Evaluating a Storm Tracking Algorithm

Valliappa Lakshmanan[1,2]*, Travis Smith[1,2]

26th Conf. on IIPS; Paper 8.2

ABSTRACT

Storm tracking algorithms have been evaluated either indirectly based on fields advected using their motion vectors or by validating each association decision in a labor-intensive manner. In this paper, we introduce three bulk statistics that are measures of mismatch error, jumps and of duration. These statistics allow us to rank commonly proposed association heuristics (centroid projection, overlap, cost functions, etc.) and choose a good storm tracking approach.

1. Introduction

Algorithms that can extract properties of storm cells and track those properties over time provide information that is important to forecasters in assessing storm intensity, growth and decay (Wilson et al. 1998). Because storm tracking algorithms are a key component of nowcasting systems, the problem of how to track storms has received a lot of attention by the research community. Several criteria for associating storm cells across time have been suggested in the literature: using extent of overlap (Morel et al. 1997), using projected centroid location (Johnson et al. 1998), minimizing a global cost function (Dixon and Wiener 1993), greedy optimization of position error and longevity (Lakshmanan et al. 2009) and checking overlap followed by a global cost function (Han et al. 2009). It is important to be able to objectively evaluate these suggested techniques in order to determine which criterion or set of criteria provide the best skill.

a. *Evaluating Storm Tracking Algorithms*

One approach to evaluating storm tracking algorithms is to use the tracking algorithm to create a short-term forecast and then compare the short-term forecast with actual data (Lakshmanan et al. 2003). However, this is an *indirect* measure of storm tracking effectiveness since there is no way to separate out the effects of storm tracking from that of storm evolution. As pointed out by Wilson et al. (1998), the key reason for poor extrapolation forecasts is not errors in forecast displacement, but the growth and decay of storms in the forecast period.

A more direct way of measuring the performance of the storm tracking component of storm identification and tracking algorithms was carried out by Johnson et al. (1998). A

*Corresponding author: V Lakshmanan, 120 David L. Boren Blvd, Norman OK 73072; lakshman@ou.edu
[1]Cooperative Institutute of Mesoscale Meteorological Studies, University of Oklahoma; [2]National Oceanic and Atmospheric Administration / National Severe Storms Laboratory

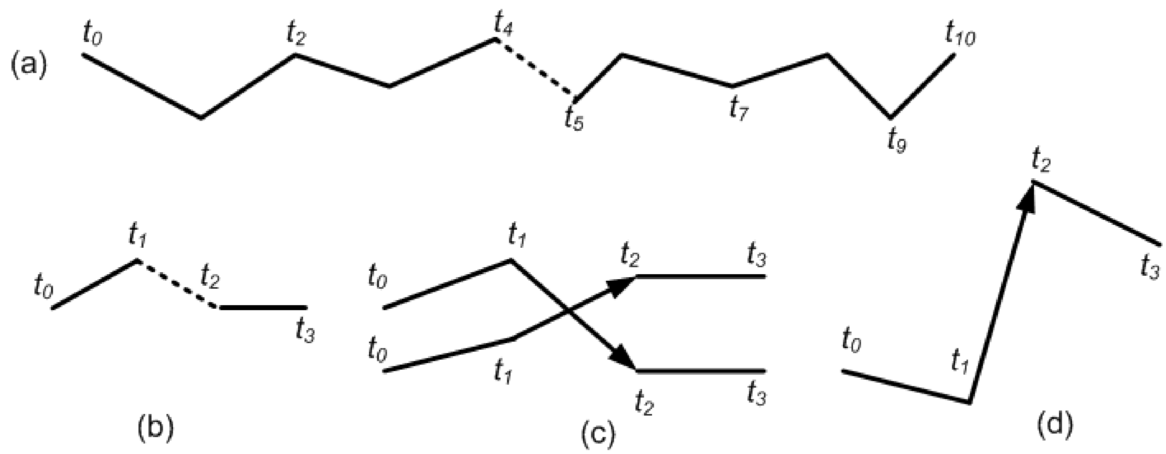

FIG. 1. Using the "percent correct" of time assocations is flawed as a way of evaluating the performance of tracking algorithms because it is an overestimate as discussed in the text and shown in (a) and is non-specific as discussed in the text and shown in (b,c,d). Dashed lines as in (a,b) indicate a "dropped" assocation while arrows indicate a wrong association which could be due to a mismatch as in (c) or due to a "jump" as in (d). Solid lines indicate a correct time association.

"percent correct" of time associations was computed by comparing the automated association of cells with a human association. This method suffers from three serious flaws: labor intensiveness, overestimation of skill (see Figure 1a) and non-specificity (See Figure 1b-d). shown in Figures 1b,c,d.

b. Storm Tracking Algorithms

The basic unit of a storm tracking algorithm is the method by which storms identified in one time frame are associated with the already labeled storms in the previous time frame – a storm that is associated with a storm in the previous time frame inherits its label (usually termed its cell ID) and its time history. A "track" consists of the locations of a storm from the time it was first assigned a cell ID to the last time at which that ID was observed.

Many heuristics have been proposed to associate storms identified at the current time frame, t_n, with storms identified at the previous time frame t_{n-1}:

i. *PRJ* (Johnson et al. 1998): Cell centroid locations at t_{n-1} are projected (PRJ[1]) to where they would be at t_n based on the position of the cell centroid at times $t_{n-k}|k > 1$. Then, each cell at t_n is assigned to the closest unassigned centroid within a certain search radius. If no centroid is close by, then the cell is given a new ID.

ii. *CST* (Dixon and Wiener 1993): A global cost (CST) function, formulated as the sum of the Euclidean distance between matched centroids and a distance metric based on some

[1]This mnemonic was not used by Johnson et al. (1998). They refered to their entire algorithm as SCIT, an acronym for Storm Cell Identification and Tracking. Because we wish to emphasize that the comparisions in this paper are carried out using a common identification algorithm (not the one in SCIT), and changing only the association algorithm, we assigned mnemonics that refer to just the association algorithm used in the various studies.

property that should be relatively consistent, is minimized. Dixon and Wiener (1993) employ the volume of the cells as this consistent property; in this paper, we'll use the area of the cells since our comparison of tracking algorithms will be on two-dimensional images.

iii. *AGE* (Lakshmanan et al. 2009): All projected cells within a size-based radius (given by $\sqrt{A/\pi}$ where A is the area of the storm) are considered "tied" in terms of position error, and such ties are resolved in favor of the longer-lived storm, i.e. based on age.

iv. *OV* (Morel et al. 1997): A storm at t_n gets the ID of the cell at t_{n-1} with which it has maximum overlap (OV) and whose ID has already not been assigned. Cells are considered in order of size, with the largest cells assigned first.

v. *OC* (Han et al. 2009): This is a combination of the OV and CST methods carried out in sequence. Cells at t_n that have 50% or greater overlap with cells from t_{n-1} are first matched. Unmatched cells are then associated using a global cost function or assigned a new ID.

vi. *NEW* (Lakshmanan and Smith 2009): A newly devised algorithm based on the evaluation technique described here and based on combining the best aspects of AGE, PRJ and CST.

We will employ the objective evaluation of storm tracking introduced in this paper to compare these heuristics on different cases.

2. Evaluation Method

We evaluate an algorithm by computing the following statistics on each track produced by that algorithm:

i. *dur* is the duration of the track. The duration is longer if there are fewer dropped assocations.

ii. σ_V is the standard deviation of the VIL of the cell in time (i.e. along a track). The σ_V is lower if there are fewer mismatches.

iii. e_{xy} is the Root Mean Square Error (RMSE) of centroid positions from their optimal line fit. The e_{xy} is lower for more linear tracks.

Central tendencies of the above statistics are computed on a large dataset of tracks:

i. $\widetilde{dur}$ is the median duration of tracks in the dataset. The better the association technique, the fewer the number of short-lived tracks that result from the technique and the greater $\widetilde{dur}$ is since the distribution of track lengths will be skewed towards longer-lived tracks.

ii. The mismatch error ($\overline{\sigma_V}$) is the mean σ_V on tracks with duration greater than $\widetilde{dur}$. Fewer mismatches are indicated by more consistent VIL values and, thus, by a lower $\overline{\sigma_V}$. This statistic is computed only on tracks with duration greater than the median duration.

iii. The linearity error ($\overline{e_{xy}}$) is the mean e_{xy} on all tracks with duration greater than $\widetilde{dur}$.

In order to perform a fair comparision of different storm tracking techniques, they were run against cells identified using the same storm identification technique with the same parameters. The techniques were evaluated on a common dataset consisting of the following WSR-88D radar data (from 18:00 UTC to 23:59 UTC on each of the days): KBIS, Bismark, ND on May 21, 1995; KCBX, Boise, ID on May 1, 1995; KIWA, Phoenix, AZ on Aug. 6, 1993, Aug. 20, 1993 and Aug. 6, 2003; KLSX, St. Louis, MO on June 8, 1993 and July 2, 1993; KLWX, Sterling, VA on Apr. 14, 1993, May 1, 1994, Oct. 6, 1995 and Oct. 6, 2005; KMLB, Melbourne, FL on Mar. 25, 1992, June 9, 1992 and June 12, 1992; and KTLX, Oklahoma City, OK on June 18, 1992 and Feb. 21, 1994. These cases are diverse geographically and in terms of the storm types. For example, they include a mesoscale convective system (KMLB, Melbourne, FL on Mar 25, 1992), a convective line (KLSX, St. Louis, MO on June 8, 1993), a stratiform event (KTLX, Oklahoma City, OK on Feb 21, 1994), isolated storms (KIWA, Phoenix, AZ on Aug 6, 1993) and a minisupercell (KLWX, Sterling, VA on Oct 6, 1995).

a. Analysis

It can be noted from the first column of graphs in Figure 2 that the mismatch error ($\overline{\sigma_V}$) is lowest when using the overlap (OV) method. The drawback of using such a conservative approach to associating cells is that the median duration of tracks is bad (white bars) in four of the five cases – only for isolated cells does the OV method have good performance on all three measures.

Similarly, it can be noted from the second column of graphs that the linearity error ($\overline{e_{xy}}$) is lowest when using the projected centroid (PRJ) method of Johnson et al. (1998). Again, this is not surprising because the centroid projection method explicitly minimizes position error after accounting for storm movement, thus emphasizing linearity at the cost of duration. Indeed, the PRJ method has bad performance in two of the five cases on the length metric.

The AGE method that was introduced "for simplicity" in Lakshmanan et al. (2009) performs surprisingly well for all cases. That method finds reasonable candidates in terms of location error and then chooses among these candidates first in terms of longevity and then (if there is a tie in terms of age) on size and finally in terms of intensity. Later experimentation determined that longevity alone was enough and it is that even simpler version that was used in this paper. The good performance of AGE indicates that the key parameters for a tracking algorithm are location error and longevity.

The new tracking technique introduced by Lakshmanan and Smith (2009) exhibits consistently good performance as evidenced by the black and gray bars in Figure 2 on all cases and metrics.

3. Summary

Although storm tracking algorithms are a key ingredient of nowcasting systems, evaluation of storm tracking algorithms has been indirect, labor intensive or non-specific. In this paper, we introduced a set of easily computable bulk statistics that can be used to directly

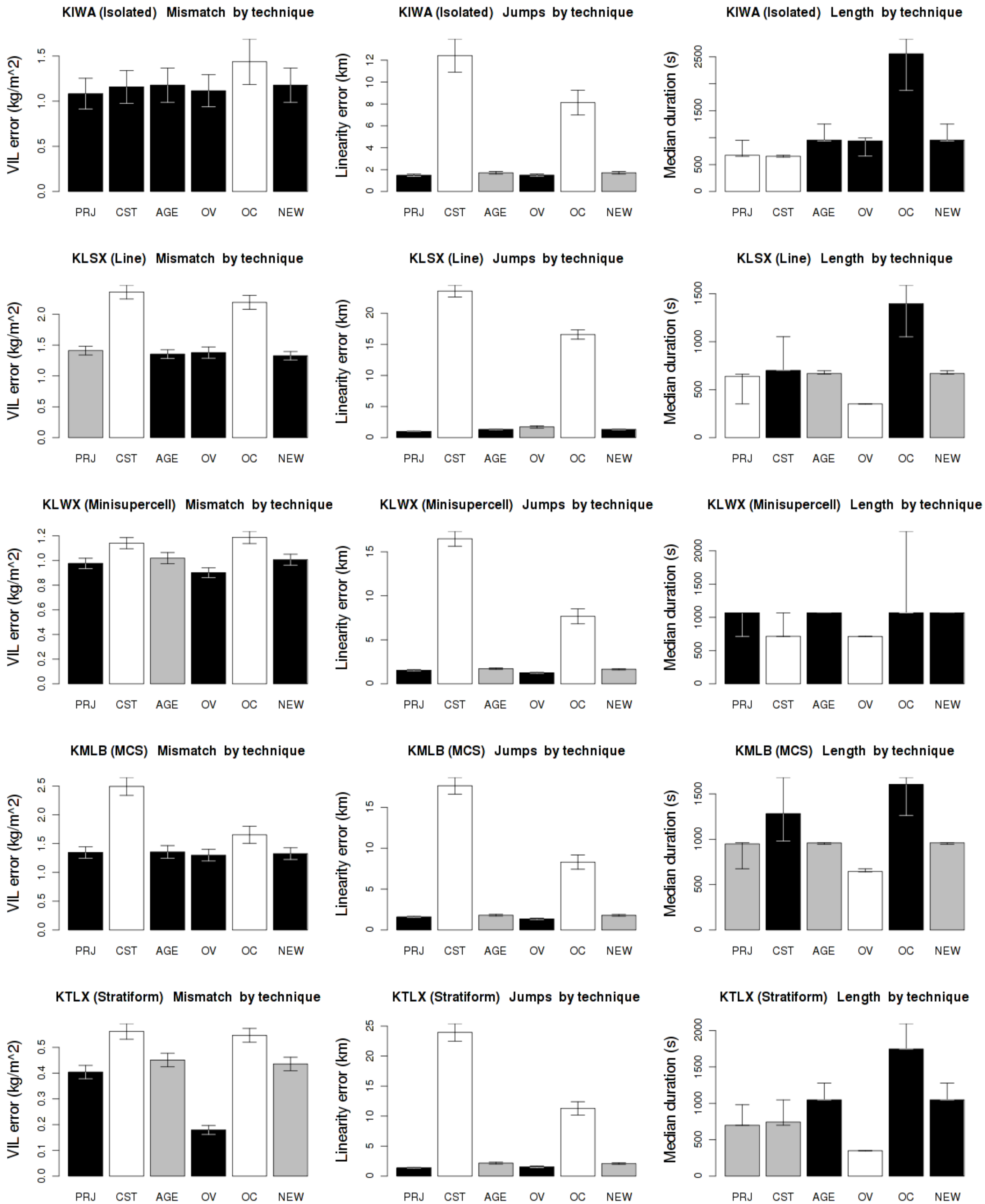

FIG. 2. Evaluation of different tracking techniques. Black bars denote good performance while white bars indicate poor performance.

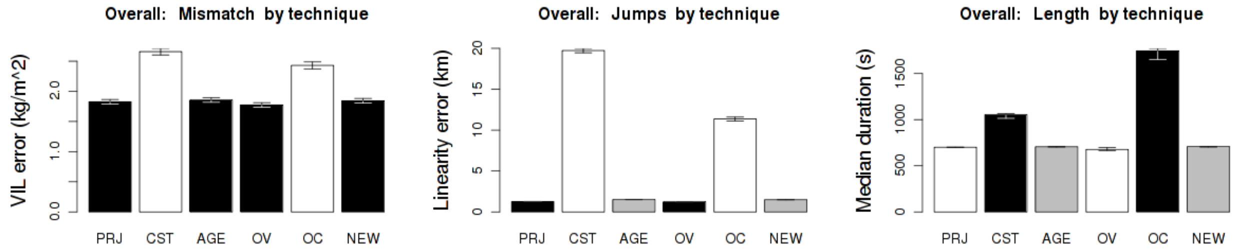

FIG. 3. Evaluation of tracking techniques on all 16 cases. Black bars denote good performance while white bars indicate poor performance.

evaluate the performance of tracking algorithms on specific characteristics. We applied the evaluation method to a diverse set of radar reflectivity data cases and noted the characteristic behavior of five different storm tracking algorithms proposed in the literature and now employed in widely used nowcasting systems. We also devised a storm tracking algorithm that performs consistently and better than any of the previously suggested techniques.

Acknowledgements

Funding for the authors was provided under NOAA-OU Cooperative Agreement NA17RJ1227. We gratefully acknowledge useful discussions with Benjamin Root and Madison Burnett. Brett Roberts carried out some of the initial studies on which this paper is based.

REFERENCES

Dixon, M. and G. Wiener, 1993: TITAN: Thunderstorm identification, tracking, analysis and nowcasting – a radar-based methodology. *J. Atmos. Ocean. Tech.*, **10**, 785–797.

Han, L., S. Fu, L. Zhao, Y. Zheng, H. Wang, and Y. Lin, 2009: 3D Convective storm identification, tracking and forecasting – an enchanced TITAN algorithm. *J. Atmos. Ocean. Tech.*, **26**, 719–732.

Johnson, J., P. MacKeen, A. Witt, E. Mitchell, G. Stumpf, M. Eilts, and K. Thomas, 1998: The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Weather and Forecasting*, **13 (6)**, 263–276.

Lakshmanan, V., K. Hondl, and R. Rabin, 2009: An efficient, general-purpose technique for identifying storm cells in geospatial images. *J. Ocean. Atmos. Tech.*, **26 (3)**, 523–37.

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *J. Atm. Res.*, **67**, 367–380.

Lakshmanan, V. and T. Smith, 2009: An objective method of evaluating and devising storm tracking algorithms. *Wea. and Forecasting*, submitted.

Morel, C., F. Orain, and S. Senesi, 1997: Automated detection and characterization of MCS using the meteosat infrared channel. *Proc. Meteor. Satellite Data Users Conf.*, Eumetsat, Brussels, 213–220.

Wilson, J., N. A. Crook, C. K. Mueller, J. Z. Sun, and M. Dixon, 1998: Nowcasting thunderstorms: A status report. *Bull. Amer. Meteor. Soc.*, **79**, 2079–2099.