Storm Tracking and Forecasting using High Resolution echoes of Short Range X-Band Radar

Sajid Shah[1], Riccardo Notarpietro[1], and Marco Branca[1]

[1]*Politecnico di Torino, Corso Duca degli Abruzzi, Torino, Italy*

(Dated: 17 July 2014)

1. Introduction

Storm Tracking has vital role in the accuracy of any weather predicting tool. In the centroids based tracking algorithms, up to our knowledge only two characteristics of storms i.e. centroids and area have been exploited until now. The contribution of each attribute i.e. centroid and area to the correct tracking, is not yet assessed. Furthermore, very little attention has been paid to tracking evaluation. Finally, a good forecast is the core purpose of any forecasting tool.

For storm identification and tracking a variety of techniques have been adopted. Storm is defined as an object composed of connected pixels in radar observation where the reflectivity of all pixels in the image is greater than a reflectivity threshold and the area of this object is greater than an areal threshold. Storm, cell and object are interchangeable terms in this work. Storm Identification had been already done in the previous work (Shah et al., 2013) where storm(s) are represented by ellipses. Ellipses properties like area, major/minor axis length, orientation and centroids (center of mass) had been calculated. The main focus of this work is on Storm tracking along with some introduction to tracking evaluation and forecasting.

A track is defined as the path followed by a storm during its complete life time or from its initiation to the last time it was observed. Storm Tracking is defined as the time association of storms identified at time instance *t* to the storms already identified and tracked at time instance *t-1*. Storms at time *t* inherit history i.e. life time and tracking id from the associated storms at *t-1*. If the number of storms at *t* is greater than the number of storms at *t-1* then either a split has occurred or a new storm has been initiated. Furthermore, if the number of storms at *t* is less than the number of storms at *t-1* then either a merge has occurred or a storm has completed its life time and is disappeared. Tracking algorithms can be divided into two main categories. Algorithms discussed in (Bjerkaas and Torsyth, 1979; Crane, 1979; Dixon and Wiener, 1993; Handwerker, 2002; Johnson et al., 1998) are centroids based while those discussed in (Lai, 1999; Li and Schmid, 1994; Matthews and Trostel, 2010; Rinehart and Garvey, 1978; Tuttle and Foote, 1998) are based on cross correlation between the reflectivity data obtained from two successive radars scans. Both techniques have benefits and drawbacks. Centroids based algorithms can track individual storms more correctly and can provide more information about individual storms (i.e. area, orientation). However their drawback is that they are designed for convective storms that is why they are not suitable for straitiform events (Pierce et al., 2004). In these algorithms the center of mass of storms (centroids) is used for tracking. Nearest in distance are the matching storms in two successive radar scans. While those algorithms which are based on cross correlation produce more accurate speed and direction information about large areas (Johnson et al., 1998). Procedures developed by (Han et al., 2009; Lakshmanan et al., 2003, 2009) are hybrid of the two mentioned categories.

The centroid based algorithms first identify individual storms and computes their characteristics i.e. centroids, area, major/minor radii and orientation and then associate storms across two consecutive radar scans. Basic idea is that two storms are best candidate for matching (temporal association) if the distances between their centroids is minimum. TITAN (Thunderstorm Identification, Tracking, Analysis and Nowcasting) discussed in (Dixon and Wiener, 1993) combines both center of mass and area of the storm for final decision of tracking. SCIT (The Storm Cell Identification and Tracking) discussed in (Johnson et al., 1998) forecast the centroid locations of cells at time *t-1*. Every cell of time *t* is associated to the closest cell of the projected cells.

Algorithms based on cross correlation computes motion vector by utilizing 2D radars reflectivity data. Perfect examples of this class are TREC (Tracking Radar Echoes by Correlation) (Rinehart and Garvey, 1978) and COTREC (Continuity of TREC) (Li and Schmid, 1994). These algorithms tend to calculate more accurately velocity and direction (Johnson et al., 1998). The drawback of these algorithms is that they are unable to identify individual storms.

Our proposed procedure belongs to the first category i.e. centroid based. In reality it is not centroids based algorithm, rather it is called combinational optimization technique because it combines more than one storm attributes in the final decision of tracking. Most of the centroids based algorithms devised until now deal with a single point i.e. center of mass, therefore, they are prone to error if the identification algorithm falsely identify the center of mass of the storm and thus will result in bad tracking. Examples of bad tracking are given in figure 1. The dotted line represents correct tracking while the solid lines represents actual tracking. These procedures don't use the reflectivity/precipitation distribution of storms for tracking. To the best of our knowledge we have used the reflectivity and shape information of storm in tracking for the first time, following centroid based technique. We present a novel procedure for storms tracking using object descriptors SALdEdA

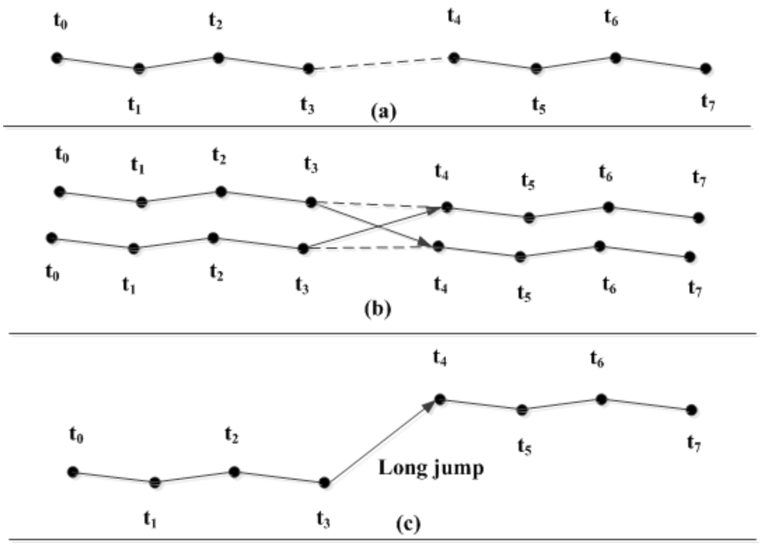


Figure 1: (a) Tracking algorithm fails to associate storms (missed association) at t_3 and t_4; (b) Storms at t_3 and t_4 are incorrectly associated; (c) Storm at t_3 is incorrectly associated with storm at t_4 due to violating maximum speed limitation.

where SALdEdA stands for Structure, Amplitude, Location, Eccentricity difference and Areal difference. Structure (S) is the difference between normalized volumes of reflectivity/precipitation objects. Amplitude (A) is the normalized difference between mean reflectivity/precipitation of objects. Location (L) is the normalized distance between the centers of mass of the objects. Eccentricity (E) tells about the circularity of objects which is a measure of how much an object is circular or elongated. Absolute difference between the eccentricities of objects is used to find how closely they are circular. Area is the number of pixels in the object. Basic idea behind mapping (associating) two objects of two consecutive time instances i.e. *t* and *t-1* is to find how much similar they are in structure, amplitude, circularity and area and how much closely they are located.

All descriptors i.e. SALdEdA are ranging from 0 to 1 where perfect match is attributed when every descriptor has 0 value and completely imperfect match occurs when every one of them is 1. SALdEdA is calculated for all possible combination of storms at time t and t-1. A global cost function is formulated which is the weighted sum of all SALdEdA descriptors. At this stage the problem of associating two storms becomes a combinatorial optimization problem. The aim is to have such matching where the global cost function is minimized. Hungarian is the candidate algorithm for solving our problem because it is relatively easy to implement. Next job after storm tracking is to objectively evaluate the accuracy of tracking algorithm. The adopted method is called percent correct which is the ratio of number of correct tracks divided by the number of total tracks. This approach is labour intensive.

Forecast is the prediction of future event(s) that yet to occur. A forecast can be short, medium or long term. Meteorological data or radar observations are inherently time-oriented. Forecast is always erroneous; good forecast is the one with smaller error. Larger the forecast leading time also called forecast horizon, larger forecasts error is expected and vice versa. Our variables of interest for the forecasting are area of the storm, major/minor axis length, angle of orientation, speed and direction. We have adopted First Order followed by Second Order Exponential smoothing strategies in order to model the above time series. Forecast of these variables is based on the assumption that growth and decay (increase and decrease) of these variables is linear.

2. Our Tracker

As mentioned before centroids based approaches use center of mass for tracking while some of them use area as well i.e. TITAN, ETITAN (Enhanced TITAN). These approaches can easily track the situation in figure 2.a but they will fail to correctly track storms of figures 2.c and 2.e. The figure 2.c shows that the candidate storms at *t* have same area and same distance from storm at *t-1*, but different reflectivity distribution. Also figure 2.e shows that both the candidate storms at time *t* have same area, same distance from center of storm *t-1* and same reflectivity distribution but their shape is different i.e. one is more circular while the other one is elongated. Having similar assumptions as of TITAN for tracking we additionally assume that the reflectivity distribution and the shape of a storm in two successive radar scans don't change abruptly until and unless a merge or a split occurs. SAL (Structure, Amplitude and Location) had been proposed by (Wernli et al., 2008) for the purpose of verification of quantitative precipitation forecasts. The basic idea of SAL was to present an object-based quality measure for the verification of forecasts. We have modified SAL and added also some others characteristics of object i.e. eccentricity and area, to find the similarity between objects (storms).

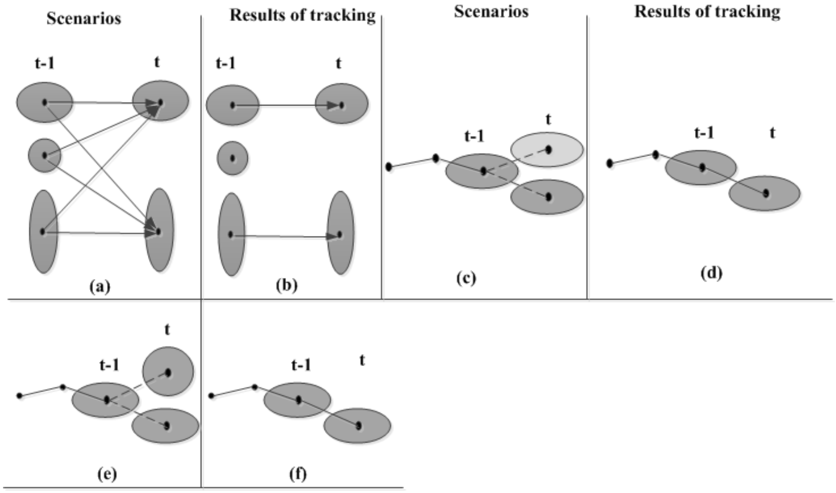


Figure 2: (a) The number of storms at t is less than number of storms at t-1; (b) Storms of (a) have been associated based on distance between centroids and areas of storms; (c) Reflectivity of the two candidate storms is different; (d) Storms of (c) have been associated considering reflectivity distribution of storms; TITAN missing this aspect; (e) Circularity of the two storms is different; (f) Storms association based on elliptical circularity;

3. Definition of Components SALdEdA

Consider a domain D_i which represents i^{th} storm, comprises of N number of pixels, where i = 1, 2, 3 . . . k, and *k*ccorresponds to the number of storms per radar scan (image/map). Current time is represented by *t* while *t-1* denotes previous time instance. Radar reflectivity of i^{th} storm in dbz is represented by Z_i. Maximum reflectivity of a storm is represented by Z^{max} . If SALdEdA is used for precipitation objects, Z should be replaced by R.

Key idea is that the change in Structure, Amplitude, Location, Eccentricity and Area of a storm is not rapid in two consecutive radar observations. Therefore, our objective is to have a matching with minimum cost and the aim is to have minimum difference between the matching storms. The objective function is define as:

$$Q = \sum C_{i,j}; C_{i,j} = w_1 S + w_2 A + w_3 L + w_4 dE + w_5 dA \tag{3.1}$$

where w_i represents weight and is in the range of [0, 1] and SALdEdA are defined below. The combinatorial optimization problem is solved by using Hungarian algorithm.

Key concept is to compare the normalized value of each SALdEdA variable. Structure (S) is the difference between normalized volumes of reflectivity/precipitation objects. Amplitude (A) is the normalized, absolute difference of mean reflectivity of two storms subjected to tracking. Location (L) is the normalized distance between the centers of mass of two storms. The dE component stands for the difference of eccentricities of two storms subjected to comparison for matching. Eccentricity describes the circularity of an object (storm).

$$S = \left| \frac{V_{i,t} - V_{j,t-1}}{V_{i,t} + V_{j,t-1}} \right| \tag{3.2}$$

$$A = \left| \frac{M(Z_{i,t}) - M(Z_{j,t-1})}{M(Z_{i,t}) + M(Z_{j,t-1})} \right| \tag{3.3}$$

$$L = \frac{|x(Z_{i,t}) - x(Z_{j,t-1})|}{d} \tag{3.4}$$

$$dE = |Eccentricity_{i,t} - Eccentricity_{j,t-1}| \tag{3.5}$$

$$dA = \left| \frac{Area_{i,t} - Area_{j,t-1}}{Area_{i,t} + Area_{j,t-1}} \right|; \quad \begin{matrix} i = 1, 2, \ldots k_1 \\ j = 1, 2, \ldots k_2 \end{matrix} \tag{3.6}$$

where V and $M(Z)$ are defined below and $x(Z)$ corresponds to the center of mass and $Area_{i,t}$ corresponds to the area of the storm. Here k_1 and k_2 are the number of storms at *t* and *t-1*.

$$V = \frac{\sum_{(x,y)\in D_i} Z_{x,y}}{Z_i^{max}} = \frac{Z_i}{Z_i^{max}} \quad (3.7)$$

$$Z = \sum_{(x,y)\in D_i} Z_{x,y}; i = 1, 2, \dots k \quad (3.8)$$

$$M(Z) = \frac{1}{N} \sum_{(x,y)\in D_i} Z_{x,y} = \frac{Z_i}{N} \quad (3.9)$$

$$Eccentricty = 1 - \frac{r_{minor}}{r_{major}} \quad (3.10)$$

here $Z_{x,y}$ is the pixel reflectivity value in dbz N is the number of pixels in a storm, r_{minor}, r_{major} are the minor and major radii of storm and $0 \leq \{S, A, L, dE, dA\} \leq 1$. The value of a variable from SALdEdA equal to zero means complete agreement while its value equal to 1 means complete disagreement.

The scope of feasible solution set is restricted by putting some limitations. For example TITAN puts restriction over the speed $L/\triangle t$ of a storm that it should be less than certain maximum speed, otherwise the association is not considered feasible for tracking. However it is not a robust method because it fails for different sizes of storms. We have adopted a more robust and flexible procedure by putting limitation over the distance between centroids of two candidate storms. Only those matching are considered feasible if the distance between the two candidate storms is less than certain MAXDIST threshold. The MAXDIST threshold is set a list of values depends upon the area of the storm shown in table 1. The assumption is that the centroids of larger storm is expected to move more than a smaller one. It is the drawback of centroid based techniques that the center of mass of a storm moves faster than actual storm motion, because in reality smaller storm moves faster than the larger one.

Table 1: MAXDIST in km verses area in km^2

	area < 3.6	3.6 ≤ area < 21.6	21.6 ≤ area < 36	36 ≤ area < 216	216 ≤ area < 316	area > 316
MAXDIST	4.2	4.8	6	6.6	7.2	9

3.1. Handling Splits and Mergers

Storm merging occurs more frequently while split occurs less frequently (Dixon and Wiener, 1993). When a merge occurs, the best matching track is extended while others are terminated. In case of split, the best match is extended while the rest are treated as newly initiated storms.

4. Storm Forecasting

Forecasting is the modelling and extrapolation of patterns found in the time series. By default radar observations are time series. Let suppose p_t is the current value of a parameter or variable of interest and $\frac{dp}{dt}$ is the rate of change then forecasting is defined as:

$$p_{t+\delta t} = p_t + \left(\frac{dp}{dt}\right)\delta t \quad (4.1)$$

where δt is the forecast leading time. Our variables of interest for the forecasting are area, major/minor axis length, angle of orientation and speed and direction of the storm. We have adopted First Order followed by Second Order Exponential smoothing strategies in order to model our variables of interest. Forecast of these variables is based on the assumption that these variables growth and decay (increase and decrease) linearly. Exponential Smoothing discussed in (Montgomery et al., 2008), is a technique which assigns geometrically decreasing weights to previous observations.

5. Results

The results have been produced to testify the efficiency and goodness of the system using datasets collected from Foggia radar site see table 2 for further details about the description of datasets. The first three datasets are convective while the last two are stratiform. For storm identification all parameters are set according to (Shah et al., 2013). For storm tracking and forecasting the colour scheme of ellipse's boundary is: { white:history, black:current and red:forecast}. Areal threshold T_a is set to 10 km^2 for all experiments. The initial threshold T_z is set to 28 dbz for convective and 20 dbz for stratiform events. Our main focus is to find the contribution of each variable of SALdEdA in storm tracking and to evaluate the goodness of over all system after combining SALdEdA variables for tracking.

Table 2: Data Sets

Events	Start Time	End Time	Duration(minutes)
April 1^{st}, 2013	16:31:04	20:00:04	210
June 3^{rd}, 2013	10:00:06	13:00:06	181
August 20^{th}, 2013	13:31:05	18:20:05	290
November 22^{nd}, 2013	04:21:06	09:20:06	300
December 27^{nd}, 2013	00:01:06	08:00:03	480

The contribution of each variable of SALdEdA for storm tracking is assessed. A variable contributes more in storm tracking if it is capable of separating storms into distinct classes or clusters. The clusters or classes are said to be distinct if there is no overlapping between two clusters or classes. The plotted results in figure 3 show that structure (S) and location contributes more. The area and structure of storm contributes in the same way. The amplitude contributes more than eccentricity. Every variable of SALdEdA is assigned a weight according to its contribution while calculating cost of every match in eqn. 3.1.

The over all performances of our tracker have been compared with manually tracked (labelled) storms. The obtained results, given in tables 3 and 4, show the significance of our procedure with 84.90% accuracy for convective events and 96.55% accuracy for stratiform events. The obtained results for stratiform events is good because of the static nature of these storms. After manual verification we come to know that mostly the error in tracking is caused by inappropriate value of MAXDIST. A most robust procedure will be investigate in the future work.

Table 3: Percentage Correct for convective events

Events	Total Tracks	Correct Tracks	Percentage Correct
April 1^{st}, 2013	43	34	79.00
June 3^{rd}, 2013	39	28	71.17
August 20^{th},2013	84	79	94.00
Over All	**166**	**141**	**84.90**

Table 4: Percentage Correct for stratiform events

Events	Total Tracks	Correct Tracks	Percentage Correct
November 22^{nd}, 2013	19	19	100.00
December 27^{nd}, 2013	68	65	95.55
Over All	**87**	**84**	**96.55**

Forecast leading time (FLT) has been set to 5,10 and 15 minutes. Minimum history required for forecasting is set to 3 minutes. A five minutes forecast with corresponding current storm and 6 minutes history is shown in figure 4. A contingency table is calculated for comparing the actual mean reflectivity (AMR) with the corresponding forecasted mean reflectivity (FMR). A "hit' is considered if both AMR and FMR existed and the relative difference between the AMR and FMR is within the range $\pm$0.05 dBz while for less than -0.05 dBz "underestimate" is recorded and for greater than 0.05 dBz over "overestimate' is counted. "False alarm" occurred if either of AMR or FMR is missing.

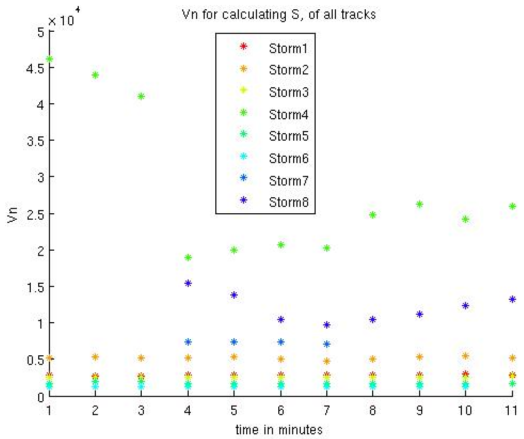
The obtained results are presented in tables 5 and 6. Very promising hit rate is achieved. It could be noted that increasing the forecasting time the hit rate decreases while "overestimate" and "underestimate" increases. As expected good results are obtained for stratiform events than convective events.

Table 5: Contingency table for convective events

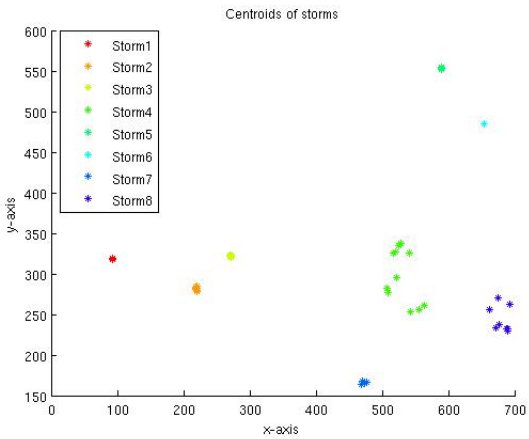
Forecast Leading Time	Hits(%)	Over Estimate(%)	Under Estimate(%)	False Alarm(%)
5	90.28	6.86	2.34	0.50
10	77.05	17.08	4.52	1.34
15	62.64	20.60	11.72	5.02

Table 6: Contingency table for stratiform events

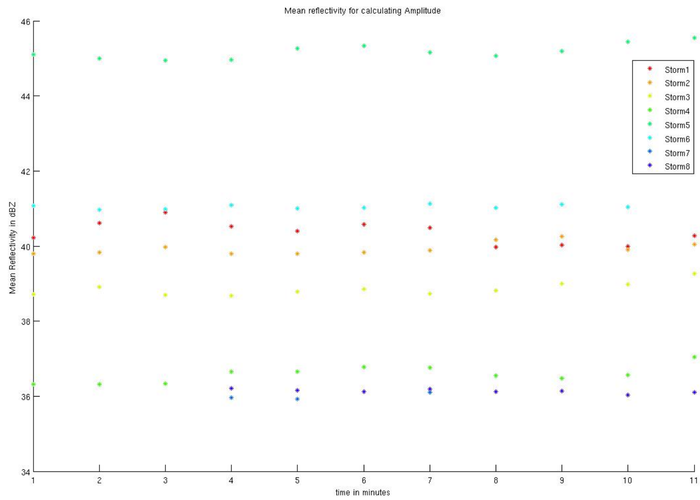
Forecast Leading Time	Hits(%)	Over Estimate(%)	Under Estimate(%)	False Alarm(%)
5	95.86	1.65	2.47	0
10	89.25	3.80	4.46	2.47
15	79.01	6.28	4.79	5.78



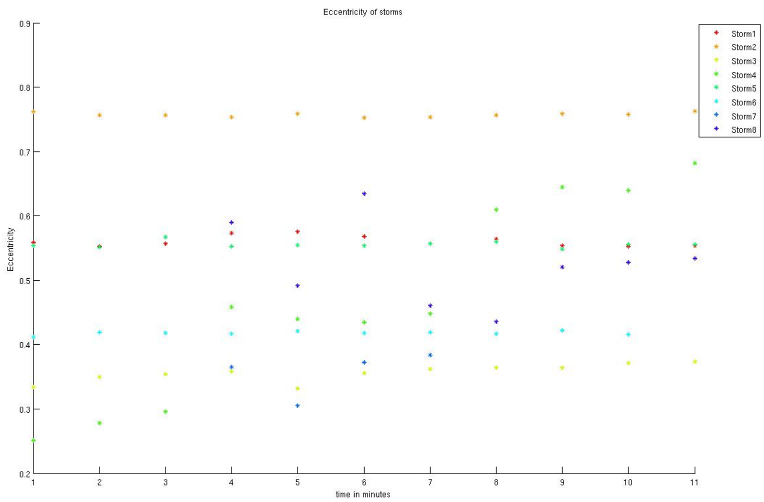
(a) V_n for calculating Structure



(b) Centroids of storms

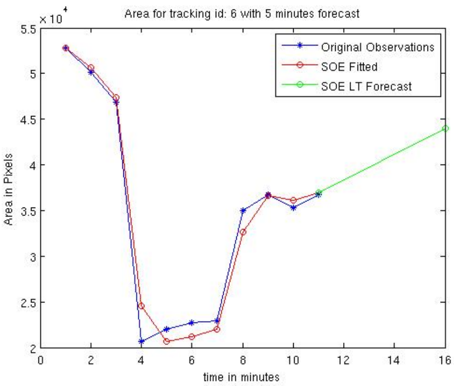


(c) Mean reflectivity for calculating Amplitude

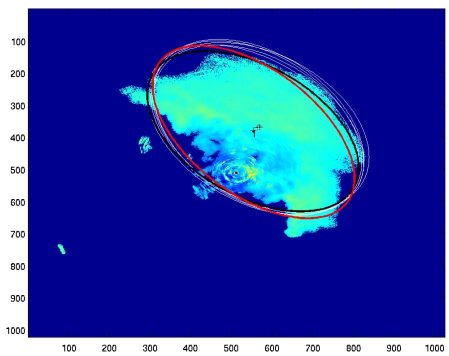


(d) Eccentricity of storms

Figure 3: Contribution of SALdEdA variables in storm tracking



(a) Smoothed values along with actual values of area in pixels of a storm and 5 minutes forecast



(b) Forecast and actual storms

Figure 4: Forecast and actual

6. Conclusion

The usage of more than one storm attributes for storm tracking is investigated in this work. The newly introduced SALdEdA components produced quite good results. Second Order Exponential smoother also produced good forecasts.

In future we aim to work on improving the procedure for putting restriction over maximum speed or maximum distance. Also, new procedure for forecasting verification is under consideration. Apart from this we aim to test our system on large datasets.

References

L. Bjerkaas, C. and E. Torsyth, D., “Real-time automotive tracking of severe thunderstorms using doppler weather radar,” ser. 11 conference, 1979.

K. Crane, R., “Automatic cell detection and tracking,” *IEEE Transaction, Geoscience Electronics*, vol. GE–17, pp. 263–276, 1979.

M. Dixon and G. Wiener, “Titan: Thunderstorm identification, tracking, analysis, and nowcasting- aradar-based methodology,” *Atmos., Oceanic Technol.,*, vol. 10, pp. 785–797, 1993.

L. Han, S. Fu, L. Zhao, Y. Zeng, H. Wang, and Y. Lin, “3d convective strom identification, tracking, and forecasting - an enhanced titan algorithm,” *Atmos. Oceanic Technol.*, vol. 26, pp. 719–732, April 2009.

J. Handwerker, “Cell tracking with trace3d- a new algorithm,” *atmos. Res*, vol. 61, pp. 15–43, 2002.

T. Johnson, J., L. Mackeen, P., A. Witt, D. Mitchell, E., J. Stumpf, G., D. Eilts, M., and W. Thomas, K., “The storm cell identification and tracking algorithm: An enhanced wsr-88d algorithm,” *Wea, Gorcasting*, vol. 13, pp. 263–276, June 1998.

E. S. T. Lai, “Trec application in tropical cyclone observation,” ESCAP/WMO Typhoon Committee Annual Review, Tech. Rep., 1999.

V. Lakshmanan, R. Rabin, and V. Debrunner, “Multiscale storm identification and forecast,” *Atmos. Research*, pp. 367–380, 2003.

V. Lakshmanan, K. Hondl, and R. Rabin, “An efficient, general-purpose technique for identifying storm cells in geospatial images,” *Atmos. Oceanic Technol.*, vol. 26, pp. 523–537, March 2009.

L. Li and W. Schmid, “Nowcasting of motion and growth of precipitation with radar over a complex orography,” *Applied Meteorology*, vol. 34, pp. 1286–1300, November 1994.

J. Matthews and J. Trostel, “An improved storm cell identification and tracking (scit) algorithm based on dbscan clustering and jpda tracking methods,” in *International Lightning Detection*, 2010.

D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. Wiley, 2008.

C. E. Pierce, E. Ebert, A. W. Seed, M. Sleigh, C. G. Collier, N. I. Fox, N. Donaldson, J. W. Wilson, R. Roberts, and C. K. Mueller, “The nowcasting of precipitation during sydney 2000: An appraisal of the qpf algorithms,” *AMS(American Meteorological Society)*, vol. 19, pp. 7–21, February 2004.

R. E. Rinehart and E. T. Garvey, “Three-dimensional storm motion detection by conventional weather radar,” *Nature*, 1978.

S. Shah, R. Notarpietro, S. Bertoldo, M. Branca, C. Lucianaz, O. Rorato, and M. Allegretti, “Automatic storm(s) identification in high resolution, short range, x-band radar images.” ICEAA, 2013.

J. D. Tuttle and G. R. Foote, “Determination of the boundary layer airflow from a single doppler radar,” *Atmos. Oceanic Technol.*, vol. 7, pp. 2079–2099, 1998.

H. Wernli, M. Paulat, M. Hagen, and C. Frei, “Sal- a novel quality measure for the verification of quantitative precipitation forecasts,” *AMS(American Meteorological Society)*, vol. 136, pp. 4470–4487, 2008.