A stochastic method for convective storm identification,
tracking and nowcasting

Lei Han\textsuperscript{a,b}, Shengxue Fu\textsuperscript{a}, Guang Yang\textsuperscript{c}, Hongqing Wang\textsuperscript{b,\*}, Yongguang Zheng\textsuperscript{d},
Yingjing Lin\textsuperscript{b}

\textsuperscript{a} College of Information Science and Engineering, Ocean University of China, Qingdao 266101, China
\textsuperscript{b} Department of Atmospheric Science, Peking University, Beijing 100871, China
\textsuperscript{c} Department of Probability and Statistics, Peking University, Beijing 100871, China
\textsuperscript{d} National Meteorological Center, Beijing 100081, China

Received 28 February 2008; received in revised form 28 May 2008; accepted 17 June 2008

Abstract

The convective storm identification, tracking and nowcasting method is one of the important nowcasting methodologies against severe convective weather. In severe convective cases, such as storm shape or rapid velocity changes, existing methods are apt to provide unsatisfied storm identification, tracking and nowcasting results. To overcome these difficulties, this paper proposes a novel approach to identify, track and short-term forecast (nowcast) of convective storms. A mathematical morphology-based storm identification method is adopted which can identify storm cells accurately in a cluster of storms. As for the difficult tracking problem, sequential Monte Carlo (SMC) method is utilized to simplify the tracking process. It is not only inherently suitable for handling complicated splits and mergers, but also capable of handling the case of storm-missing detection. In order to provide more accurate forecast of a storm position, this method takes the advantages of the cross-correlation method. The qualitative and quantitative evaluations show the efficiency and robustness of the proposed approach.

© 2008 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.

Keywords: Convective storm; Doppler radar; Nowcasting

1. Introduction

Strong convective weather events are potentially damaging and dangerous to lives and property. But these convective weather systems are difficult to be observed using the ordinary observing instruments because of their short lifetime and small spatial scale. Doppler weather radar can provide a high temporal and spatial resolution of continuous measurement of convective weather events. The new generation weather radar network is under construction in China which will greatly benefit the severe weather warning operations. The convective storm identification, tracking and forecasting method is one of the important nowcasting techniques, and can also provide continuous storm data suitable for the study of the physical mechanism of storm evolution [1].

Many nowcasting algorithms based on radar data have been developed over the past decades. These algorithms can be classified into two categories: the cross-correlation method [2] and the centroid-type method [3-6]. The cross-correlation method uses the reflectivity data to calculate the motion vector field which can be used to forecast the storm movement. The advantage of the cross-correlation method is that it provides more accurate speed and direction information of the reflectivity echo [5]. Its weakness is that it is unable to identify and track individual storms. Such a typical algorithm is tracking radar echoes by correlation (TREC) [2]. The centroid-type method is able to track individual isolated storms effectively and to provide the time his-
tory data of each storm, such as vertically integrated liquid water (VIL), volume, and top height. Two well-known centroid-type algorithms are TITAN (thunderstorm identification, tracking, analysis, and nowcasting) [4] and SCIT (storm cell identification and tracking) [5]. TITAN uses a single threshold to identify storms. In the tracking phase, TITAN utilizes a combinatorial optimization method to match the two storm sets across successive scans. Additional effort is needed to treat splits and mergers. SCIT uses multiple thresholds to identify storms and the nearest neighbor method to track storms. SCIT is unable to handle splits and mergers. For forecasting, both methods use the centroid displacement to forecast the storm motion.

When there are dense severe convective storms whose shape or velocity changes rapidly, or in the case that clusters of storms occur frequently, existing methods have the following difficulties: (i) Storm identification. If the echoes are distributed densely, there will be a false merger and clusters of storms. False merger means that two individual storms are treated by the identification algorithm as a single (merged) storm. TITAN is apt to identify a cluster of storms as one storm, and it cannot handle the false merger problem. SCIT loses the internal structure information of storms when it isolates adjacent storm cells from the cluster using several reflectivity thresholds. (ii) Storm tracking. SCIT is apt to fail when the storm shape or velocity changes rapidly. Besides, existing methods cannot handle the storm-missing detection problem. Missing detection means an existing storm is not detected by the identification algorithm and thus cannot be tracked correctly. (iii) Nowcasting. All existing centroid-type methods use the storm centroid displacement to forecast the storm motion, which may result in large errors from the actual storm movement if the shape or size of the storm changes rapidly.

To surmount these difficulties, this paper introduces a novel centroid-type approach to identify, track and nowcast convective storms. A mathematical morphology-based storm identification method is adopted which can identify false merger and isolate adjacent storm cells from a cluster of storms. Sequential Monte Carlo (SMC) method, rooted in statistics, is utilized to simplify the tracking process. It is not only inherently suitable for handling complicated splits and mergers, but also capable of handling the case of storm-missing detection. In order to provide more accurate forecast of storm motion, this paper incorporates the cross-correlation method into the proposed method through utilizing the motion vector field calculated by the cross-correlation method.

2. Storm identification

The identification algorithm only uses the reflectivity values to identify storm. And the radar coordinate data have been transformed into Cartesian coordinates.

Under the cases where storms are distributed sparsely, existing methods mentioned above perform well. But it is difficult for these methods to identify false merger and isolate adjacent storms from a cluster of storms. We adopt a mathematical morphology-based storm identification method [7]. This method first applies the single threshold identification followed by implementing an erosion process to resolve false merger problem. During multi-threshold identification stages, dilation operation is performed against the storm cells which are just obtained by the higher threshold identification, until the storm edges touch each other or touch the edges of the previous storms identified by lower threshold. The main steps are listed in Fig. 1. Detailed information can be found in Ref. [7].

3. Tracking

Recently, particle filters, also known as sequential Monte Carlo (SMC), have proved to be powerful tools for tracking problems [9-12]. These methods are rooted in ideas from statistics, control theory and computer vision. The advantage of these methods lies in their simplicity, flexibility, and systematic treatment of nonlinearity and non-Gaussianity. The basic principle is to use an iterative sampling process to dynamically estimate the interested target state. Every iteration step includes three parts: sampling, prediction and measurement. Because the theoretical explanation of SMC is very complicated, this paper only describes how to implement SMC in storm tracking domain.

We simplify SMC when performing storm tracking across consecutive scans. Only random sampling and one time-step prediction remained. As for the case of storm-missing detection, SMC is fully used.

3.1. Tracking storms across consecutive scans

Because the composite reflectivity (CR) image can represent the storm properties well, detected storms are then tracked in CR images. Here, the composite reflectivity image can be treated as a 2D grid, in which each grid value

- Apply the single-threshold identification algorithm with the first threshold. $T_{min}$
- Erode the 3D storms which are just identified above.
- For $i = 2, N$, do:
  - Identify sub-storms with the $i$th threshold. $T_{min} + 5 \times (i - 1)$
  - Dilate the sub-storms
  - Erode the sub-storms

Fig. 1. Algorithm for automatic storm identification based on mathematical morphology. The first step is the same as TITAN. The second step is to disconnect the weak connection between the two adjacent storms. And the third step is gradually increasing the threshold $T_{i} = T_{min} + 5 \times (i - 1)$, $i = 2, \ldots, N$, to identify storms followed by implementing both the dilation and erosion operations, so we can gradually isolate storm cells with different intensity from a cluster of storms. Here, $N$ is the number of thresholds set by users, typical 2-6.
is calculated as the maximum radar reflectivity observed by all elevation scans above it.

### 3.1.1. Sampling

For each detected storm at time-step \( t - 1 \), draw particles (also called samples) which are uniformly distributed within the storm areas (as shown in Fig. 2). The number of the particles is given by

\[
N_{P_k} = \text{floor}(A_k), \quad k = 1, \ldots, N_{t-1},
\]

where \( A_k \) is the area of the \( k \)th storm (normally between 30 and 1000 km\(^2\)). \( N_{t-1} \) is the number of storms detected at \( t - 1 \). \( \text{floor}(A_k) \) indicates the maximum integer no greater than \( A_k \).

The state of the particles is denoted by \( x_{t-1} \), where \( x = (x, y) \), namely, the position of the particle. Besides, the storm ID and an equal weight are also assigned to each particle. The storm ID can help us to distinguish particles belonging to different storms during prediction and update stages. For the \( k \)th-detected storm at \( t - 1 \), we denote the particle set which is sampled from this storm as \( \{ x_{j,t-1}, \omega_{j,t-1} = 1/N_{P_k}, k \}_{j=1}^{N_{P_k}} \).

### 3.1.2. Prediction

Each particle should be passed through the following one-order linear system model to obtain its predicted position:

\[
z_t = x_{t-1} + V_t \times \Delta t + \omega(t),
\]

where \( V_t \) is the velocity at pixel \((x, y)\) which is calculated using the cross-correlation method [2], \( \Delta t \) is the radar sampling interval, typically 5–10 min, and \( \omega(t) \) is the Gaussian noise with zero mean and variance \( \sigma^2 \).

### 3.1.3. Measurement

Denote \( \theta_{t,k}^m \) as the event: the \( m \)th storm at time-step \( t \) evolves from the \( k \)th storm at time-step \( t - 1 \). Once the observation \( Y_t \) arrives, the probability of \( \theta_{t,k}^m \) is calculated through a special ratio

\[
p(\theta_{t,k}^m|Y_t) = \max \left( \frac{N_{F_{m,k}}}{A_{m}}, \frac{N_{F_{m,k}}}{A_k} \right), \quad m = 1, \ldots, N_t, \quad k = 1, \ldots, N_{t-1},
\]

where \( A_m \) is the area of the \( m \)th detected storm at time-step \( t; A_k \) the area of the \( k \)th detected storm at time-step \( t - 1 \); \( N_{F_{m,k}} \) the number of predicted particles which fall into the \( m \)th storm at time-step \( t \), and these \( N_{F_{m,k}} \) particles all belonged to the \( k \)th storm at time-step \( t - 1 \).

If \( p(\theta_{t,k}^m|Y_t) \) is larger than a threshold \( T_r \) (0.5 in this paper), the \( m \)th storm at time-step \( t \) is considered to be evolved from the \( k \)th storm at time-step \( t - 1 \). This constitutes the temporal–spatial association inference process.

For instance (as shown in Fig. 2), storms 22 and 23 at time-step \( t \) correspond to storms 12 and 13 at time-step \( t - 1 \), respectively. The corresponding trajectory paths are extended. Storm 24 is considered to be a newly formed storm because there are no predicted particles falling into it. Then, a new trajectory path is established.

On the contrary, storm 11 possibly disappeared at time-step \( t \) because all predicted particles from storm 11 do not fall into any detected storm at \( t \). Because the missing detection may occur, a deferred logic is utilized [13]. Here, deferred logic means that instead of making a decision at the current time, the tracking decision is deferred until several scans (not greater than \( N \) scans) have been examined. The reasons of missing detection may be that a few radial lines are not sampled correctly during the radar scanning process, and the radar cannot “see” the storm when the storm is just between the two elevation angles, or the storm reflectivity does not meet the threshold only for a short time.

Now, we use Fig. 3 to explain the implementation of deferred logic using the standard SMC. As shown in Fig. 3, storm 11 is detected at \( t - 2 \) but missed at \( t - 1 \), and reoccurs at \( t \) with a new label: storm 31. If we do not correctly handle this missing detection case, storm 31 at \( t \) will be treated as a newly born storm, which is actually an existing old storm.

Firstly, at \( t - 2 \), draw particles within storm 11 and predict these particles using Eq. (2). Then, at \( t - 1 \), for each particle predicted from storm 11, its weight is adjusted by

\[
w_j,t-1 = \frac{I_j,t-1}{T_{m_{\min}}} \quad j = 1, \ldots, N_{P_k},
\]

where \( I_j,t-1 \) is the measured reflectivity at the position of particle \( j \); \( N_{P_k} \) the number of particles drawn from storm 11 at \( t - 2 \); and \( T_{m_{\min}} \) the reflectivity threshold in the detection algorithm. Then normalize weight so that \( \sum_{j=1}^{N_{P_k}} w_j,t-1 = 1 \).
The centroid position at \( t - 1 \) is not available because of the missing detection, so we need to estimate it. The virtual centroid of the storm is estimated by

\[
\hat{x}_{t-1} = \sum_{j=1}^{NP} w_{j,t-1} x_{j,t-1}.
\]

(5)

Now, we simulate \( p(x_t | Y_{t-1}) \) through

(1) Resampling the particle set \( \{x_{j,t-1}, w_{j,t-1}\}_{j=1}^{NP} \) to obtain the new particle set \( \{x_{j,t-1}, 1/NP\}_{j=1}^{NP} \). We use these new particles to represent the "virtual storm" at \( t - 1 \). The purpose of resampling is to prevent the deterioration of particles [8].

(2) Predicting these particles by Eq. (2).

At time-step \( t \), if \( p(\theta^A_t | Y_t) \) obtained by Eq. (3) shows that storm 31 at time-step \( t \) is associated with the "virtual storm" at \( t - 1 \), thus associated with storm 11 at \( t - 2 \). Then the storm trajectory can be rebuilt through the virtual centroid. Otherwise, storm 11 is considered to have disappeared at \( t - 1 \), and the related particles are discarded (here we assume that particles are only reserved for one time-step, \( NR = 1 \)).

3.2. Handling splits and mergers

Splits and mergers are handled simultaneously when calculating \( p(\theta^A_t | Y_t) \). No additional effort is needed. For the splitting case as shown in Fig. 4(a), the sampling and prediction steps are the same as described in Section 3.1. Then in the measurement step, \( p(\theta^A_t | Y_t) \) obtained by Eq. (3) shows that both detected storms 21 and 22 at \( t \) are associated with storm 11 at \( t - 1 \). This is considered to be a split case where \( p(\theta^A_t | Y_t) = \frac{NP}{4} \). Following the same process, and through \( p(\theta^A_t | Y_t) \), we can also easily identify mergers, while the direction is opposite. Fig. 4(b) shows a schematic case of merger.

Compared to the existing methods, such as TITAN and SCIT, the proposed method is inherently suitable for handling splits and mergers and is robust even in very complicated cases (a challenging example is presented in experiment section).

4. Nowcasting

This section will introduce how to incorporate the cross-correlation method into our centroid-type method to obtain storm position forecast. The computation of the cross-correlation is also based on the CR image.

After identifying and tracking storms, the nowcasting can be made for a number of storm parameters based on the linear fit extrapolation [4,5] (as shown in Fig. 5). The parameters include storm top, maximum reflectivity, volume, centroid, storm motion (speed and direction), and projected area. The most important parameter is the storm motion because it decides the impacted area in the forecast period.

All the existing centroid-type methods use the storm centroid displacement to forecast the storm motion, which may result in the following problem. As shown in Fig. 6, the shape and size of the storm can change rapidly, causing random displacement of the storm cell centroid. This can cause large deviation from the correct storm velocity leading to forecasting errors.

The advantage of the cross-correlation method is to provide more accurate speed and direction information of the reflectivity echo [5]. Its weakness is that it is unable to identify and track individual storms. The following algorithm combines the merits of correlation tracking and centroid tracking in the context:

1. Apply TREC to generate the motion vector field. For every detected storm, we retrieve the motion vector \( V \) at the point where the storm centroid locates.

2. Using \( V \) as the storm velocity to forecast the storm position.

Fig. 4. Handling the split (a) and merger (b).

Fig. 5. Forecast based on time history data (from [4]).
5. Experiment and analysis

5.1. Qualitative analysis

In order to evaluate the performance of the proposed method, and to give an impression of the results provided by the proposed method, this section presents three real cases for storm identification, tracking and forecasting separately. The following images are all composite reflectivity images.

As for storm identification, Fig. 7 shows a typical case of a cluster of storms at 1729 LST 31 May 2005 (from Tanggu radar station). For comparison, Fig. 7(a) and (b) shows the results of TITAN and the proposed method, respectively. The brown curve indicates the boundary of identified storms. Because of using a single threshold (40 dBZ), TITAN identifies this cluster of storms as one big storm (in white rectangular), which is obviously unreasonable.

Fig. 7(b) shows the result of the proposed method. The adjacent storms in the cluster of storms are isolated well.

Fig. 8 shows a challenging case of split. A storm split into three sub-storms at 1923 LST 31 May 2005. These sub-storms have irregular shapes and different movement trends. The proposed method successfully associates all three sub-storms at 1923 with their parent storm at 1917 LST 31 May 2005. The white curve indicates the storm centroid history trajectory. Note that these three sub-storms inherit their parent storm’s trajectory path, which is absolutely necessary to forecast the position of these sub-storms.

Fig. 9 shows another example of a supercell observed at 1627 and 1645 LST 24 June, 2006. Fig. 9(a) shows the forecasting result using the centroid displacement at 1627 LST. The storm is predicted to move southward. The cyan curve indicates a 18-min forecast of storm position and size. Fig. 9(c) shows the forecasting result using the proposed method at 1627 LST. The storm is predicted to move eastward. Fig. 9(b) and (d) shows the actual position of the storm. The 18-min forecast shown in Fig. 9(a) is greatly deviated from the actual moving direction of the echo; meanwhile the forecast shown in Fig. 9(c) correctly reflects the moving trend of the echo. This verifies that the short-term forecast can be improved by incorporating the cross-correlation method into centroid-type method, which is important for nowcasting operations.

5.2. Quantitative evaluation of the short-term forecasting

The contingency table approach [4] has been used to evaluate the short-term forecasts using the proposed
method. It should be noted that a comparison between different tracking algorithms is hardly possible because they track different objects [1, 14].

Three cases of convective weather are chosen. They are case 1 (1300–2200 LST 7 July 2004), case 2 (1400–2000 LST 31 May 2005) and case 3 (1600–2000 LST 24 June 2006). The sampling rate is one volume scan per 6 min under the VCP21 mode. Table 1 shows the 18-, 30- and 60-min forecast evaluation results.

The results for case 3 are poorer than the other two cases because of many small erratic storms with a short lifetime on 24 June 2006, which makes storm tracking and forecasting very difficult. The 18-min result for case 1 is the best because the storms in case 1 were distributed sparsely and evolved smoothly, which makes it easy for tracking and forecasting. In case 2 on 31 May 2005, there were many dense clusters of storms which had relatively stable structures, and quite a few supercells lasted for a long time which extended the valid forecast period. From the evaluation results, the overall performance of the proposed method is satisfactory.

Table 1 also shows that both POD and CSI decrease with increasing forecast period, while FAR increases. This indicates that the forecast accuracy decreases at a very rapid rate. The forecast accuracy within 30 min is reliable, but is unreliable for 60-min forecast. The reason is that extrapolation techniques only utilize the history data to forecast and there is no accounting for storm initiation, growth and dissipation.

6. Conclusion and discussion

This paper proposes a novel convective storm identification, tracking and nowcasting method. A mathematical morphology-based storm identification method is adopted which can resolve the false merger problem and isolate storm cells well in a cluster of storms. Sequential Monte Carlo (SMC) method is utilized to track storms, which can handle complicated splits and mergers as well as the case of storm missing detection. For the important storm motion forecast, the proposed method takes the advantage of the cross-correlation method by using the motion vector field. Results of experiments presented in this paper show the efficiency and robustness of the proposed method.

It should be noticed that the proposed method is an automated algorithm which gives access to the temporal development of various properties of storms. Tanggu radar station, which can observe convective weather in Beijing,
has been in operation for a number of years. Based on the long history of Tanggu radar data, the future work is to use the proposed method to study the climatological characteristics of convective storms in Beijing and its vicinity.

Acknowledgments

This work was supported by National Basic Research Program of China (Grant No. 2004CB418300) and Ph.D. Programs of the Ministry of Education of China (Grant No. 20040001008). The authors thank Dr. Mike Dixon, NCAR, for providing information of TITAN and fruitful discussions and suggestions.

References