

## **Tracking and characterization of convective cells through their maturation into stratiform storm elements using polarimetric radar and lightning detection**

Jiaxi Hu<sup>\*(1)</sup>, Daniel Rosenfeld<sup>(2)</sup>, Dusan Zrnic<sup>(1)</sup>, Earle Williams<sup>(4)</sup>, Pengfei Zhang<sup>(1)</sup>, Jeffrey C. Snyder<sup>(1)</sup>, Alexander Ryzhkov<sup>(1)</sup>, Eyal Hashimshoni<sup>(2)</sup>, Renyi Zhang<sup>(3)</sup>, Richard Weitz<sup>(3)</sup>.

(1) Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma,  
and NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma

(2) Department of Atmospheric Sciences, The Hebrew University of Jerusalem, Jerusalem,  
Israel.

(3) Department of Atmospheric Sciences, Texas A&M University, College Station, Texas,  
USA.

(4) Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

*\* Corresponding author:* 120 David L Boren Blvd, Norman, OK 73072.

E-mail: [jjiaxi.hu@noaa.gov](mailto:jjiaxi.hu@noaa.gov)

**Keywords:** Cloud tracking, cloud microphysics, radar observation.

## **Abstract**

Polarimetric radars make it possible to retrieve information on hydrometeors types, sizes and concentrations. Additional information on cloud electrification can be obtained from Lightning Mapping Arrays (LMAs). To study the development time and height of the hydrometeors and electrification require tracking their evolution within the lifecycle of convective cells. A new methodology for multi-cell identification and tracking (MCIT) is presented in this study. The algorithm in this study is different from traditional tracking methods; this new algorithm is applied to time series of radar volume scans. It tracks local maxima of vertically integrated liquid (VIL) water by identifying the two cells in consecutive radar scans that have maximum common VIL. The vertical profile of the polarimetric variables is used for constructing the time-height cross section of the cells' microphysical properties around the peak reflectivity as a function of height. The LMA sources that occur within the cell area are integrated as a function of height as well for each time step, as determined by the radar volume scans. The result of the tracking can provide insights on the evolution of storms, hydrometer types, precipitation initiation and cloud electrification under different thermodynamic and aerosol conditions. The details of the MCIT algorithm, its products and their performance for different type of storms are described in this paper.

## **1. Introduction**

Convective clouds serve a fundamental role in heat transfer and precipitation. Observing their properties and evolution with time is essential for quantifying their behavior and its dependence on the thermodynamic and aerosol properties of their environment. This requires the identification and tracking of the convective clouds and the precipitation and lightning that they

produce (Gascón et al. 2015; Melcón et al. 2017; Merino et al. 2019; Michaelides et al. 2018; Sanchez et al. 2017).

Algorithms that identify and track convective storms have been used to study the evolution of convective storms in various scenarios (Dixon and Wiener 1993) and in weather forecasting, especially severe weather detection and short-term forecasting (Joe et al. 2004; Wilson et al. 2004). Two well-known types of tracking algorithms have appeared — the cross-correlation method (Li et al. 1995; Rinehart and Garvey 1978; Tuttle and Foote 1990) and the centroid method (Bellon et al. 1980; Bjerkaas and Forsyth 1979; Crane 1979; del Moral et al. 2018; Dixon and Wiener 1993; Handwerker 2002; Johnson et al. 1998; Kyznarova and Novak 2009; Lopez and Sanchez 2009; Munoz et al. 2018; Zan et al. 2019). The strength of a cross-correlation tracking algorithm is better identification of the shifting reflectivity echoes' speed and direction at the expense of the ability to identify and track isolated cells<sup>1</sup>. In contrast, the centroid tracking method can identify isolated cells and provide temporal properties of storms (Johnson et al. 1998). Because centroid tracking algorithms depend on threshold values to identify storm cells, they cannot capture well the initial and ending stage of cloud cells and are inappropriate for identifying and tracking stratiform precipitation. The thresholding techniques can also make incorrect identification of cloud cells if there are multiple cells with similar strength in close proximity to one another (i.e., within one cloud cluster).

An algorithm first developed by (Rosenfeld 1987) for the study of cells that are embedded in convective rain systems is neither a centroid nor a cross-correlation tracking algorithm. The algorithm uses watershed principles to identify cells by the outline of radar echoes and uses the largest common volume rain rate [measured in cubic meters of rain water which fall through the

---

<sup>1</sup> The term “cell” used in this paper refers to echoes associated with convective precipitation (whether a convective storm or “shower”) as seen in radar reflectivity data collected by a weather radar.

lowest elevation scan or through cloud base level over some time period ( $\text{m}^3/\text{hr}$ )] to track them. This method overcomes the disadvantages of the centroid tracking thresholding and is able to identify cells for both isolated and clustered cloud systems. The objective of this study is to develop a methodology that will document the life cycle of a convective cloud to the greatest possible extent.

Here we introduce the multi-cell identification and tracking (MCIT) algorithm, which is based on the concepts of Rosenfeld (1987), but with the advantages of modern image processing and polarimetric radars. The cell identification is based on the local maximum of VIL and the tracking between time steps is based on the maximum common VIL of the overlapped cells. The details of the MCIT algorithm will be discussed later in the methodology section. The goal of the MCIT algorithm is to create an automatic tool to identify and track cells from their first echo (regardless of whether they are isolated or clustered) and to provide cell lifecycle properties as revealed by polarimetric, lightning, and thermodynamic data. The algorithm produces the time-height evolution of hydrometeors and electrification in the tracked cells, giving users the ability to study more easily the microphysical evolution of each cell, something that is generally easier to do in a Lagrangian reference frame. This procedure is demonstrated in this study on various cloud cell types, from shallow warm showers to supercells. It is the first time that the time-height evolution of such a comprehensive package of cloud and precipitation properties is developed and shown. Although the evolutions of the microphysical properties demonstrated here is in line with expectations based on known cloud physics considerations and previous observations, the application of the new capability presented in this study also opens new research possibilities.

The structure of this study includes a Methodology section and a Results section. The Methodology section includes details about the MCIT algorithm structure and the assignment of



cell lifetime and instantaneous properties. The Results section includes examples of the time-height evolution of the properties of tracked cells. This is done for four types of cells: warm rain showers, an isolated thunderstorm, a cell that belongs to a squall line, and a supercell storm.

## 2. Methodology

This section explains the MCIT algorithm in detail. It includes data pre-processing, cloud cell identification, cloud cell tracking and post processing. Specifically, the parameters and the differences between the MCIT algorithm and the algorithm in Rosenfeld (1987) are described.

### a. Data and region of interest

#### 1. Polarimetric radar data

$Z$ : The equivalent reflectivity factor, which is proportional to the 6<sup>th</sup> moment of the raindrop size distribution.

$Z_{DR}$ : The ratio of radar reflectivity factors at horizontal and vertical polarizations expressed in dB, which is an indicator of the bulk shape of hydrometeors being sampled (e.g., raindrop shape and, following well-known models of raindrop shape, raindrop size). The combination of  $Z$  and  $Z_{DR}$  can provide information on hydrometeor type and rain rates. For example, the combination of high  $Z$  and low  $Z_{DR}$  is a common indicator of hail at the frequencies at which WSR-88D radars operate.

$K_{DP}$ : The specific differential phase shift on propagation, which is an indicator for rain water content and is not affected by attenuation, miscalibration, or partial beam blockage.

$\rho_{hv}$ : The copolar correlation coefficient between horizontally and vertically polarized returns. It is generally lower for mixed-phase hydrometeors than for single-phase hydrometeors. For instance, rain and snow echoes tend to produce high  $\rho_{hv}$  values (i.e.,

near 1.0), but mixed-phase hydrometeors such as those found in the melting layer can produce  $\rho_{hv}$  values generally below 0.97. Nonmeteorological scatterers (e.g., insects and birds) can also have much lower values of  $\rho_{hv}$ .

$N_w$  and  $D_m$ : The normalized raindrop concentration and the mean volume diameter, respectively, retrieved from  $Z$  and  $Z_{DR}$  as described by Tabary et al. (2011), and Ryzhkov et al. (2014).

## 2. Lightning data.

Two lightning networks are included – LMA (Lightning Mapping Array) and NLDN (National Lightning Detection Network). The LMA can provide both intracloud (IC) and cloud-to-ground (CG) lightning sources and radiated power in three dimensions. LMA sites are limited to certain regional locations. For research purposes, we have chosen Houston, TX LMA (HLMA) and Lubbock TX LMA (WLMA) as examples in this study. Both HLMA (12 sensor sites) and WLMA (12 sensor sites) solution files are output every 10 minutes with standard two pass filtering in order to reduce noise (Chmielewski and Bruning 2016). The WLMA solution files are created if there are 5 or more network sites reporting observations and HLMA solution files are created if there are 6 or more network sites responded. The geographic location of HLMA sites, WLMA sites and their corresponding nearby radar can be found in figure S1 and figure S2. For demonstration purpose, only the LMA source event information is used in this paper. The flash level data from LMA is not calculated.

NLDN can provide only accurate CG lightning locations and peak current. The merit of NLDN data is it can provide the polarity of each lightning event. The NLDN data covers the entire U.S. CONUS (Biagi et al. 2007).

### 3. Assimilation Data

The sounding data is from Eta Data Assimilation System (EDAS) (Air Resources Laboratory 2004) is utilized for thermodynamic information in this study. The spatial resolution is 40 km, and data are available every 3 hours. External assimilation data are not limited to EDAS. The flexibility of MCIT algorithm allows additional data input for further scientific needs.

#### *b. Definition of a cell*

A convective storm is initiated typically as an isolated or a cluster of convective cells that typically maintain their identity as reflectivity peaks during their growing stage. Upon maturation the cells areas expand on expense of their peak reflectivity, and eventually merge into low gradient stratiform area in the dissipation stage. Most cell tracking method (e.g., (Dixon and Wiener 1993)), are designed to track only the well defined cells during their growing stages, while ignoring reflectivities smaller than ~30 dBZ. Since the fraction of rainfall that is contributed by the stratiform component is highly variable and often exceeds 50% of the total (Schumacher and Houze 2003), it is important to track the maturing storm elements that can be attributed to the cells that contributed to them the most. Therefore, we track the full lifecycle of cells including the dissipation stage to the possible extent. The tracked entities are best termed as storm segments that can be attributed to the cells that initiated them. When we use the term "cell" hereafter, it pertains to the tracked storm segments.

A convective cell in its growing stage is composed of an updraft in which most of the condensation occurs in a vertical column. The cloud water is subsequently converted into precipitation, which falls through the updraft or which weakens the updraft and/or initiates a downdraft leading to the demise of the convective cell. Therefore, a vertically-integrated column of radar-retrieved precipitation represented by VIL is used as a marker for the core of a convective cell. VIL can be used as a quick and effective indicator of strong precipitation cells (Greene and Clark 1972b). VIL is calculated with a fixed Z-R relationship from the WDSS II software package (Lakshmanan et al. 2007). The VIL equation is as follows:

$$VIL = \sum 3.44 \times 10^{-6} \left[ \frac{Z_i + Z_{i+1}}{2} \right]^{\frac{4}{7}} \Delta h$$

Where  $Z$  (dBZ) is the radar reflectivity and  $h$  is the height expressed in meters.  $\frac{Z_i + Z_{i+1}}{2}$  is set to 56 dBZ if this term is over 56 dBZ in order to eliminate ice phase hydrometeor contributions (Greene and Clark 1972a). A cell peak is defined as a local maximum of VIL within a radar echo that is at least 5 grid boxes in size, where each grid box is 0.5 km x 0.5 km in horizontal extent and 0.3 km in vertical extent. Coordination and resolution details are discussed in the next section. The VIL map is smoothed by a mean filter in the horizontal over a square of size 3-by-3 grid points. Basically, grid point  $i$  is determined by the mean of data contained in the square centered on  $i$ . If grid point  $i$  has no detectable radar echo (i.e., and has designation “NaN” for Not a Number), then it will be preserved as “NaN”. At the edges of the VIL map, if the available grid points are not sufficient to build a 3-by-3 square, the smoothing program uses as many grids point as possible. Since the MCIT algorithm works better in relative strength, the algorithm uses  $10 \log_{10}(VIL)$  in units of dB. Two neighboring cells may appear to overlap but are still considered distinct as long as there are two local maxima in VIL separated by a valley that is at least 2 dB lower than the

weaker of the two maxima/cells. If the valley between the two peaks is not at least 2 dB lower than the maxima, the weaker cell is merged into the stronger one.

### *c. The MCIT algorithm*

The purpose of the MCIT algorithm is to identify and track multiple cells and save the time-height cross sections of various radar variables and lightning properties throughout their lifecycles. A flow chart of the MCIT algorithm (Figure 1) identifies the four stages of the cell-tracking analysis. A detailed description of the four MCIT stages follows below.

#### 1) STAGE 1: PRE-PROCESSING RADAR AND LIGHTNING DATA

The first stage is to process Level 2 radar data, LMA and NLDN data sets into Cartesian coordinates. The radar data for each case is from a single WSR-88D radar closest to the storm. The MCIT algorithm requires pre-processed 2-D VIL data and ERA interim 6-hourly wind field reanalysis as input. The WDSS II software package (Lakshmanan et al. 2007) is used to pre-process Level 2 WSR-88D radar data, analyze the data onto a Cartesian grid and calculate VIL. The interpolation method option is using a beam spreading technique in this study to determine the way data values are combined if they overlap in both space and time. The beam spread weighting is inversely proportional to the beam spread when the values of radar variables are interpolated onto the Cartesian coordinate. In other words, the wider the beam spreads, the less weight is assigned when interpolating.

The variables that are converted to Cartesian coordinates are reflectivity ( $Z$ ), differential reflectivity ( $Z_{DR}$ ), specific differential phase shift on propagation ( $K_{DP}$ ), copolar correlation coefficient ( $\rho_{hv}$ ) and VIL from the WDSS II package. The algorithm provides output once for each volume scan (usually every ~5 minutes), so that short-lived cells can be tracked. Vertical resolution

better than 1 km (Rosenfeld 1987) is desirable, but this cannot be achieved everywhere within the radar domain because of beam broadening and incomplete data coverage owing to gaps in elevation angles that exceed the beamwidth. In this study, the Cartesian grid is defined with vertical grid point spacing of 300 m. Since we desire the horizontal resolution to be less than the meso-gamma scale (2 km; (Orlanski 1975)), the horizontal grid spacing is 500 m. Owing to the trade-off between the maximum height of the radar domain (i.e., the height of the  $19.5^\circ$  elevation angle beam) and the increasing size of the radar volume with range, we use the range interval between 40 and 120 km. The maximum range was selected to be 120 km to limit the deterioration in resolution with distance. At the maximum range of 120 km the center of the lowest beam, at a tilt of  $0.5^\circ$ , is at a height of 1850 m and a beam spreading of 2 km between the half power points. The half power point of the upper part of the beam reaches the height of 4 km at the range of 120 km. To minimize contamination of ice hydrometeors at the base scan we limited the range to 120 km. Noted that Cartesian grid resolution (x-y-500m, z-300m) is not guaranteed once radar beam is far from radar sites due to beam broadening, beam spreading interpolation technique within the WDSS II package is used to interpolate coarse radar beam signals to finer Cartesian grid (Lakshmanan et al. 2007).

The hydrometeor classification algorithm (HCA) and hail size discrimination algorithm (HSDA) developed at the National Severe Storms Laboratory (NSSL) are used to provide hydrometeor classifications using radar observations (e.g.,  $Z$ ,  $Z_{DR}$ , and  $\rho_{hv}$ ) along with melting-level height using a fuzzy-logic scheme (Ortega et al. 2016; Park et al. 2009). The hydrometeor classification is calculated with the Cartesian gridded radar data. The classification categories and associated abbreviations are as follows: No Echo (NE), Light/Moderate Rain (L/MR), HR (Heavy Rain), Rain/Hail (R/Ha), Big Drops (BD), Anomalous Propagation (AP), Biological Objects (BI),

Unknown Hydrometeors Classification (UK), Dry Snow (DS), Wet Snow (WS), Ice Crystal (CY), Graupel (GR), Large Hail (LH), and Giant Hail (GH). HCA details can be found in Park et al. (2009) and Ortega et al. (2016) for discrimination between large and giant hail.

LMA and NLDN data are also converted to Cartesian coordinates and placed on the same grid as the radar data. Note that both LMA and NLDN data sets have much higher spatial and temporal resolution than the radar data, so for each Cartesian grid includes an integrated LMA and NLDN data spatially and temporally. Since the NLDN data include only X,Y locations for cloud-to-ground lightning , i.e. 2D data, the NLDN data are mapped onto cells in a 2D manner as identified by VIL without any vertical distribution.

## 2) STAGE 2: CELL IDENTIFICATION AND INSTANTANEOUS PROPERTY ASSIGNMENT

The second stage of the MCIT algorithm uses the calculated VIL data, one scan at a time, as follows:

- (i) A watershed algorithm (Meyer 1994) identifies the boundary of local maximum VIL clusters of at least 5 grid boxes in size. The criterion for cells to be at least 5 grid boxes ( $1.25 \text{ km}^2$ ) is to avoid noise and over-identification of cells. The size threshold of 5 grid boxes are empirical and can be changed for further needs. Each local maximum VIL cluster is assigned a local ID number. The MCIT algorithm is written in MATLAB, which offers a 'watershed' function. The command used was simply "watershed(input, options)".
- (ii) Local maxima of VIL are considered as peaks of cells. All neighboring cells are tested for potential merging if the depth of the saddle point between them is shallower than 2 dB (Rosenfeld 1987). The 2-dB threshold is an empirical parameter in the MCIT algorithm and can be easily changed to fit different research needs.

- (iii) The watershed algorithm ascribes grid boxes to each peak by locating watershed ridgelines in the VIL field. The sign of the VIL is reversed before the watershed algorithm is applied to it. The reason for reversing the sign is that the watershed algorithm assumes that water flows from high to low elevations. Therefore, the peaks get the lowest values when multiplied by -1.
- (iv) Each watershed region is assigned a local ID number when it also satisfies the size criterion of at least 5 grid boxes in this study.
- (v) Next, we determine if local cells are neighbors and should be considered for merging. For each local cell with local ID  $j$ , the program searches for peaks of other cells within a radius twice the cell's longest dimension ( $R_j$ ). Cell  $j$  and the peaks of the other cells within the radius of  $R_j(k_1, k_2, \dots k_n)$  are examined. If there are grid boxes not assigned with local ID numbers of either  $j$  or  $k_i$  between the peak of cell  $j$  and cell  $k_i$ , then cell  $j$  and cell  $k_i$  are not neighbors. Only when the grid boxes between the peak of cell  $j$  and cell  $k_i$  consist of local ID numbers of  $j$  and  $k_i$  exclusively can cells  $j$  and  $k_i$  be treated as neighbors and then tested for merging by the VIL 2-dB comparison. After searching all the neighboring peaks and after the merging process, all the grid boxes with valid VIL values have been assigned to cells with a unique local ID number.
- (vi) For every local ID number that remains after merging, the program calculates and saves its instantaneous properties in a data structure for later use. The lightning properties that are added to the cell properties are:
- LMA source count
  - NLDN total flash count
  - NLDN positive flash count



- NLDN negative flash count
- (vii) The program also saves the horizontal map of VIL and the local identification ID from the current time step. These VIL and ID maps are the input for stage 3.

### 3) STAGE 3: MULTI-CELL TRACKING

The third stage of the MCIT algorithm tracks cells by comparing two consecutive VIL maps from stage 2. The temporal resolution is about 5 minutes for VIL maps, which is the time between radar volume scans that underlie the VIL grids. The assumption is that the duration of the targeted cells that are to be tracked is longer than 5 minutes. A tracking error may occur if a targeted cell dissipates between two tracking steps and a newly formed cell develops around the location of the targeted cell between these two tracking steps. The false track is typically easy to identify owing to the sharp change in the reflectivity and height of the cell.

The cell tracking flow chart can be found in 《Rosenfeld (1987), Table 3》. The tracking algorithm summary and difference from Rosenfeld (1987) are as follows:

- (i) The MCIT algorithm compares a pair of sequential VIL maps for overlapping cells. This is done after shifting the VIL map at time step  $n+1$  back to account for the mean cell motions between the time of maps  $n$  and  $n+1$ . The shifting vector is updated for each cell at each time step based on the remaining distance between the cell peaks in the two maps after the shift back. The detailed description of the shifting vector is discussed in stage 3 subsection (iii). The program keeps track of all the overlapping cells. One cell from map  $n+1$  can overlap with more than one cell from map  $n$ . This may be a result of merging of cells in map  $n$ . On the other hand, one cell in map  $n$  can overlap with more than one cell from map  $n+1$ . This can be a result of the cell splitting in the map  $n$  to several cells in the

map  $n+1$ . The most important quantity for tracking is the overlapping or common integrated VIL. For each pair of overlapped cells between time step  $n$  and  $n+1$ , their common VIL values are integrated within the overlapped pixels and saved for tracking criteria in the step next.

- (ii) This step does the actual tracking and is based on the logic in Rosenfeld (1987), Table 3, with the difference that the present method uses the common VIL and VIL peaks instead of a common rain volume and reflectivity peaks. The program searches for all the overlapped cells at time step  $n+1$  relative to targeted cell  $(i,n)$ , i.e., cell  $i$  at time step  $n$ . The criteria to identify one of the cells at time step  $n+1$  to be the continuation of cell  $(i,n)$  are:
- 1) The integrated common VIL is at least 50% of the total VIL of the smaller of the two compared cells.
  - 2) The VIL peaks of both old and new cells are inside the common area of the two compared cells.
  - 3) If the previous condition (peaks of old and new located within the common area of the two cells) is not fulfilled it is required that the integrated common VIL must be more than 75% of the total VIL of the smaller of the two compared cells.
  - 4) If none of the conditions above is fulfilled, but the integrated common VIL is at least 25% of the total VIL of the smaller cell, then the cell from time step  $n$  can either be a merger or a split.

According to many tests and experiments, the new tracking code performs similarly to the original code (Rosenfeld 1987) in that the tracking results are not sensitive to the percentage limits of integrated common VIL. The percentage numbers are usually close to 100% or smaller than 50%.

There are four possible situations for candidate cells to emerge at time step  $n+1$ : simple continuation, simple birth, split, and merger. A simple continuation status occurs when  $\text{cell}(i,n)$  is the only source for  $\text{cell}(i,n+1)$  and  $\text{cell}(i,n+1)$  is the only source for  $\text{cell}(i,n)$ . A simple birth situation is when  $\text{cell}(i,n+1)$  has no source cell from the previous time step. If  $\text{cell}(i,n)$  has more than one continuation cell, only the one with maximum integrated common VIL gets the identity of  $\text{cell}(i,n)$  and all the rest of candidate continuation cells are labeled splits. A merger situation is similar to splits;  $\text{cell}(i,n+1)$  has more than one source cell and is therefore labeled as a merger.

At this stage of the MCIT algorithm, each cell has been identified and tracked. A global ID number is assigned to each cell. Unlike the local ID numbers, the global ID number is a unique number assigned to each cell through its lifetime and is used as the identifier to assign each cell lifetime and instantaneous properties. The commonly used parameters are also available in the appendix.

Two types of tracking examples are presented here to show the performance of the MCIT algorithm, i.e. isolated tracking and clustered tracking. An isolated tracking example (Figure 2) is shown from June 8, 2013, using the KHGX (Houston, TX) radar. The figure contains both the VIL panels (Figure 2a - f) and the corresponding tracking panels (Figure 2g - l). The extracted tracking steps are from 19:51:04 UTC to 20:17:00 UTC to the northeast of Houston. The cells included in Figure 2 are about 70 to 90 km from the KHGX radar site. Each color and the related black number in figure 2g -l links to a unique cell. By examining each VIL map and comparing with the corresponding partition map, cells are well identified. The general direction of motion of these cells is from the west to east. Cells 1, 2, 3, 4 and 7 are relatively stronger cells in these maps deduced through inspection of

their VIL peak magnitudes. Cells 5 and 9 are relatively weaker cells and have shorter lifetimes compared with other cells.

A clustered tracking example (Figure 3) is shown for August 8, 2016, as seen by the KLBB (Lubbock, TX) radar. Figure 3 has the same layout as Figure 2 for both VIL and the corresponding tracking partitions. The example is from 23:34:28 UTC to 00:00:33 UTC east of KLBB and 100 to 120 km from the radar. These clustered cells are usually more difficult to identify and track, especially when convective cells are surrounded by stratiform storm elements. By comparing pairs of the VIL panels with the partition panel, one can be confident in the multiple cell identification capability of the MCIT algorithm. The general direction of motion of these clustered cells is from the west to the northeast. To validate the tracking capability of the MCIT algorithm, each pair of adjacent time steps is examined. The major cells are 2, 3 and 11 in these clustered cells. Cells 1 and 12 are mergers and merge to cells 2 and 11 respectively. Cells 4 and 8 are surrounded by stratiform storm elements but uncoupled from other convective cells. Cells 5, 6, 7, 13, 14, 15, 20, 21 and 22 are close to the coordinate boundary in this example and will not be discussed here.

- (iii) The cell motion vector represents the movement of the cell peaks between time steps. According to Rosenfeld (1987), this motion is well correlated with the wind at the steering level (700 mb – 850 mb), so that cells in the same region usually have similar motion vectors. Therefore, shifting back the map by the statistics of the cell motion vectors in a given region generally works well for tracking individual cells. The shifting vector or back-shifting vector is constructed from a group of motion vectors to represent the general motion of cells and provides the tracking code with a reasonable first guess. The shifting

vector of the maps between time steps is a key part of the MCIT algorithm and is updated at each time step as follows.

Based on the motion vector from previous time steps, each pair of tracked cells between two consecutive time steps has a motion vector and the program saves all the motion vectors. The initial shifting vector uses the mean wind at the steering level.

The full radar map then is divided into 9 equal rectangular parts of the entire domain. Each part in this paper represents a rectangular area with dimensions  $100\text{km} \times 85\text{km}$ . The histogram of the u and v components of the cell motion vectors is constructed for each of the 9 sub domains. The median u and v components are used for constructing the shifting vector for all the cells within the sub domain. The median u and v components for the full domain are also recorded. The u and v data are collected from all cell motions in order to construct shifting vectors. All histograms and shifting vectors are saved.

After the initial shift between two consecutive VIL maps by the background wind, the remaining shifted distance of the paired identified continuation cells will be considered as a candidate motion vector to update the shifting vector for the next time step. We need to have at least 25 motion vectors, up to 2 hours search range backward in time, to update the shifting vectors. We calculate the median value of the updated motion vectors. The median updates in the motion vector are accepted only if the magnitude of its median  $\pm 2$  pixels in both x and y dimensions range covers at least 70% of the matched cells. If not, we use the previous shifting vectors with no updating.

#### 4) STAGE 4: ASSIGNMENT OF LIFECYCLE PROPERTIES

The last stage of the MCIT algorithm is the construction of the time-height evolution of the observed polarimetric variables and LMA properties of the tracked cells. The peak reflectivity within the cell area at each height is taken to form the vertical reflectivity profile. The vertical profiles of the other polarimetric variables are constructed by taking samples from the same locations of these peak reflectivities at each height. The LMA lightning sources that occur at the various heights anywhere within the cell area are counted. These time-height histories of the cells are saved in a database with a record for each cell.

For example, cells 590, 3401, 2 and 1790 were detected by the MCIT algorithm on June 11, 2016 at KLBB; July 17, 2015 at KHGX; Aug 08, 2016 at KLBB and on June 11, 2016 at KLBB, respectively. Selected properties of these cells as a function of time are plotted in Figure 4. Both cell 3401 and 590 started as isolated cells. Cell 3401 is a warm rain cloud and cell 590 is a deep convective cloud with mixed-phase processes. Cell 2 from Figure 3 is a squall line cell and cell 1790 is a supercell. It is generally found that the reflectivity peaks first, followed by echo top height, rain volume rate, and the cell's maximum area. The lightning rate (lma cc sum) peaked near the time of the echo top maximum height (Williams 1985; Yoshida et al. 2009). Positive lightning flashes are much rarer than negative lightning flashes as is widely known, and the pattern of total CG lightning flashes looks similar to that of negative flashes. This is mostly the case for the cells shown in Figure 4.

With these data in hand, the fourth stage of the automatic procedure calculates and saves the lifecycle properties as follow:

- 1) Time-Height properties – Composites of the time-height evolution of each cell's radar variables (i.e.,  $Z$ ,  $Z_{DR}$ ,  $K_{DP}$ , and  $\rho_{hv}$ ), results of hydrometeor classification, the mean raindrop volume diameter ( $D_m$ ), the normalized concentration ( $N_w$ ) retrieved from the polarimetric radar variables and the counts and powers of lightning sources from LMA data. This is demonstrated here for four different types of cells (Figures 5, 6, 7 and 8). Each point in time and height in the different panels shows a polarimetric radar variable and the result of classification at the location of the reflectivity peak of the cell at that height and time. The values in the LMA panels show the integrated source counts within the cell area at that height for each point in time and height.
- 2) Volume rain rate (RVR) [ $m^3/hr$ ] – The integrated rain rate, which is calculated by the method from (Sachidananda and Zrnić 1987) over the cell area at each time step, is saved as shown in Figure 4.
- 3) Maximum values – Maximum values throughout the entire life cycle of the cell for  $Z$ ,  $Z_{DR}$ ,  $K_{DP}$ ,  $\rho_{hv}$  (minimum only), cell area, echo top height, volume rain rate and etc. The time of each parameter's maximum during each cell's life cycle is also saved. Lifetime evolutions of these values during the cell lifetime are presented in Figure 4.
- 4) Location – The azimuth and range of each cell from the radar site is also saved at each time step. To ensure accuracy, all the cells in this paper are located between 40 and 120 km from the radar sites (KHGX in Houston/Galveston and KLBB in Lubbock, in this paper). A minimal range of 40 km was set to avoid excessive truncation of echo tops, because the radar's vertical coverage extends up to 14 km AGL starting at that range.
- 5) Cell status and complexity – The four flags discussed in stage 3 (simple birth, simple continuation, and merger and split) are saved for each cell. At each time step, the cluster in

which each cell is embedded can have more than one cell. The complexity keeps a record of the number of cells that make up the cluster at each time step during each cell's life cycle.

- 6) Thermodynamic parameters – The sounding data is from Eta Data Assimilation System (EDAS) (Air Resources Laboratory 2004). The spatial resolution is 40 km, and data are available every 3 hours. Melting level and Convective Available Potential Energy (CAPE) are saved for each cell. Since the temporal and spatial resolution is relatively coarse compared to the radar data, the program searches for the sounding data closest to each cell.
- 7) Satellite retrievals – By using the same method as in (Rosenfeld et al. 2016), for each cell a record is kept of its cloud base CCN concentration [ $\text{cm}^{-3}$ ], cloud base supersaturation, cloud base temperature and cloud base vertical velocity. Although the satellite has only one data record for each case (around 13:30 solar time) and the CCN concentration is slowly varying, it is safest to use the retrieved CCNs for cells present at times close to the satellite scan time (+/- 2 hours of satellite overpass time).

### **3. Results**

Here we present selected results that are intended to show the capabilities of the MCIT algorithm. The selected well tracked cells must be at their growing stage when first identified and go through maturation and reach the dissipation stage, as demonstrated in figure 4. To exclude poorly tracked storm segments, additional quality criteria for well tracked cells are applied, including the lifecycle duration is over 30 minutes and the first echo top height is less than 9 km. Four types of convective storm segments are shown: i) Isolated convective storm with mixed-phase processes; ii) isolated convective storm with mainly warm rain processes; iii) convective



storms in a large cluster; iv) supercells. Examples of different types of convective storms demonstrate the applications of this algorithm.

*a. Isolated deep convective storms with mixed-phase processes*

A common type of convective cloud/storm is a quasi-barotropic atmosphere is the isolated convective cloud with mixed phase processes and lightning events. Since these clouds/storms are mainly isolated with few splits or mergers, they are relatively easy to identify and track, especially when they are deep with mixed-phase processes that indicate a stronger updraft and more significant VIL signals compared to shallower clouds with little or no mixed-phase precipitation present.

An example is selected from June 6, 2016, near Lubbock, TX (KLBB radar). The cloud cell id is 590 and was initiated at 32.82 °N 101.83 °W at 18:54 UTC, 93 km from the radar station at an azimuth angle of 181°. Figure 5a shows the time–height evolution of maximum reflectivity in dBZ. In this paper, the first echo height is defined as the height of the maximum reflectivity in the vertical at the first time it exceeds 30 dBZ. The first echo height of cell 590 is detected at around 6.5 km (i.e., about 2 km above the environmental 0° C level), and it continues to grow for 25 minutes until the echo top reaches about 16 km. Dissipation in this paper is characterized as the period after which the reflectivity core has reached its maximum height and during which the reflectivity decreases. The dissipation starts after 30 minutes of initial identification and lasts for another 55 minutes. The red numbers under each Z column are the complexity values, which shows how many cells is the cluster in which the cell resides, and the magnitude is  $\log_2(N)$ , where N is the number of cells in the cluster in which the tracked cell resides.

Convective processes can be inferred from the time-height cross section. For example, Figure 5a shows a strong reflectivity core (i.e.,  $Z > 60$  dBZ) at  $t=30$  min and  $H=8$  km, which descends and reaches the surface after 5 to 10 minutes along the marked black line. The low value of  $Z_{DR}$  in the core of this precipitation shaft marked by the black line indicates that this core is primarily composed of hailstones which start melting at a height of 3 km as indicated by the abrupt increase of  $Z_{DR}$  with values up to about 2 dB; this increase can be caused by complete melting of smaller-size graupel / hail particles, by partial melting of larger-size hailstones which acquire an external water coating, and by the emergence of very large raindrops having ice cores inside (Ryzhkov et al. 2013). The low values of  $\rho_{hv}$  (less than about 0.95, Figure 5d) through the precipitation shaft indicate that the hail is not completely melted all the way to the ground. The green line in Figure 5b marks a  $Z_{DR}$  column that represents the largest raindrops which form in the convective updraft, fall from the updraft area ahead of smaller-size raindrops, and reach the surface some 5 to 10 minutes before the main precipitation shaft (Kumjian and Ryzhkov 2012). The high values of  $Z_{DR}$  (colored red,  $Z_{DR} \sim 4$  dB) extend up to 7 km and are indicative of large, supercooled raindrops or other “wet” hydrometeors. The relatively low  $\rho_{hv}$  there indicates the existence of mixed phase (e.g., partially frozen raindrops or wet hail) rather than purely liquid supercooled raindrops. The low  $K_{DP}$  values within the  $Z_{DR}$  column (Figure 5c) along with the large reflectivities suggest that the column is composed of sparse large falling raindrops lacking small drops (which results in low rainwater content) or, in parts, relatively small concentrations of wet hail.

The hail shaft transforms into a heavy rain shaft some 10 minutes later, as indicated by the large  $K_{DP}$  values, marked by the blue line. The  $K_{DP}$  cross section (Figure 5c) shows high values that extend well below the melting level in the main precipitation shaft (black line). This serves as an indication that hail or large graupel continues to fall and melt into rain at a height of  $\sim 3$  km

between 45 and 60 minutes into the cell lifetime. The increase of the hydrometeor melting level from time 60 minutes onward is evident in the  $Z_{DR}$  cross section by the increasing height of the top of the high  $Z_{DR}$  region. The retrieved large raindrop sizes and their respective low concentrations at low levels are shown in Figures 5e and 5f, respectively during the times before the onset of the major precipitation after 35 min. These signals are likely caused by differential sedimentation of large drops falling from the updraft. The indicated values for  $D_m$  and  $N_w$  are not valid for frozen hydrometeors, and only data classified as L/MR, HR and BD (Figure 5h) are plotted in color in the  $D_m$  and  $N_w$  panels (Figure 5e and 5f). It is possible to discern the locations of raindrops with various  $D_m$  and  $N_w$  using the polarimetric variables  $Z_{DR}$  and  $K_{DP}$ . For instance, at lower levels (e.g., below 4 km) during the cell's dissipation stage ( $t = 60$  to  $t = 90$  minutes), a systematic pattern of decreasing  $D_m$  (Figure 5e) and increasing  $N_w$  (Figure 5f), likely caused by fallout of large raindrops and dominance of smaller raindrops before dissipation, is apparent. The masked grid boxes in the  $D_m$  and  $N_w$  panels (black dots in Figure 5e and 5f) due to mixed/solid phase processes are suggested in the  $\rho_{hv}$  values (Figure 5d;  $\rho_{hv} < 0.9$  is a strong signal of hail and graupel) and in the DHCA (Figure 5h) panels.

The number of lightning sources (Figure 5g) begins increasing rapidly approximately 5 minutes after the collapse and freezing of the  $Z_{DR}$  column. The lightning flashes begin aloft (probably as intracloud flashes) and descend to the ground with the main precipitation hail shaft (Figure 5h). Flash rate can be calculated with the MCIT algorithm package, but in this paper we are concentrating on the lightning source time-height patterns without grouping sources into flashes.

The DHCA panel (Figure 5h) shows the pattern of different inferred hydrometeor habits. Most of hail and large graupel undergoes wet growth within the updraft at a height of above 4.5

km and starts melting below that level in the downdraft area containing the heaviest precipitation from  $t = 30$  to  $t = 55$  minutes after detection. Most solid phase hydrometeors comprising this echo are small hail categorized as R/Ha, dry graupel (GR), dry snow (DS) and crystals (CY). No large hail (LH), i.e., hail exceeding 2.5 cm diameter, is shown in DHCA. Dry graupel (GR) is the inferred dominating hydrometeor above 6 km. During the mature stage ( $t = 30$  to  $t = 50$  minutes), most hydrometeor signals below 6 km are R/Ha type, indicating rain and small hail falling to ground during this time period. Inferred heavy rain (HR) and light / moderate rain (L/MR) dominate the dissipation stage near the surface ( $z < 4.5$  km) from  $t = 55$  minutes to the end.

*b. Isolated convective cell with mainly warm rain processes*

A typical isolated convective cell with most precipitation formed by the warm rain process can occur in a slightly unstable marine air mass. Warm rain clouds usually have relatively low echo top height that does not extend much above the freezing level, a slower growth rate of echo top height, and less intense precipitation than mixed-phase convective clouds and storms.

An example is selected on July 17, 2015, near the Houston/Galveston KHGX radar. The cloud cell id is 3401 and was initiated at 28.85 °N 94.98 °W, at 17:05 UTC, 48 km from the radar with an azimuth angle of 172°. Cloud cell 3401 was detected over the ocean. The first-echo height of cell 3401 is detected at near 4 km (Figure 6a). The cell continued to grow for 20 minutes until the echo top reached around 7 km and subsequently dissipated in 15 minutes.

The reflectivity core is relatively weak (i.e.,  $Z < 50$  dBZ) compared to the previous example. The black line marks the descent of the reflectivity core to the ground in less than 10 minutes and represents the rain shaft of cell 3401. There are no clear  $K_{DP}$  and  $Z_{DR}$  enhancements in this cell. The low  $K_{DP}$  and moderate reflectivity magnitudes indicate that this cell has low rainwater content.

Inspection of the  $Z$ ,  $Z_{DR}$ ,  $K_{DP}$ , and  $\rho_{hv}$  panels show no indication of mixed-phase hydrometeors. Figure 6g shows no lightning activity (both LMA and NLDN) in this cell. In the warm rain scenario, pre-existing rain embryos quickly grow by collision-coalescence to large raindrops which are sparser at the growing stage ( $t = 0$  to  $t = 15$  minutes) of the cloud. This is suggested by the relatively high  $Z_{DR}$  (Figure 6b) and  $D_m$  (Figure 6e) combined with low  $N_w$  (Figure 6f) values below 3 km during this period. A closer look at the descending pattern of  $Z_{DR}$  (Figure 6b,  $t = 10$  to  $t = 15$  minutes) shows that the higher  $Z_{DR}$  occurs ahead of the main rain shaft as indicated by the large  $Z$  values (Figure 6a,  $t = 15$  to  $t = 25$  minutes). These high  $Z_{DR}$  values at low level are likely due to the early fallout (i.e., higher fall speeds) of the heavier and larger raindrops. During dissipation ( $t = 15$  to  $t = 35$  minutes) more numerous smaller raindrops are inferred to dominate the drop-size distribution. This is likely because most large raindrops have already fallen out. This transition is indicated by the relatively low  $Z_{DR}$  (Figure 6b),  $D_m$  (Figure 6e) and high  $Z$  (Figure 6a) and  $N_w$  (Figure 6f) values during this stage. The results of the DHCA (Figure 6h) indicate that this warm rain cloud consists mainly of L/MR, which is consistent with the typical warm rain process discussed above.

### *c. Multi-cell convective storms*

Multi-cell clouds/storms usually refer to clusters consisting of several clouds/storms each of which is at a different stage of development. This type of storm can bring heavy precipitation, hail, frequent lightning, and high wind, but they rarely produce tornadoes. Multi-cell clouds tend to have a longer lifetime and more intense precipitation than isolated convective clouds, except for supercells.

An example of a tracked cell from a squall line is selected on August 8, 2016, near the Lubbock, TX (KLBB) radar. It is cell #2 in the horizontal tracks (Figure 3) that were discussed in the methodology section. The cell was initiated at 34.25 °N 100.81 °W, at 18:52 UTC, 114 km from the radar at the azimuth angle of 54°. The first-echo height of this cell is detected at about 8 km and the cloud continues to grow for 30 minutes until the echo top reaches 18 km.

Three discernible cycles are visible in  $Z$  (Figure 7a). The main precipitation shafts are marked by the black lines, and each line represents the regeneration of this squall line. The combination of negative  $Z_{DR}$  (Figure 7b) and low  $\rho_{hv}$  (Figure 7d) coinciding with very high reflectivities ( $> 60$  dBZ in Figure 7a) is a strong indication of large hail and graupel (Park et al. 2009). The second and third black lines in each panel are well-defined signatures of hail shafts that reach ground level. The strongest hail shaft belongs to the third cycle between 120 to 135 minutes, as indicated by the lowest  $\rho_{hv}$  region in Figure 7d. Figure 7c indicates occasional high rainwater content at low levels, apparently as a result of the partial melting of the ice hydrometeors. The high (colored red)  $K_{DP}$  combined with very low or even negative  $Z_{DR}$  and low  $\rho_{hv}$  suggest heavy precipitation combined with wet hail and graupel. Between adjacent cycles delimited by the black lines, the regeneration of the squall line is most obvious in the  $Z_{DR}$  panel. For instance, within the time interval between  $t = 100$  min and  $t = 120$  min, the large positive (yellow to red) values in the  $Z_{DR}$  panel (Figure 7b) with relatively smaller reflectivity (Figure 7a) indicate that the liquid phase raindrops are dominating during this time slot. Before  $t = 115$  min, when the hail shaft of the third cycle starts, large raindrops begin falling at  $t = 110$  min as inferred by the large values of  $Z_{DR}$  (Figure 7b). The intensity of each regeneration is stronger than in the previous cycle. This is indicated by the evolution of  $Z$ ,  $Z_{DR}$ , and  $\rho_{hv}$  along the black lines. No obvious pattern can be found in  $D_m$  and  $N_w$  panels probably due to the overwhelming presence of ice or mixed-phase

hydrometeors, as suggested by the DHCA product (Figure 7h), which inhibits the retrieval of  $D_m$  and  $N_w$ .

The LMA observations (Figure 7g) shows extremely high lightning activity (LMA source counts) from 30 min to 155 min after cell initiation. The lightning activity follows the pattern of the hail shafts. The lightning activity reaches maximum source counts at the same time the third hail shaft reaches maximum intensity.

From the DHCA figure (Figure 7h), the three major precipitation shafts marked by the black lines are dominated by inferred R/Ha and LH (2<sup>nd</sup> and 3<sup>rd</sup> precipitation shafts) and reach ground level during at least two periods marked by two most intense precipitation shafts ( $t = 80$  to  $t = 100$  min and  $t = 120$  to  $t = 135$  min). LH is inferred to reach the ground during the 3<sup>rd</sup> precipitation shaft in figure 7h, consistent with the  $\rho_{hv}$  panel (Figure 7d), showing the  $\rho_{hv}$  value is less than 0.9. A gradual transition is evident around every 5 minutes with increasing height from inferred pure rain (L/MR, HR, and BD) to small hail mixed with rain (R/Ha), then to GR, DS and CY.

#### *d. Supercell*

In the thunderstorm spectrum, supercells are the least common type of thunderstorm, but they have the highest propensity to produce severe weather, including damaging winds, very large hail, and tornadoes. What makes a supercell unique from other convective storm types is that it contains a deep and persistent rotating updraft called a mesocyclone. If the environment is favorable, supercell thunderstorms can last for several hours, and so are often easily tracked with algorithms of the kind considered here.

An example is selected on June 11, 2016, near Lubbock, TX (KLBB radar). The cell id is 1790 and was initiated at 33.77 °N 101.03 °W, at 20:18 UTC, 74 km from the radar station with an azimuth angle of 80°. The first-echo height of cell 1790 is detected near 7 km and continues to grow for 30 minutes until the echo top reaches 18 km. The supercell echo top quickly increases to over 18 km in 30 minutes with a strong reflectivity core (e.g.,  $Z > 60$  dBZ). The dissipation begins after 55 minutes of cell 1790's detection and lasts for 70 minutes. This supercell's lifecycle is longer (greater than 2 hours) than the lifecycle of an individual cell in a squall line (e.g., less than 1 hour) as in the previous storm example. Of particular note in the present case is the long duration (i.e., 30 min) of continuous high radar reflectivities and inferred hail, which is common characteristic of a supercell (Browning 1977; Moller et al. 1994) .

The cross section of reflectivity indicates that the major precipitation shaft starts around 25 minutes after the first echo. It is marked by the black lines in each panel of Figure 8. The low  $Z_{DR}$  and  $K_{DP}$  (Figure 8) at the core of high reflectivity indicate a strong signal of a hail shaft from 25 to 55 min. The areas of high  $K_{DP}$  (red) and low  $\rho_{hv}$  at lower level (e.g., below 6km) likely indicate wet hail (Figure 8c and 8d). The  $\rho_{hv}$  cross section (Figure 8d) indicates that the first detection of hail occurred at  $t = 20$  min, as inferred by the low values of  $\rho_{hv}$ . The low  $Z_{DR}$  along the line of maximum  $Z$  indicates that the hailstones reached ground level. According to the inferred radar hail signature, the hail production and fallout in this supercell lasted for 40 minutes and was much longer than that in the squall line example ( $< 20$  min) as shown in Figure 7. The onset of the dissipation stage is indicated by the rising of the hydrometeor melting level from  $t=50$  min onward, which is inferred by the increasing height of the relatively large  $Z_{DR}$  (yellow to red areas in Figure 8b). This signature is due to hail melting below 0 °C isotherm. Larger ice hydrometeors melt further below the 0 °C isotherm. Once the supercell starts to weaken, the hail weakens and starts



melting at greater heights. The decreasing raindrop size (Figure 8e) and increase in their concentration (Figure 8f) from 70 min also indicate the decay of this cell.

The lightning panel (Figure 8g) shows that lightning initiated right after the hail/graupel became evident (15 min) and intensifies once the hail shaft is obvious (25 min). During the mature stage of this supercell ( $t=25$  min to  $t = 55$  min), the maximum height of the lightning activity is close to the echo top height, and the base of lightning activity practically reaches the surface. As the supercell dissipation starts, the top height of the lightning activity descends to about 10 km while the echo top height remains at 15 km. At the same time, the base of the lightning activity rises to a height of  $\sim 5$  km, indicating a decrease of cloud-to-ground lightning activity and more stratiform cloud-to-cloud lightning during the supercell's dissipation stage. The lightning height top descends with the 25 dBZ reflectivity contour line at nearly the same rate (Figure 8a and 8g around 3m/s), which is likely due to the descent of large hydrometeors as a result of the weaker updraft characterizing the weakening cell. The fall speed is estimated by the mean height of the descending 25 dBZ contour. This pattern of lightning reaching the radar echo top during the active convective phase of a cell and descending as the cell starts to decay was also observed in other thunderstorms that were tracked through their decaying stages.

From the DHCA figure (Figure 8h), a clear hail shaft is inferred from  $t = 25$  to  $t = 55$  minutes after detection of “big drops” (BD), and its transition to L/MR (at  $t = 5$ –10 minutes) and HR ( $t = 5$ –20 minutes) and then to solid phase hydrometeors (R/Ha, LH and GR from  $t = 15$  minutes) is evident. The inferred HR at low levels (e.g., below 4 km altitude) is associated with GR and R/Ha at heights above 4 km. Comparing figure 8g and 8h, the core of LMA source counts descends after the core of GR falls downward. After surface rain became weaker (L/MR from  $t = 105$  min to the end), the upper level (e.g., above 4.5 km) GR is replaced by DS.

#### **4. Summary and Conclusion**

The MCIT algorithm uses 2-dimensional VIL maps as input for cell identification. A watershed algorithm divides VIL maps into individual cells. The tracking of cells between time steps is implemented by comparing the maximum common VIL, which is determined by the shifting vectors calculated from the median shifting vector from previous time steps. Each cloud cell is given a unique global ID number, and the life cycle information from polarimetric radar, lightning, satellite retrieval and sounding data is saved in the output data structure.

The MCIT package was applied here as a demonstration on different types of convective storms. The output from the MCIT algorithm is a comprehensive and expandable dataset of cell microphysical properties, which can be used to study storm structure and electrification,

This study presents the time-height evolution of hydrometeors and electrification in convective clouds, from shallow warm showers to supercells. The evolution of the tracked cells conforms with physical expectations based on prior knowledge (Dotzek et al. 2009; Putsay et al. 2009). This indicates that the tracking algorithm performs well. For example, the first raindrops that fall are large and scarce, whereas raindrops become more numerous and the mean size of the drop-size distribution becomes smaller in the mature phase of the convective cell (e.g., panels e and f of Figures 5–8). Lightning is initiated only when the precipitation freezes (appearance of solid phase hydrometeors like graupel and hail) and starts to descend, coincident with rising echo-top height, a manifestation of the charge separation mechanism normally involving negative charge on the large ice hydrometeors in the descending precipitation shaft and ascent of the normally positive charge on the ice crystals in the rising cloud top (e.g., panels a, b, g and h of

Figures 5, 7, and 8). The height of lightning sources descends with the maturation of the storm with the descending graupel and snowflakes (fig 8g and 8h).

Although the evolution patterns are in line with-known cloud physics and previous observations, it is valuable to have them combined in the time-height evolution of convective cells and the potentially new insights are expected. For example, the dataset output from the MCIT algorithm is useful for the study of individual convective storms. For instance, the effect of aerosols on cloud-tracked properties can be studied by analyzing the tracked cells under different CCN scenarios. The application of MCIT for these objectives is currently in progress.

*Acknowledgments:* Funding for this study is provided by the National Science Foundation, grant 1523325. The authors thank Eric Bruning, Don MacGorman and Crystal Nassir for providing LMA and NLDN data. Gratitude is also extended to John Krause for assistance with the DHCA code.

**Appendix.** The list of tracked cells attributes

lma_1_step	The time step of the cell tracking with first LMA detection.
gid	Cell global ID number.
region_flag	Region flag, 1 = ocean, 2 = coastal, 3 = inland.
Urban_Dis	Distance from the location of cell's initial detection to Houston center (km).
Urban_Azi	Radar azimuth angle at cell's initial detection step (degree).
Wind_950	Reanalysis wind magnitude at 950 mb (m/s).
Wind_950_Azi	Reanalysis wind azimuth angle at 950 mb (degree).
H_ZDR_sl_ini	Top altitude of first Z <sub>DR</sub> shaft (km).
ZDR_sl_time	Z <sub>DR</sub> t <sub>0</sub> used in the Z <sub>DR</sub> super positioning method.
life_time_minutes	Cell life time in minutes
H_max	Maximum Echo top height in km.
NCLmax	Cluster complexity (N of cells within the cluster that the cell resides in.
cell_ini_station_dis_km	Distance between the radar to the location of cell initial detection (km).
simple_life_nt	Number of time steps while cell stays isolated.
Hour	The time (UTC) cell first get detected.
cell_ini_time_step	Cell first detection time step relative to its case total time steps.
cell_time	Cell initial detection time (yyyymmddhhmmss).
lon_ini	Cell initial detection longitude.
lat_ini	Cell initial detection latitude.
azimuth_ini	Azimuth angle at cell initial detection time (degree).

cape	Convective available potential energy at cell's initial detection time (J/kg).
melt_m	Melting level at cell's initial detection time (m).
tb	Temperature at cloud base at cell's initial detection time (K).
lma_N_sum	Integrated LMA source count.
lma_sum	Integrated LMA power (PBW).
CCN_Mean_cm	Mean CCN number concentration at cell initial detection time ( $\text{cm}^{-3}$ ).
SS_Mean	Mean super saturation at cell initial detection time (%).
SAT_CBH_Mean	Mean cloud base height at cell initial detection time (m).
z_1	First echo height of cell (km).

## References

- Air Resources Laboratory, cited 2018: Eta Data Assimilation System (EDAS40) Archive Information
- Bellon, A., S. Lovejoy, and G. L. Austin, 1980: Combining Satellite and Radar Data for the Short-Range Forecasting of Precipitation. *Monthly Weather Review*, **108**, 1554-1566.
- Biagi, C. J., K. L. Cummins, K. E. Kehoe, and E. P. Krider, 2007: National Lightning Detection Network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003–2004. *Journal of Geophysical Research: Atmospheres*, **112**.
- Bjerkaas, C. L., and D. E. Forsyth, 1979: Real-Time Automated Tracking of Severe Thunderstorms Using Doppler Weather Radar. *Bulletin of the American Meteorological Society*, **60**, 533-533.
- Browning, K. A., 1977: The Structure and Mechanisms of Hailstorms. *Hail: A Review of Hail Science and Hail Suppression*, G. B. Foote, and C. A. Knight, Eds., American Meteorological Society, 1-47.
- Chmielewski, V. C., and E. C. Bruning, 2016: Lightning Mapping Array flash detection performance with variable receiver thresholds. *Journal of Geophysical Research: Atmospheres*, **121**, 8600-8614.
- Crane, R. K., 1979: Automatic Cell Detection and Tracking. *Ieee T Geosci Remote*, **17**, 250-262.
- del Moral, A., T. Rigo, and M. C. Llasat, 2018: A radar-based centroid tracking algorithm for severe weather surveillance: identifying split/merge processes in convective systems. *Atmospheric Research*, **213**, 110-120.
- Dixon, M., and G. Wiener, 1993: Titan - Thunderstorm Identification, Tracking, Analysis, and Nowcasting - a Radar-Based Methodology. *J Atmos Ocean Tech*, **10**, 785-797.
- Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. *Atmospheric Research*, **93**, 575-586.

- Gascón, E., A. Merino, J. L. Sánchez, S. Fernández-González, E. García-Ortega, L. López, and L. Hermida, 2015: Spatial distribution of thermodynamic conditions of severe storms in southwestern Europe. *Atmospheric Research*, **164-165**, 194-209.
- Greene, D. R., and R. A. Clark, 1972a: Vertically Integrated Liquid Water—A New Analysis Tool. *Monthly Weather Review*, **100**, 548-552.
- Greene, D. R., and R. A. Clark, 1972b: Vertically Integrated Liquid Water - New Analysis Tool. *Monthly Weather Review*, **100**, 548-&.
- Handwerker, J., 2002: Cell tracking with TRACE3D - a new algorithm. *Atmospheric Research*, **61**, 15-34.
- Joe, P., D. Burgess, R. Potts, T. Keenan, G. Stumpf, and A. Treloar, 2004: The S2K severe weather detection algorithms and their performance. *Weather Forecast*, **19**, 43-63.
- Johnson, J. T., P. L. MacKeen, A. Witt, E. D. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, 1998: The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Weather Forecast*, **13**, 263-276.
- Kumjian, M. R., and A. V. Ryzhkov, 2012: The Impact of Size Sorting on the Polarimetric Radar Variables. *J Atmos Sci*, **69**, 2042-2060.
- Kyznarova, H., and P. Novak, 2009: CELLTRACK - Convective cell tracking algorithm and its use for deriving life cycle characteristics. *Atmospheric Research*, **93**, 317-327.
- Lakshmanan, V., T. Smith, G. Stumpf, and K. Hondl, 2007: The Warning Decision Support System-Integrated Information. *Weather Forecast*, **22**, 596-612.
- Li, L., W. Schmid, and J. Joss, 1995: Nowcasting of Motion and Growth of Precipitation with Radar over a Complex Orography. *J Appl Meteorol*, **34**, 1286-1300.
- Lopez, L., and J. L. Sanchez, 2009: Discriminant methods for radar detection of hail. *Atmospheric Research*, **93**, 358-368.
- Melcón, P., A. Merino, J. L. Sánchez, L. López, and E. García-Ortega, 2017: Spatial patterns of thermodynamic conditions of hailstorms in southwestern France. *Atmospheric Research*, **189**, 111-126.
- Merino, A., J. L. Sánchez, S. Fernández-González, E. García-Ortega, J. L. Marcos, C. Berthet, and J. Dessens, 2019: Hailfalls in southwest Europe: EOF analysis for identifying synoptic pattern and their trends. *Atmospheric Research*, **215**, 42-56.
- Meyer, F., 1994: Topographic Distance and Watershed Lines. *Signal Process*, **38**, 113-125.
- Michaelides, S., and Coauthors, 2018: Reviews and perspectives of high impact atmospheric processes in the Mediterranean. *Atmospheric Research*, **208**, 4-44.
- Moller, A. R., C. A. D. III, M. P. Foster, and G. R. Woodall, 1994: The Operational Recognition of Supercell Thunderstorm Environments and Storm Structures. *Weather and Forecasting*, **9**, 327-347.
- Munoz, C., L. P. Wang, and P. Willems, 2018: Enhanced object-based tracking algorithm for convective rain storms and cells. *Atmospheric Research*, **201**, 144-158.
- Orlanski, I., 1975: Rational Subdivision of Scales for Atmospheric Processes. *Bulletin of the American Meteorological Society*, **56**, 527-530.
- Ortega, K. L., J. M. Krause, and A. V. Ryzhkov, 2016: Polarimetric Radar Characteristics of Melting Hail. Part III: Validation of the Algorithm for Hail Size Discrimination. *J Appl Meteorol Clim*, **55**, 829-848.
- Park, H., A. V. Ryzhkov, D. S. Zrnica, and K. E. Kim, 2009: The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS. *Weather Forecast*, **24**, 730-748.

- Putsay, M., I. Szenyan, and A. Simon, 2009: Case study of Mesoscale Convective Systems over Hungary on 29 June 2006 with satellite, radar and lightning data. *Atmospheric Research*, **93**, 82-92.
- Rinehart, R. E., and E. T. Garvey, 1978: 3-Dimensional Storm Motion Detection by Conventional Weather Radar. *Nature*, **273**, 287-289.
- Rosenfeld, D., 1987: Objective Method for Analysis and Tracking of Convective Cells as Seen by Radar. *J Atmos Ocean Tech*, **4**, 422-434.
- Rosenfeld, D., and Coauthors, 2016: Satellite retrieval of cloud condensation nuclei concentrations by using clouds as CCN chambers. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 5828-5834.
- Ryzhkov, A., M. Diederich, P. F. Zhang, and C. Simmer, 2014: Potential Utilization of Specific Attenuation for Rainfall Estimation, Mitigation of Partial Beam Blockage, and Radar Networking. *J Atmos Ocean Tech*, **31**, 599-619.
- Ryzhkov, A. V., M. R. Kumjian, S. M. Ganson, and P. F. Zhang, 2013: Polarimetric Radar Characteristics of Melting Hail. Part II: Practical Implications. *J Appl Meteorol Clim*, **52**, 2871-2886.
- Sachidananda, M., and D. S. Zrnić, 1987: Rain Rate Estimates from Differential Polarization Measurements. *J Atmos Ocean Tech*, **4**, 588-598.
- Sanchez, J. L., A. Merino, P. Melcón, E. García-Ortega, S. Fernández-González, C. Berthet, and J. Dessens, 2017: Are meteorological conditions favoring hail precipitation change in Southern Europe? Analysis of the period 1948–2015. *Atmospheric Research*, **198**, 1-10.
- Schumacher, C., and R. A. Houze, 2003: Stratiform rain in the tropics as seen by the TRMM precipitation radar. *J Climate*, **16**, 1739-1756.
- Tabary, P., A. A. Boumahmoud, H. Andrieu, R. J. Thompson, A. J. Illingworth, E. Le Bouar, and J. Testud, 2011: Evaluation of two "integrated" polarimetric Quantitative Precipitation Estimation (QPE) algorithms at C-band. *J Hydrol*, **405**, 248-260.
- Tuttle, J. D., and G. B. Foote, 1990: Determination of the Boundary-Layer Air-Flow from a Single Doppler Radar. *J Atmos Ocean Tech*, **7**, 218-232.
- Williams, E. R., 1985: Large-Scale Charge Separation in Thunderclouds. *Journal of Geophysical Research-Atmospheres*, **90**, 6013-6025.
- Wilson, J. W., and Coauthors, 2004: Sydney 2000 Forecast Demonstration Project: Convective storm nowcasting. *Weather Forecast*, **19**, 131-150.
- Yoshida, S., T. Morimoto, T. Ushio, and Z. Kawasaki, 2009: A fifth-power relationship for lightning activity from Tropical Rainfall Measuring Mission satellite observations. *Journal of Geophysical Research-Atmospheres*, **114**.
- Zan, B., Y. Yu, J. Li, G. Zhao, T. Zhang, and J. Ge, 2019: Solving the storm split-merge problem—A combined storm identification, tracking algorithm. *Atmospheric Research*, **218**, 335-346.

## LIST OF FIGURES

**Figure 1.** Flow chart of the MCIT algorithm.

**Figure 2.** Example of scattered clouds with six identification and tracking steps for the KHGX radar on June 8, 2013, is shown. The first two rows show the horizontal VIL maps. Units are dB. Different colors in the lowest two rows divide the first two rows of VIL maps into cells. Each black number is a unique tracking id number for each cell with time.

**Figure 3.** A squall line example with six identification (a-f) and tracking steps (g-l) for the KLBB radar on August 1, 2015 is shown here. The first two rows show the horizontal VIL maps. Units are dB. Different colors in the last two rows divide the first two rows of VIL maps into cells. Each black number is a unique tracking id number for each cell with time.

**Figure 4.** The properties of an isolated thunderstorm (cell 590), a warm rain shower (cell 3401), a cell that belongs to a squall line (cell 2) and a super cell storm (cell 1790) as a function of time:  $Z_{bmax}$  is peak reflectivity at cloud base in dBZ;  $H_{max}$  is the cell's echo top height in  $10^{-1}$  km above sea level;  $A_{base}$  is the area at cloud base in  $10 \cdot \log_{10}(\text{km}^2)$ ;  $RVR$  is rain volume rate in  $10 \cdot \log_{10}(\text{m}^3 \text{ h}^{-1})$ ;  $LMA\_CC$  is the integrated LMA source count in  $10 \cdot \log_{10}(n)$ .

**Figure 5.** The time-height properties of cell 590 are shown here. Figure 5.a – h show the time-height evolution of  $Z$ ,  $Z_{DR}$ ,  $K_{DP}$ ,  $\rho_{hv}$ ,  $D_m$ ,  $N_w$ ,  $LMA\_CC$  and  $DHCA$ . For figure 5a – f, each grid box represents the magnitude of the variables at the same location of maximum  $Z$  of each time-height dimension for this cell. Figure 5g shows the integrated VHF lightning sources for this cell at each time-height dimension. Figure 5h shows hydrometeor classifications. Each panel title includes the specific UTC time of the cell shown in the figure when the cell is first detected. For instance, 201606111854 means 18:54, June 11, 2016 and 093 km/181° 32.82°N -101.83°W means this cell is first detected 93 km away from the KLBB radar with an azimuth angle of 181 degree located at 32.82°N -101.83°W.

**Figure 6.** As in Fig 5, but for cell 3401 on July 17, 2015.

**Figure 7.** As in Fig 5, but for cell 2 on June 8, 2016.

**Figure 8.** As in Fig 5, but for cell 1790 on June 11, 2016.



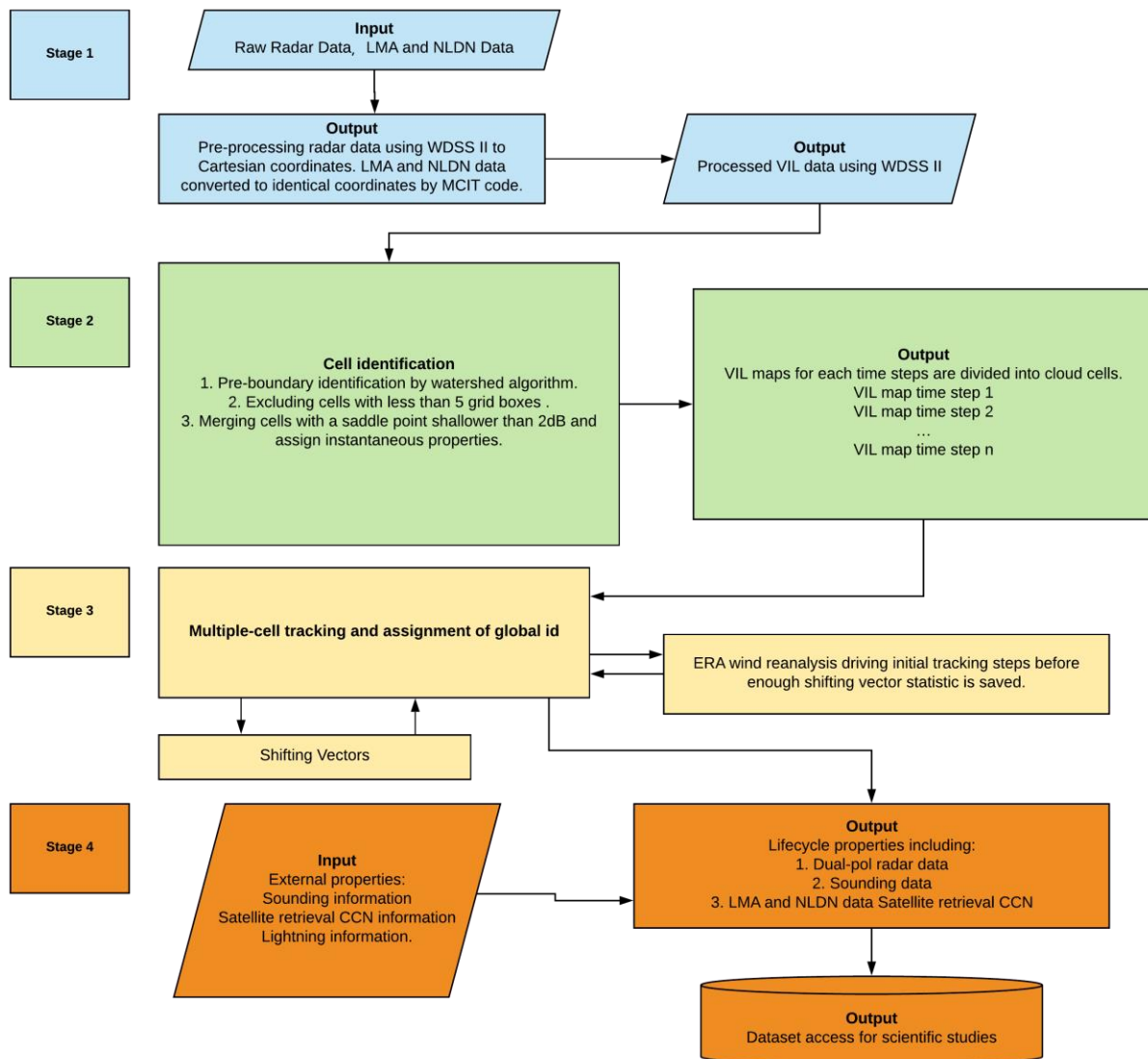


Figure 1 Flowchart of the MCIT algorithm.

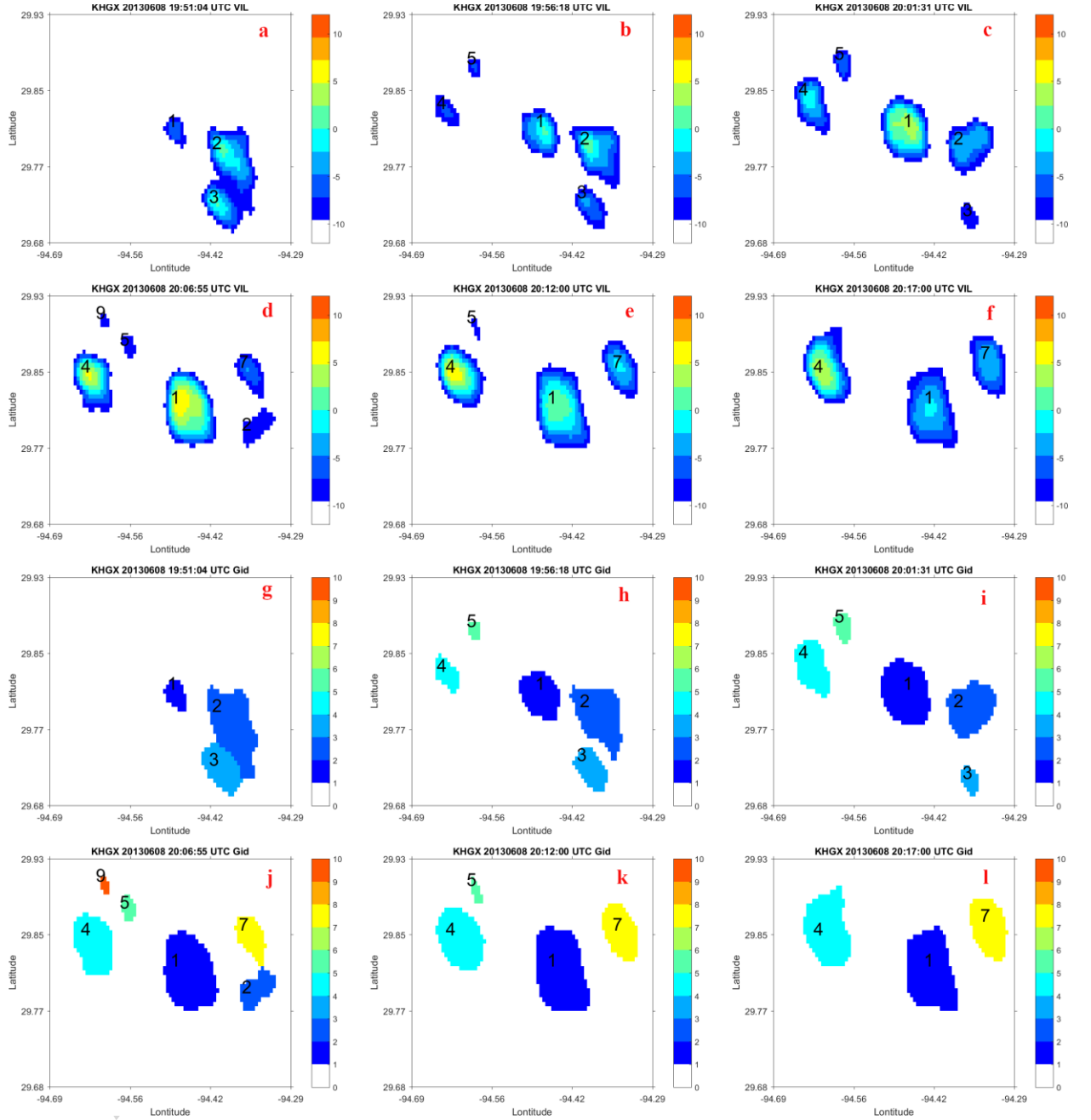


Figure 2. An example of identification and tracking of isolated clouds in a sequence of six radar scans from the KHXG radar on June 8, 2013. The first two rows (a-f) show the horizontal VIL maps. Units are dB. Panels of g – l shows the identity numbers same cells as in a – f. Each cell is marked by a unique color representing the identify number.

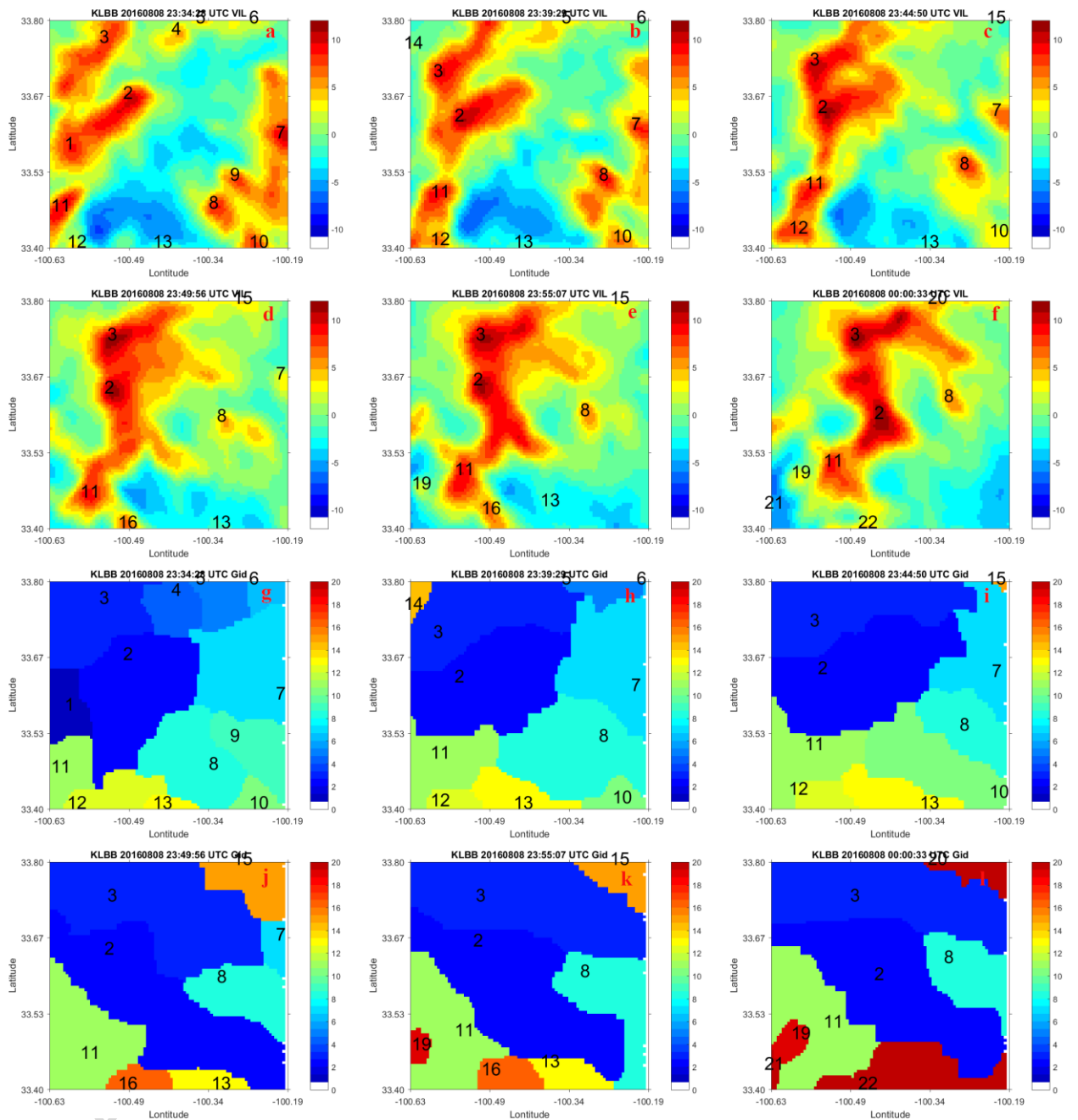


Figure 3. A squall line example with six identification (a-f) and tracking steps (g-l) for the KLBB radar on August 1, 2015 are shown here. The first two rows show the horizontal VIL maps. Units are dB. Different colors in the last two rows divide the first two rows of VIL maps into cells. Each black number is a unique tracking id number for each cell with time.

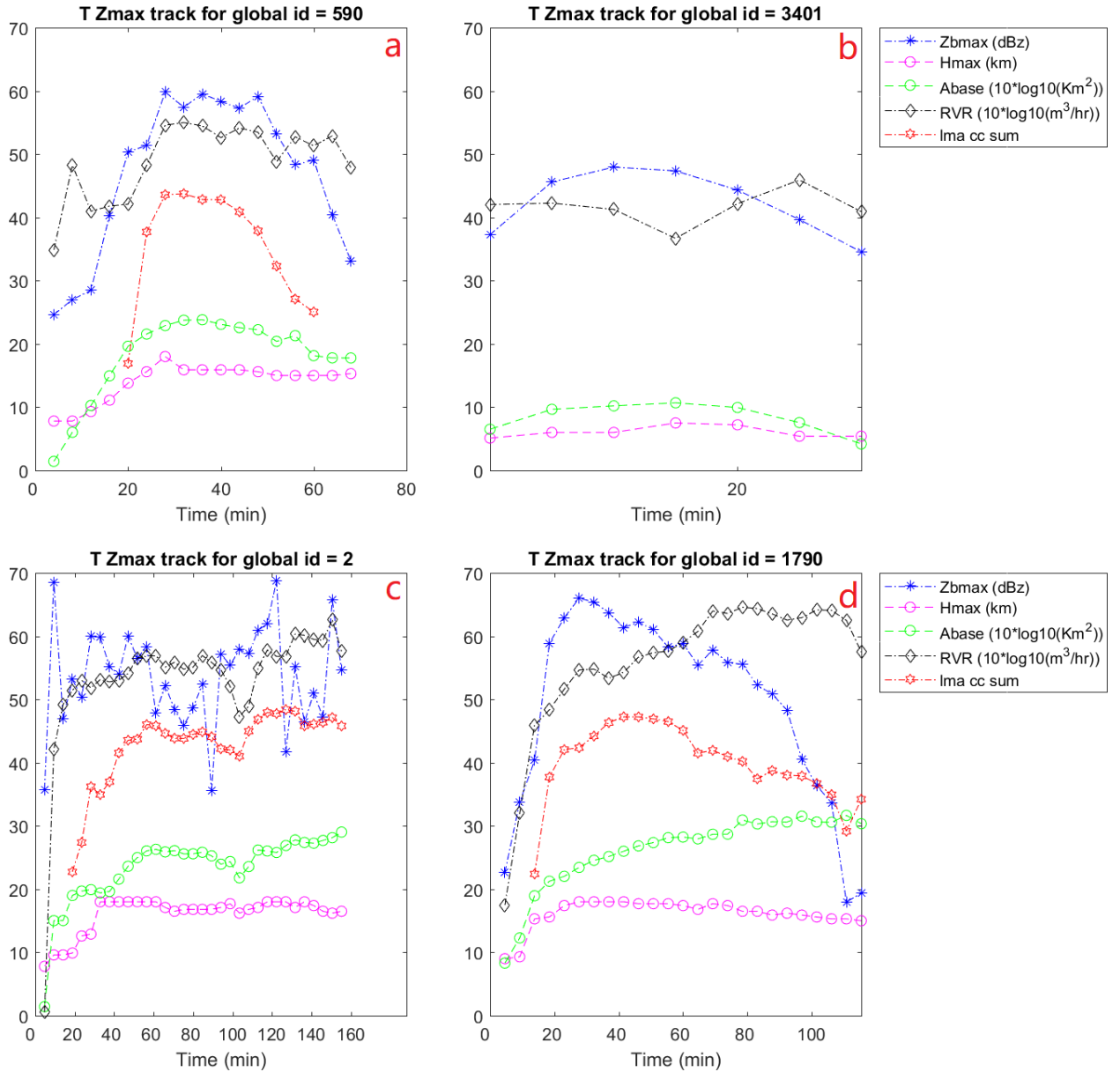


Figure 4. The properties of an isolated thunderstorm (cell 590), a warm rain shower (cell 3401), a cell that belongs to a squall line (cell 2) and a super cell storm (cell 1790) as a function of time: Zbmax is peak reflectivity at cloud base in dBZ; Hmax is the cell's echo top height in 10<sup>-1</sup> km above sea level; Abase is the area at cloud base in 10\*log10(km<sup>2</sup>); RVR is rain volume rate in 10\*log10(m<sup>3</sup> h<sup>-1</sup>); LMA\_CC is the integrated LMA source count in 10\*log10(n).



Figure 5. The time-height properties of an isolated thunderstorm (cell 590) are shown here. Figure 5a – h show the time-height evolution of  $Z$ ,  $Z_{DR}$ ,  $K_{DP}$ ,  $\rho_{hv}$ ,  $D_m$ ,  $N_w$ , LMA\_CC and DHCA. For figure 5a – f, each grid box represents the magnitude of the variables at the same location of maximum  $Z$  of each time-height dimension for this cell. Figure 5g shows the integrated VHF lightning sources for this cell at each time-height dimension. Figure 5h shows hydrometeor classifications. Each panel title includes the specific UTC time of the cell shown in the figure when the cell is first detected. For instance, 201606111854 means 18:54, June 11, 2016 and 093 km/181° 32.82°N -101.83°W means this cell is first detected 93 km away from KLBB site with an azimuth angle of 181 degree located at 32.82°N -101.83°W.



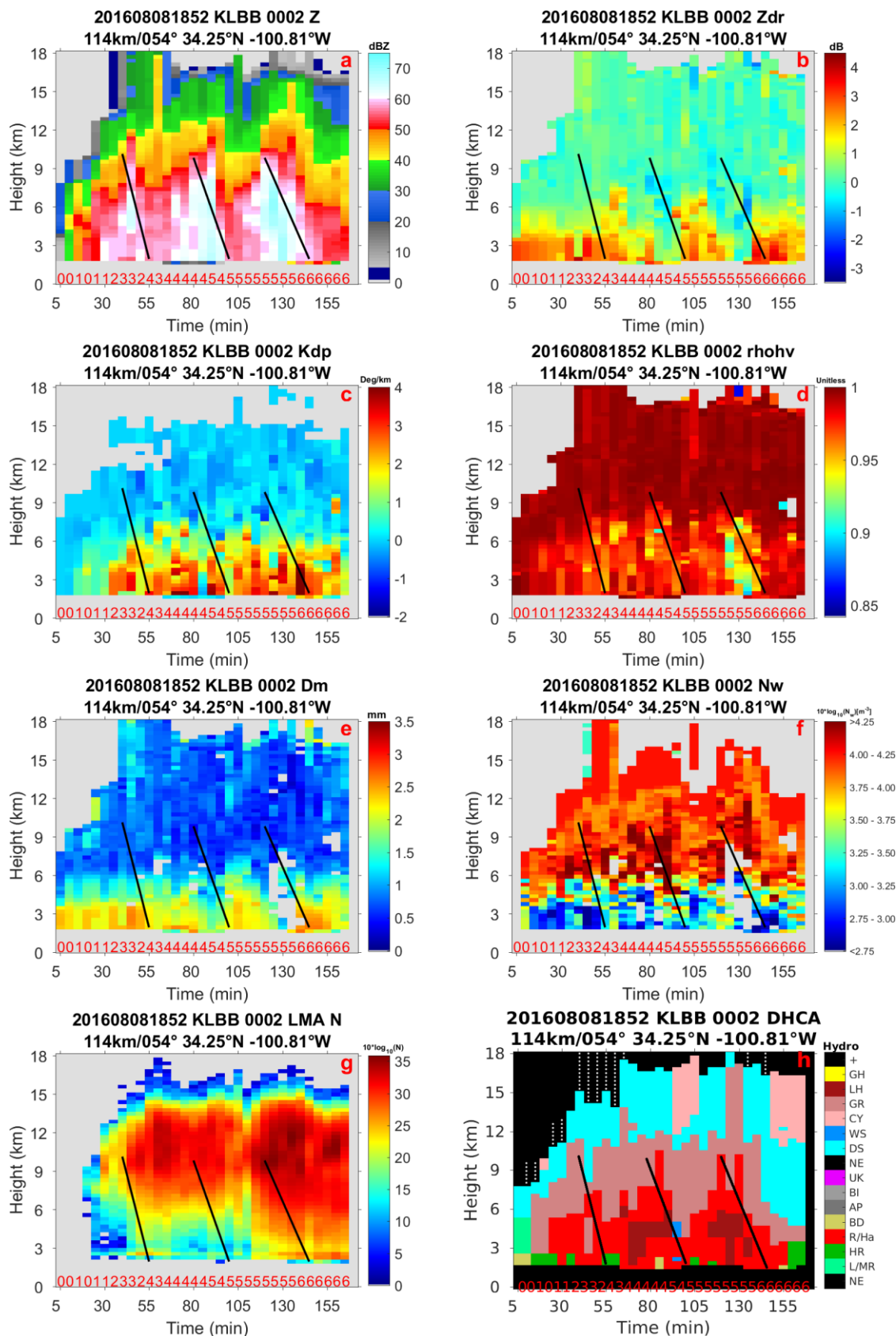


Figure 7. As in Fig 5, but for a multi-cell squall line (cell 2 on June 8, 2016).



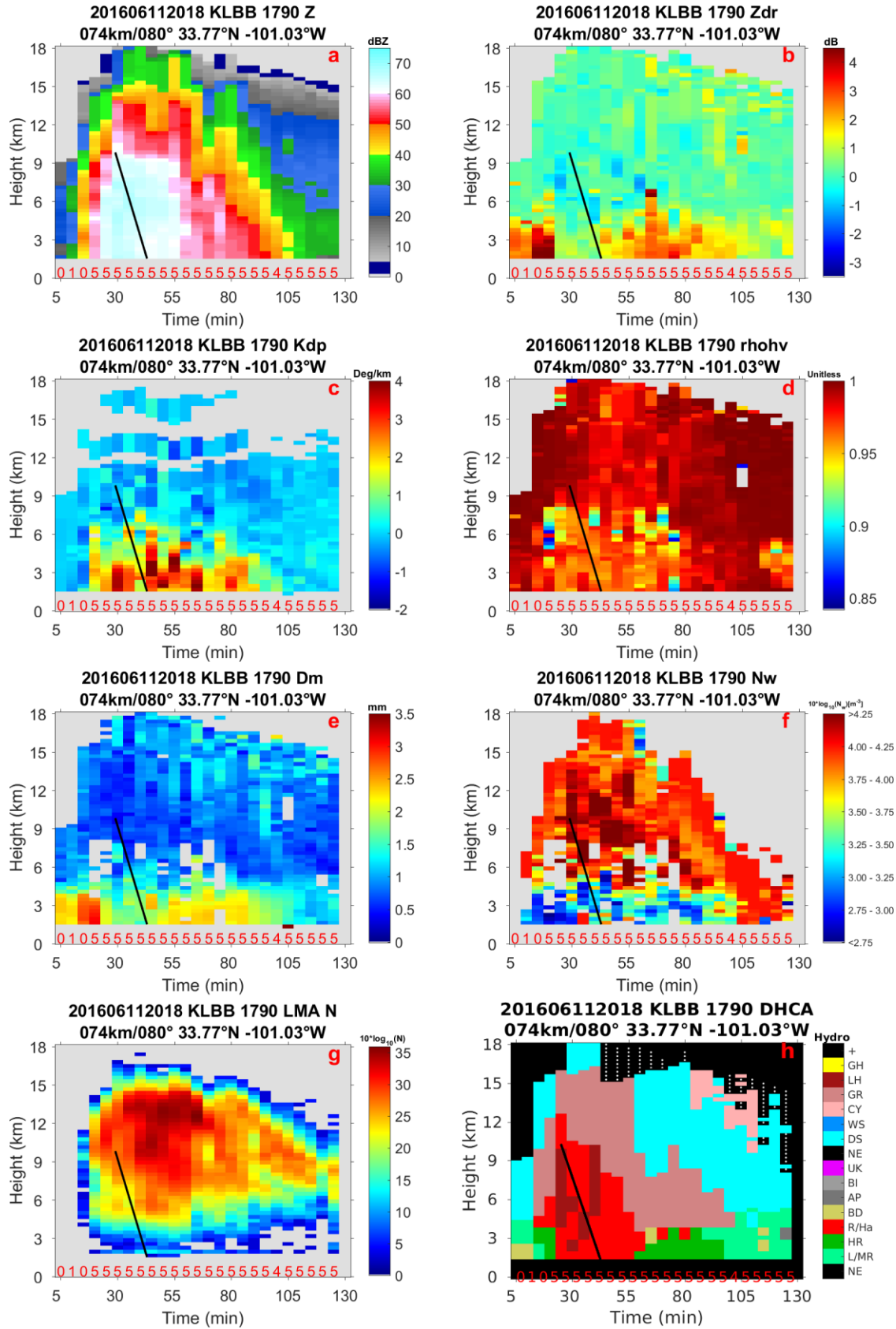


Figure 8. As in Fig 5, but for a super-cell (cell 1790 on June 11, 2016).

Supplementary

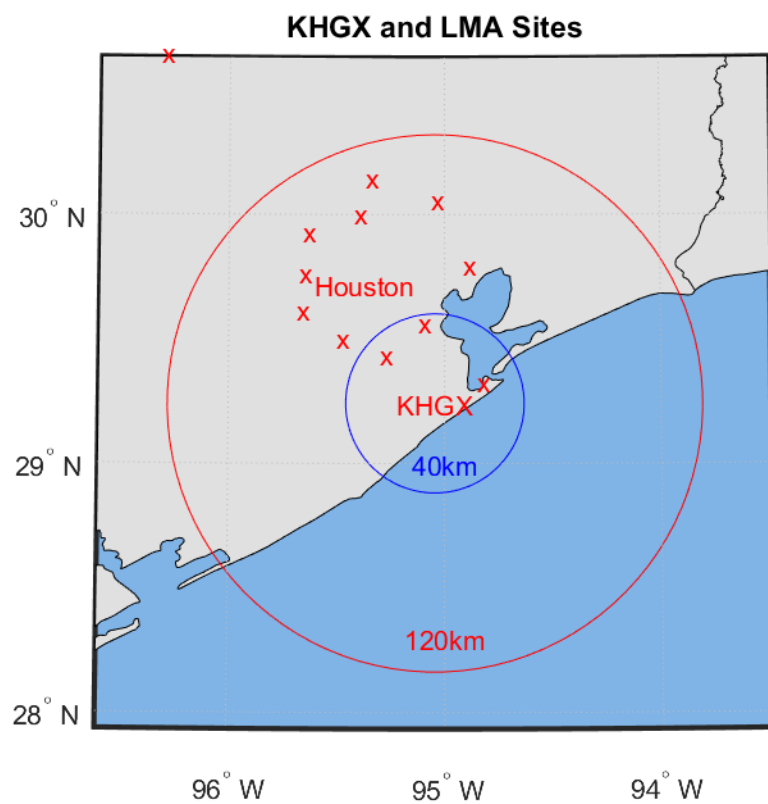


Figure S1. KHGX tracking domain is shown. Blue and red circles represent 40 km and 120 km search radius used in this study and are labelled correspondingly. The red crosses represent HLMA sensor locations.

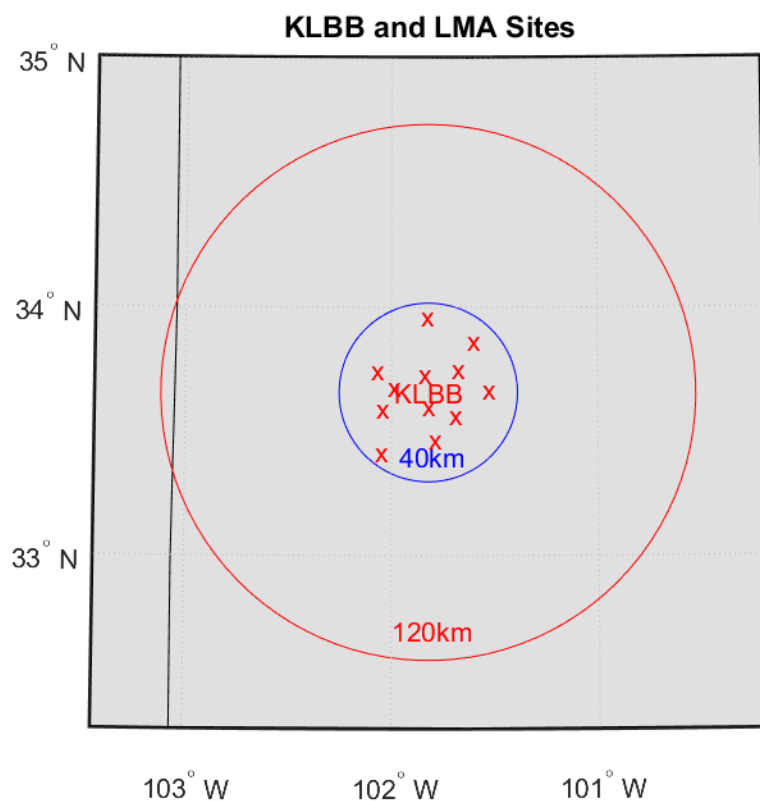


Figure S2. KLBB tracking domain is shown. Blue and red circles represent 40 km and 120 km search radius used in this study and are labelled correspondingly. The red crosses represent WLMA sensor locations.

### Highlights

- Development of a novel cloud cell tracking algorithm.
- Tracking cloud types include: convective and stratiform, isolated and clustered.
- Tracking cloud cells through maturation into stratiform storm elements.
- Use a synergy of polarimetric radar, satellite, lightning detection and model data.
- Algorithm has flexible output options and possibilities for further studies.

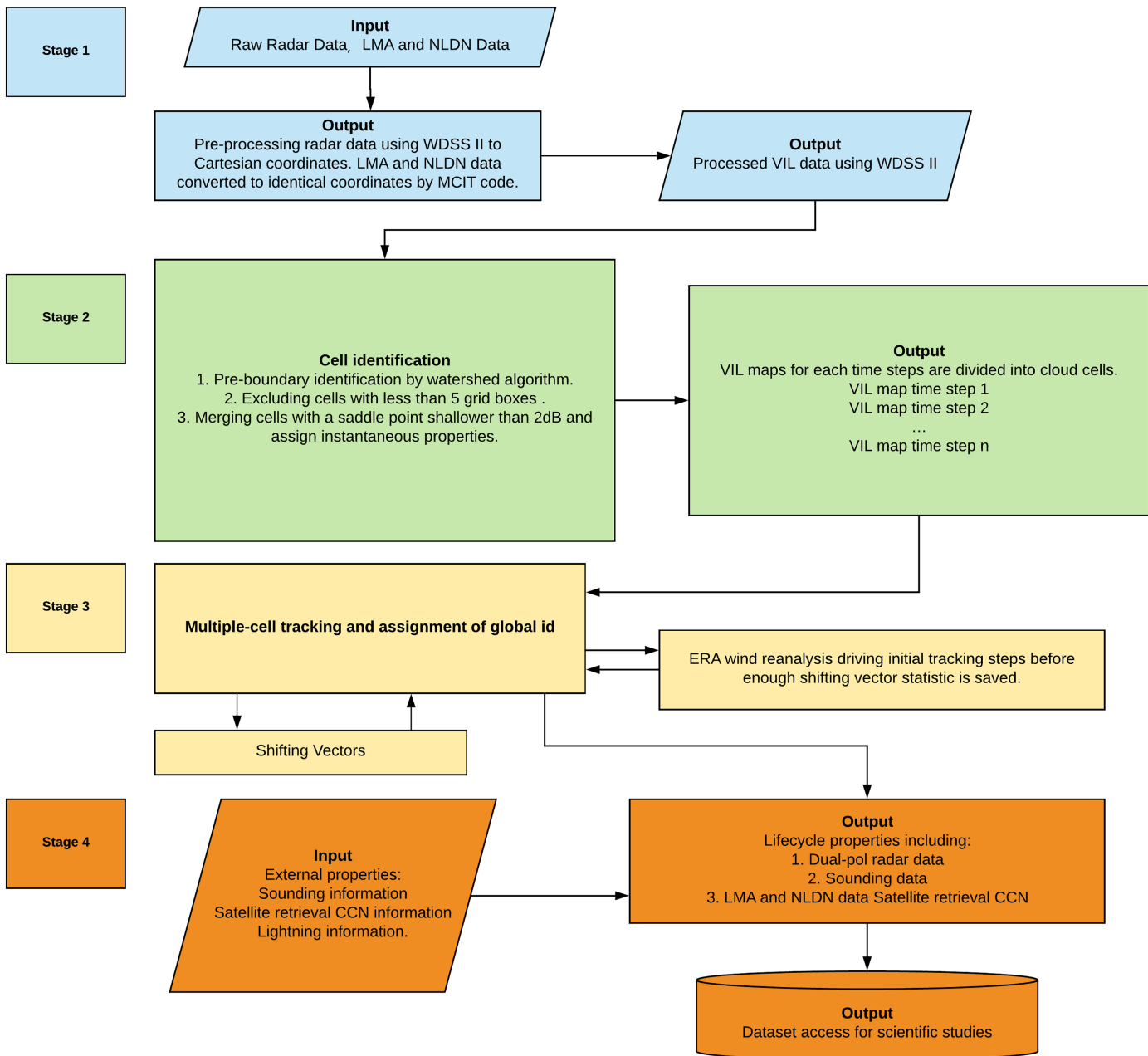


Figure 1

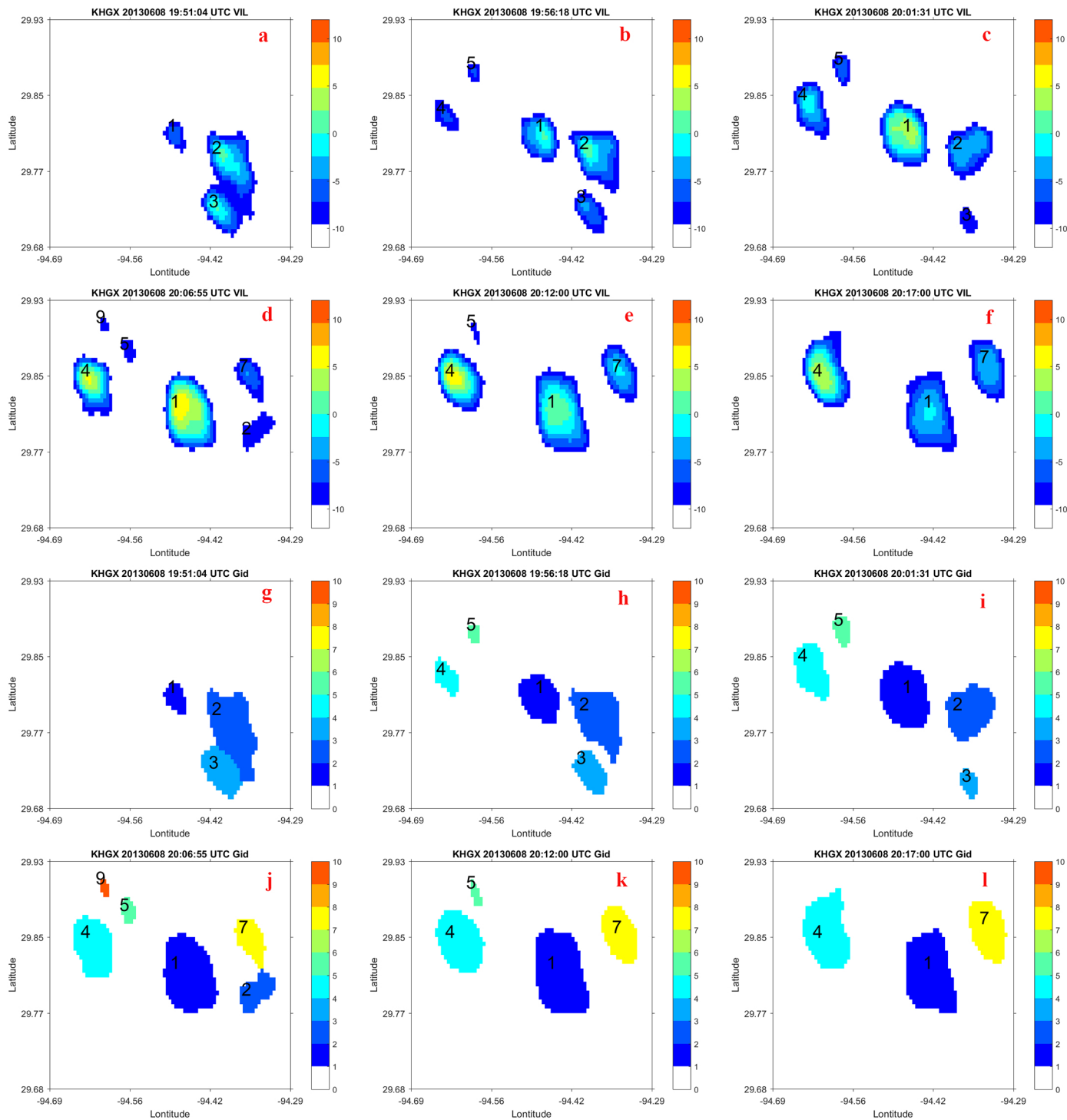


Figure 2

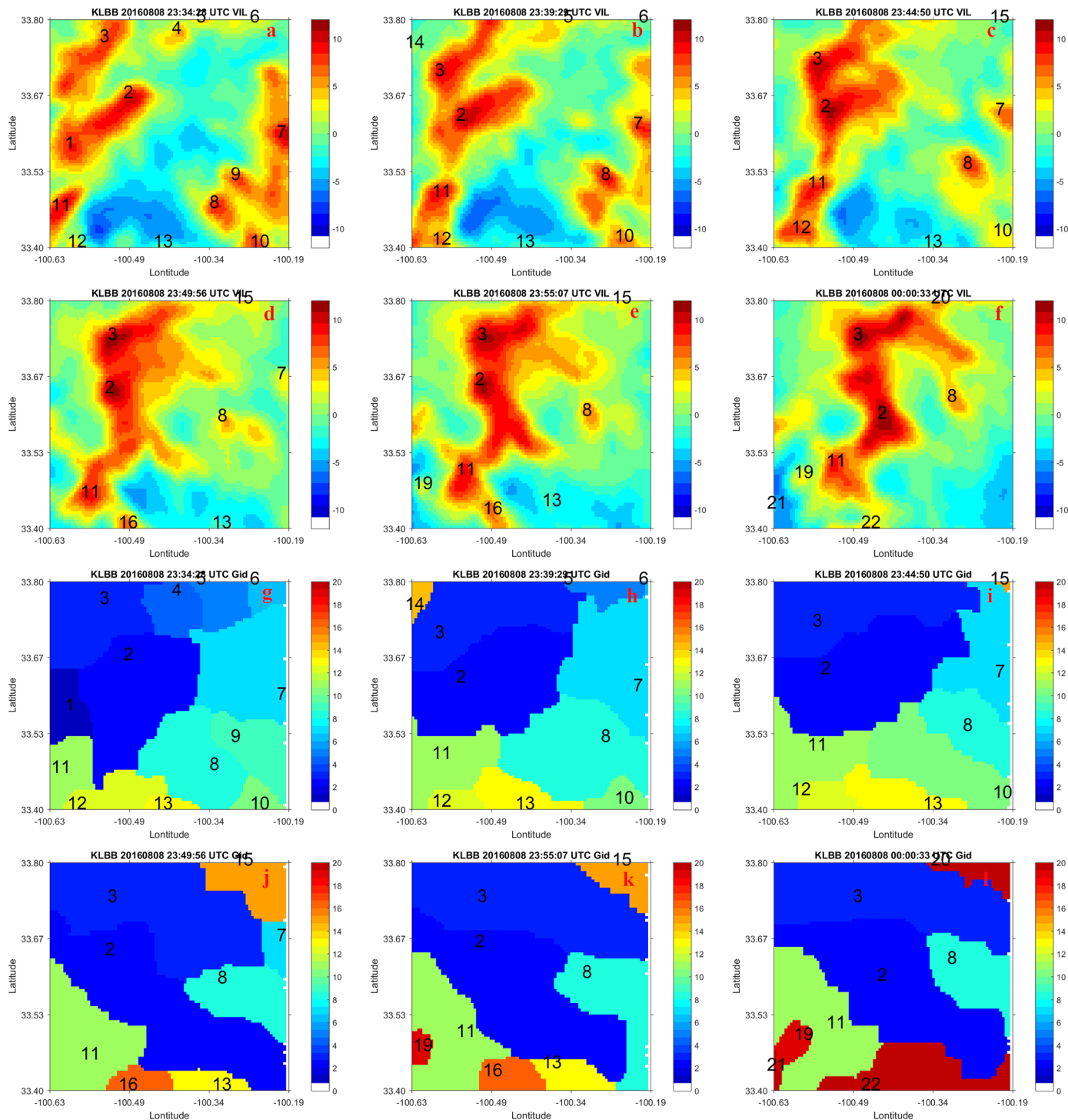


Figure 3

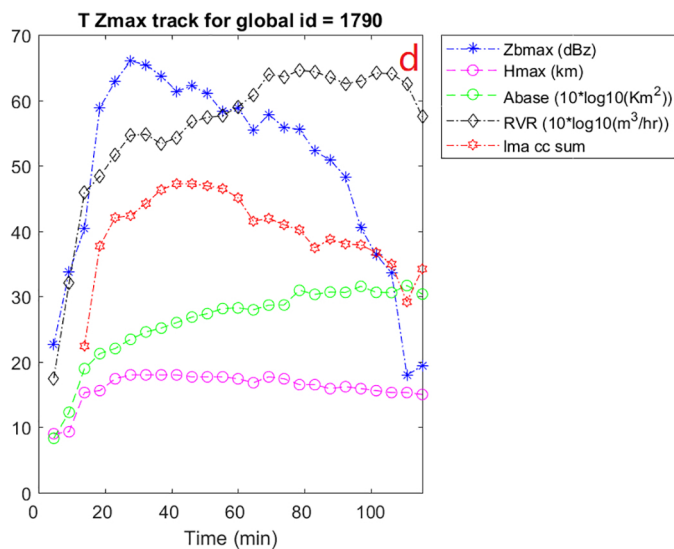
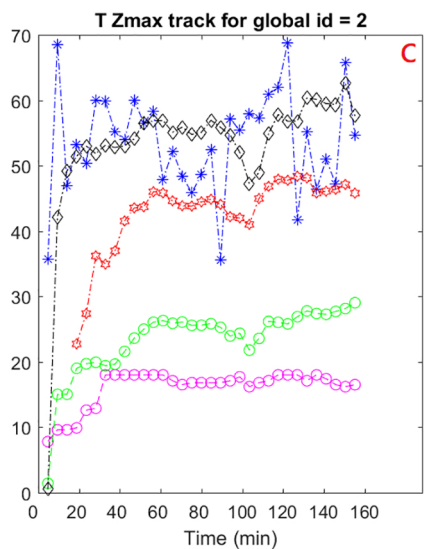
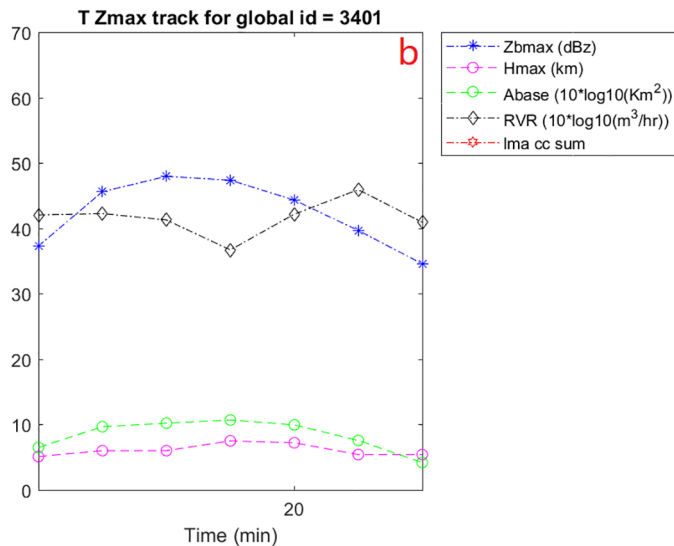
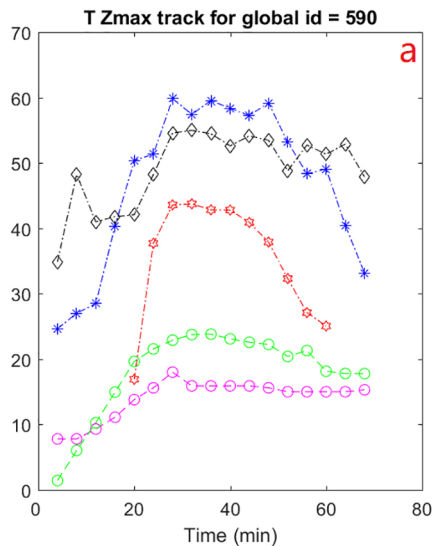


Figure 4





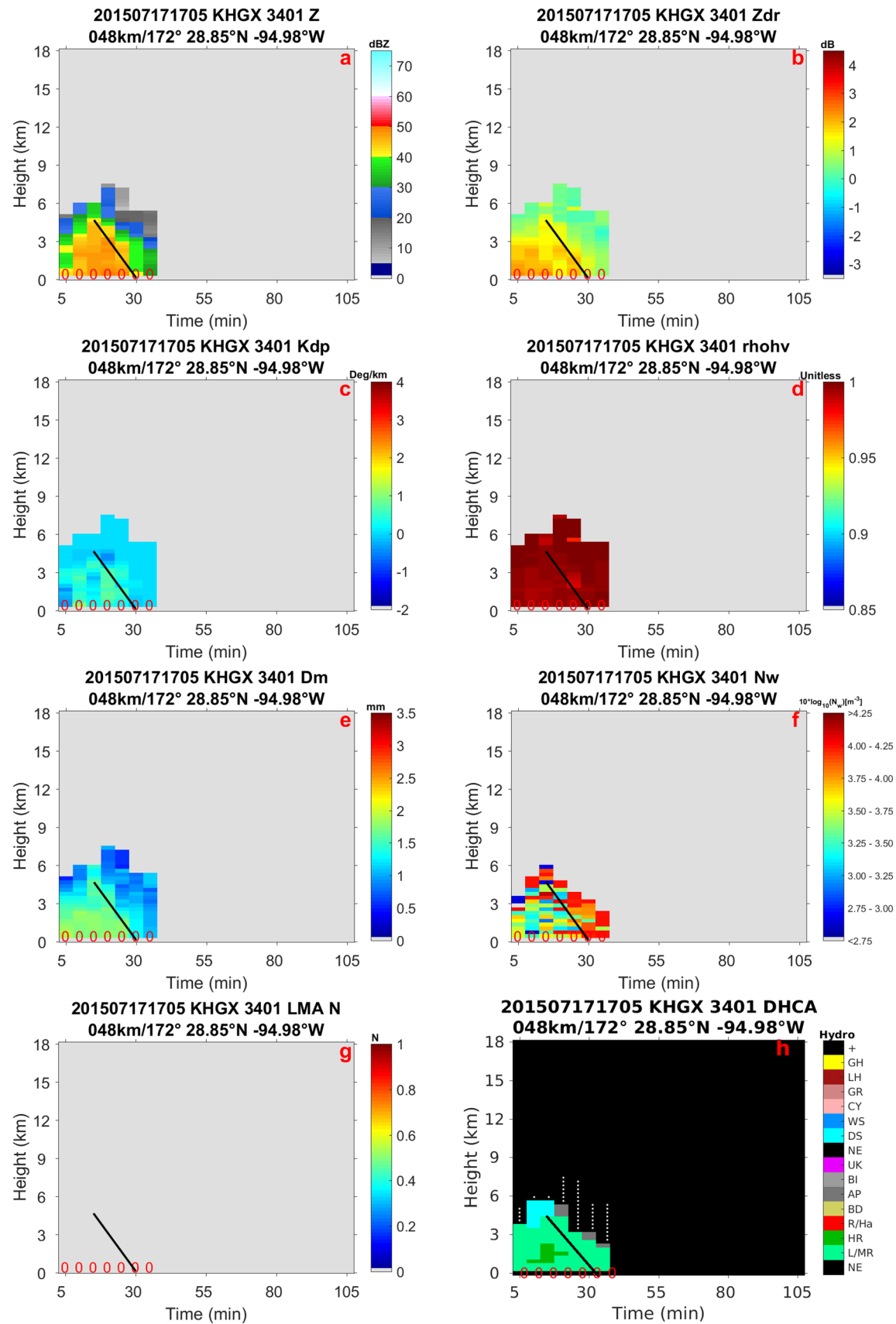


Figure 6

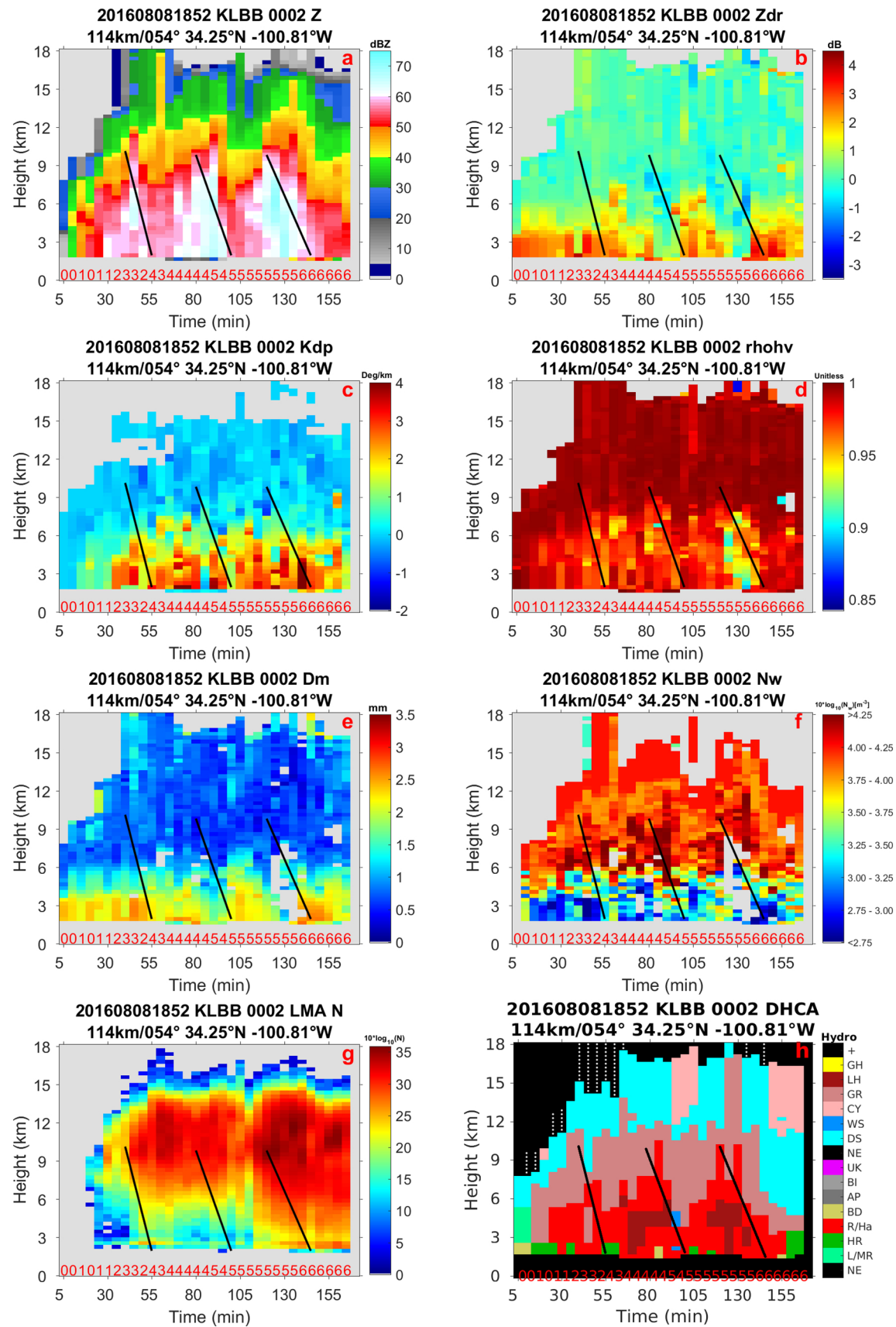


Figure 7

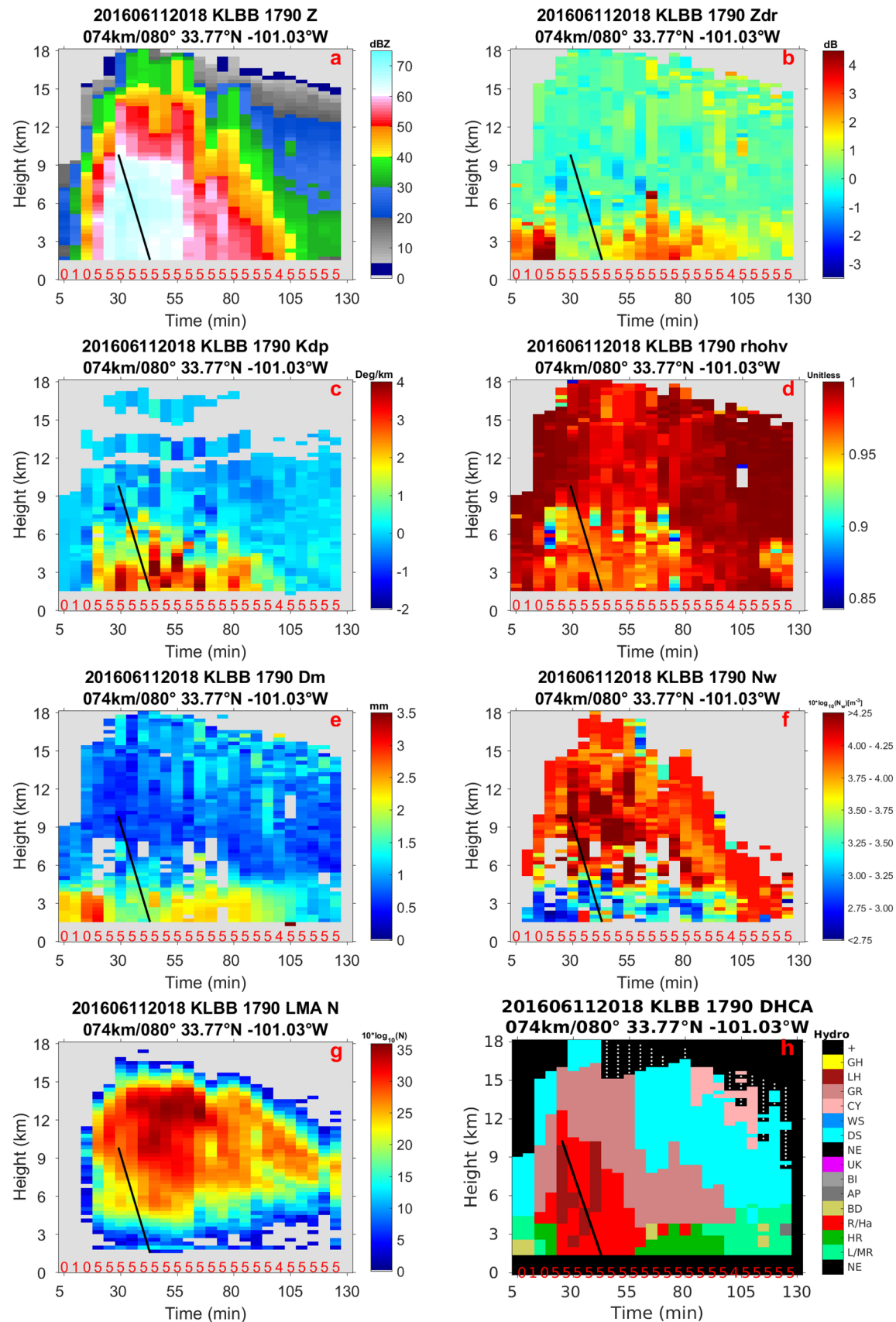


Figure 8