7th International Workshop on Climate Informatics
September 20-22, 2017
CI 2017
Hosted by the National Center for Atmospheric Research in Boulder, CO

CREATING POST-EVENT STORM TRACKS FOR SEVERE WEATHER CLIMATOLOGIES

Valliappa Lakshmanan[1,2], Benjamin Herzog[1,2], Darrel Kingfield[1,2]

***Abstract*—Commonly employed storm tracking algorithms do not use information on the subsequent positions of a storm because it is not available at the time that associations between frames are carried out, but post-event analysis is not similarly constrained. Therefore, it should be possible to obtain better tracks for post-event analysis than what a real-time algorithm is capable of. In this paper, we describe a statistical procedure to determine storm tracks from a set of identified storm cells over time. We find that this procedure results in fewer, longer-lived tracks at all scales.**

I. MOTIVATION

Even though storm tracking methods such as the Storm Cell Identification and Tracking Algorithm (SCIT [1]), Thunderstorm Identification, Tracking and Nowcasting (TITAN [2]) and Segmentation-Motion Estimation (w2segmotion [3], [4]) are constrained to work in a purely causal fashion, these algorithms have been widely employed by the meteorological research community to carry out case studies and formulate spatio-temporal relationships, for example by [5], [6], [7], [8].

Using a storm tracking algorithm that is constrained to work in real-time to carry out post-event analysis is sub-optimal. There is more information (about which cells persist and the direction in which they move) that is available if the entire set of storm cell identifications over the complete dataset is used to determine thunderstorm tracks. In this paper, we describe a way of clustering a set of storm cell identifications over time into trajectories where a trajectory is the line (or curve) that best fits the position of an individual storm cell over time.

This work was carried out in order to improve spatiotemporal relationships between radar-derived storm characteristics and the subsequent onset of specific weather hazards such as cloud-to-ground lightning, hail and tornadoes [9]. Such hazard probabilities can be derived from storm attributes using the method of [4] on a multi-year reanalysis dataset created as described in [10], but the reliability and skill of these probabilities is limited by the quality of the storm tracks used to train the data mining algorithms.

Corresponding author: V Lakshmanan, lakshman@ou.edu [1]Cooperative Institute of Mesoscale Meteorological Studies, University of Oklahoma [2]National Severe Storms Laboratory, Norman, OK

II. METHOD

Given a cluster of storm cells (x_t, y_t) at multiple times, the best constant-speed straight-line trajectory fit u, v for the cluster is the best fit slope of the line that connects the points in the cluster. [11] introduced a non-parametric, rank-invariant method for obtaining the best-fit slope in a dataset whereby one computes the median of the slopes of every pair of sample points. [12] modified the definition so that the median is computed only of points at different times ($t_2 \neq t_1$). Once the median of the slopes (u and v) are obtained, and assuming that t_0 is the time of the earliest storm cell in the cluster, the value of x_0 can be obtained by computing the median value of $x(t) - u(t - t_0)$ over all the storm cells in the cluster. This value was shown by [12] to be the value that makes the Kendall rank correlation coefficient [13] between the actual storm cell locations and the fitted values on the line approximately zero.

The clustering method we use is a variant of K-Means clustering where the cluster center is defined to be Thiel-Sen fit to the set of points in the cluster and distance between a storm cell at (x, y, t) and the cluster is defined to be the Euclidean distance between the storm cell location and the Theil-Sen estimate at that time.

The clustering method is as follows:

1) Find an initial estimate of tracks in the dataset. This can be obtained from any robust storm tracking algorithm, even a real-time one such as that of [1], [2], [4].
2) Treating each track (set of storm cells with the same id) as a cluster, compute the Theil-Sen slope and constants (u, v, x_0, y_0, t_0) for each cluster.

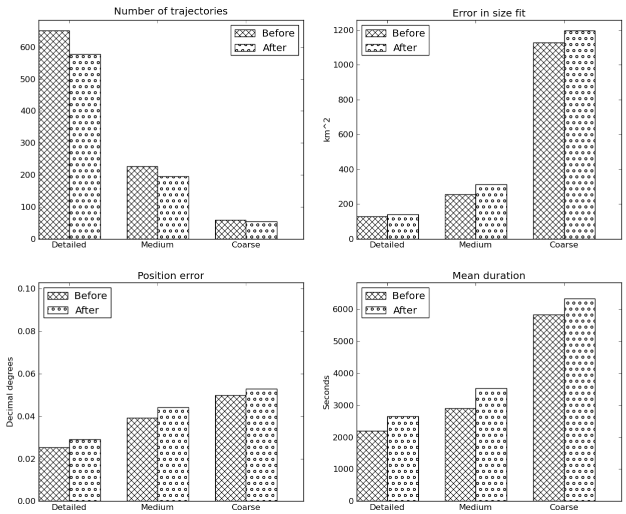

Fig. 1. Effect of clustering storm tracks at different scales.

3) For every storm cell in the dataset, find the nearest cluster. If the nearest cluster is different from the cluster the cell is currently part of, and if the distance is less than some reasonable threshold D, move the storm cell to the nearest cluster.
4) Compute the Theil-Sen fit for each cluster, prune the set of clusters to remove substantially identical trajectories and carry out Step 3, repeating steps 3 and 4 until there are no more changes or until the number of iterations reaches some maximum (we used 3 iterations as this maximum number).

III. EVALUATION

Following [14], we carried out a statistical analysis of the set of storm tracks extracted from the radar data of June 17, 2012. At the most detailed (200 km^2) scale, the number of trajectories is cut by about a third as a result of postanalysis (See Figure 1). The error in size fit (computed by fitting the sizes of the storm cells within a trajectory to a "growth-and-decay" parabola and looking for deviations from that fit – see [14] for details), an indicator of how likely it is that two separate tracks are wrongly combined, increases by a very small amount. The position error, an indicator of how likely it is that storm cells are added to tracks they are not part of, also increases but remains limited to be below the 0.1 decimal degree limit imposed by D. The fourth panel of Figure 1 demonstrates the benefit of postanalysis – the mean duration of the tracks increases by about 50%, from an average of about 2000 seconds to an average of over 3000 seconds. At the moderate (600 km^2 scale) and coarse (1000 km^2 scale), the behavior is similar. For a very small cost in terms of potentially wrong associations, one gets a significant improvement in the form of longer-lived tracks.

ACKNOWLEDGMENTS

Funding for the authors was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-OU Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce.

REFERENCES

[1] J. Johnson, P. MacKeen, A. Witt, E. Mitchell, G. Stumpf, M. Eilts, and K. Thomas, "The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm," *Weather and Forecasting*, vol. 13, no. 2, pp. 263–276, 1998.
[2] M. Dixon and G. Wiener, "TITAN: Thunderstorm identification, tracking, analysis, and nowcasting — a radar-based methodology," *Journal of Atmospheric and Oceanic Technology*, vol. 10, no. 6, pp. 785–797, 1993.
[3] V. Lakshmanan, R. Rabin, and V. DeBrunner, "Multiscale storm identification and forecast," *Atmospheric Research*, vol. 67, pp. 367–380, 2003.
[4] V. Lakshmanan and T. Smith, "Data mining storm attributes from spatial grids," *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 11, pp. 2353–2365, 2009.
[5] B. Antonescu, S. Burcea, and A. Tănase, "Forecasting the onset of cloud-to-ground lightning using radar and upper-air data in Romania," *International Journal of Climatology*, vol. 33, no. 6, pp. 1579–1584, 2013.
[6] N. Peleg and E. Morin, "Convective rain cells: Radar-derived spatiotemporal characteristics and synoptic patterns over the eastern Mediterranean," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D15, 2012.
[7] M. Kohn, E. Galanti, C. Price, K. Lagouvardos, and V. Kotroni, "Nowcasting thunderstorms in the Mediterranean region using lightning data," *Atmospheric Research*, vol. 100, no. 4, pp. 489–502, 2011.
[8] T. Weusthoff and T. Hauf, "The life cycle of convective-shower cells under post-frontal conditions," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 633, pp. 841–857, 2008.
[9] K. Kuhlman and G. Stumpf, "Probabilistic hazard information (PHI): Experiments in short-term hazardous weather information at NSSL," in *Proceedings: National Severe Weather Workshop*, 2009.
[10] J. Cintineo, T. Smith, V. Lakshmanan, H. Brooks, and K. Ortega, "An objective high-resolution hail climatology of the contiguous United States," *Weather and Forecasting*, vol. 27, no. 5, pp. 1235–1248, 2012.
[11] H. Theil, "A rank-invariant method of linear and polynomial regression analysis I, II and III," in *Nederl. Akad. Wetensch. Proc.*, vol. 53, pp. 386–392, 521–525, 1397–1412, 1950.
[12] P. Sen, "Estimates of the regression coefficient based on Kendall's tau," *Journal of the American Statistical Association*, vol. 63, no. 324, pp. 1379–1389, 1968.
[13] M. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.
[14] V. Lakshmanan and T. Smith, "An objective method of evaluating and devising storm-tracking algorithms," *Weather and Forecasting*, vol. 25, no. 2, pp. 701–709, 2010.