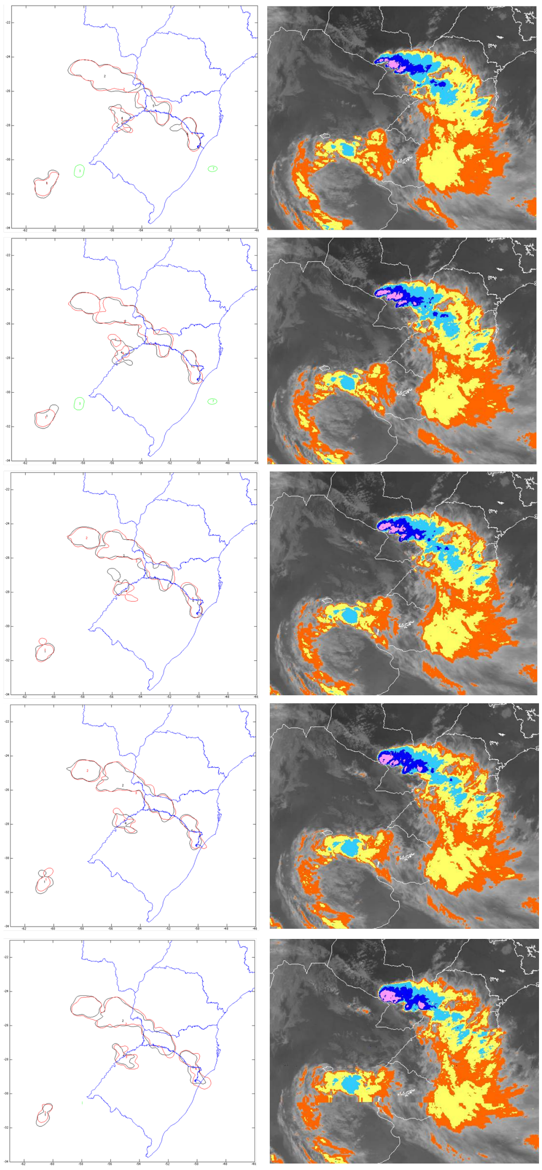

Figure 7 – Panel showing a short track (five time steps) in the left, starting at 00:00 UTC on October 25, 2011. In the right are shown enhanced IR satellite images for the same period. The steps are in 15 minutes interval.

The comparison of the polygons obtained from the lightning density contours against the satellite images indicates that the interval of 15 minutes is fairly representative of the thunderstorm activity. Short intervals might be used, increasing the details of the storm path and flash rate, for example.

However, even though the results seem satisfactory, some adjustments are still necessary to better deal with the split and merge conditions, especially in the case of new cells that are inside the previous contour (not shown).

5. FINAL REMARKS

Results are preliminary. The development of the tracking is not complete and more criteria for different environment conditions (meteorological or even related to the location system) must be evaluated. Improvements on the merge and split conditions have been made, but there are still some aspects that need adjustments.

The multi-level and displacement direction features are expected to bring more practical results to the different meteorological conditions (e.g. local storms versus Mesoscale Convective Systems). The use of lightning density contours and overlapping areas seems to be a good way to track thunderstorm activity, especially in the case of hazardous thunderstorms.

Some advantages of the lightning density contours overlap are observed. In relation to satellite images, for example, one can emphasize the ability to better discriminate individual cells inside large convective systems (e.g. Mesoscale Convective Systems, frontal systems etc.), which is really important for public safety. Other advantage, as mentioned in other previous works, is the spatial coverage (when compared to products using radar data) and the temporal resolution, which might be below to 10 minutes.

Finally, storm tracking based on lightning data also depends on the system performance, which should be taken into account when analyzing the evolution of the flash count inside a thunderstorm, for example. This is even more important when the severity in terms of flash rate is used with nowcasting purpose.

ACKNOWLEDGEMENTS

This research is sponsored by FAPESP (São Paulo Research Foundation), under the grant n. 2012/07586-8.

REFERENCES

Betz, H. D., and Schmidt, K., Oettinger, W. P., and Montag, B., 2008: Cell-tracking with lightning data from LINET. *Adv. Geosci.,* 17: 55–61.

Comaniciu, D. and Meer, P., 2002: Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24:603–619.

Ester, M., Kriegel, H. P., Sander, J., and Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining*. Portland, OR, pp. 226-231.

Grant, M. D., 2010: A self-consistent method for the analysis of lightning stroke data sets containing misclassified strokes: the variation of lightning over southern Africa. *Ph.D. Thesis*. University of the Witwatersrand, Johannesburg.

Hartigan, J. A.; Wong, M. A., 1979: Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society*, Series C (Applied Statistics) 28 (1): 100–108.

J.A. Hartigan, 1975: Clustering algorithms. John Wiley & Sons, Inc..

Lakshmanan, V., Smith, T., Stumpf, G. J., Hondl, K., 2007: The warning decision support system - integrated information, *Weather and Forecasting*, vol. 22, no. 3, pp. 596-612,

Liu C., and Heckman S., 2011: Using Total Lightning Data and Cell Tracking in Severe Weather Prediction, *91st AMS meeting*.

Lopes, J., Peres, E., Herrera, J., Porras, L., 2011: Methodology for thunderstorm tracking using lightning location systems data in Colombia. *International Symposium on Lightning Protection (SIPDA)*. IEE USP: Fortaleza.

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

STRAUSS, C.; STEPHANY, S., 2011: Spatio-temporal clustering with a temporal sliding window for tracking of convective activity. *Workshop dos cursos de Computação Aplicada do INPE (WORCAP)*, São José dos Campos: INPE.