Identification, Representation, and Analysis of Convective Storms

By

Weibo Liu

Submitted to the graduate degree program in Geography and Atmospheric Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson Dr. Xingong Li

_____

Dr. Stephen L. Egbert

_____

Dr. David A. Rahn

_____

Dr. Johannes J. Feddema

_____

Dr. James R. Miller

Date Defended: July 15, 2016

The Dissertation Committee for Weibo Liu

certifies that this is the approved version of the following dissertation:

Identification, Representation, and Analysis of Convective Storms

_____

Chairperson Dr. Xingong Li

Date approved: September 2, 2016

Abstract

Large amount of time series of spatial snapshot data have been collected or generated for the monitoring and modeling of environmental systems. Those data provide an opportunity to study the movement and dynamics of natural phenomena. While the snapshot organization is conceptually simple and straightforward, it does not directly capture or represent the dynamic characteristics of the phenomena. This study presents computational methods to identify dynamic events from time series of spatial snapshots. Events are represented as directed spatiotemporal graphs to characterize their initiation, development, movement, and cessation. Graph-based algorithms are then used to analyze the dynamics of the events.

The method is demonstrated using the time series radar reflectivity images during one of the deadliest storm outbreaks that impacted 15 states of southeastern U.S. between April 23 and 29, 2011. As shown in this case study, convective storm events identified using our methods are consistent with previous studies and our analysis indicates that the left split/merger occurs more than right split/merger in those convective storm events, which confirms theory, numerical simulations, and other observed case studies.

This study also examines the spatial and temporal characteristics of thunderstorm life cycles in central United States mainly covering Kansas, Oklahoma, and northern Texas during the warm seasons from 2010 to 2014. Radar reflectivity and cloud-to-ground lightning data were used to identify thunderstorms. The thunderstorms were stored in a GIS database with a number of additional thunderstorm attributes. The spatial and temporal characteristics of thunderstorm occurrence, duration, initiation time, termination time, movement speed, and direction were analyzed. Results revealed that thunderstorms were most frequent in the eastern part of the study

area, especially at the borders among Kansas, Missouri, Oklahoma, and Arkansas. We also linked thunderstorm features to land cover types and compared thunderstorm characteristics between urban and surrounding rural areas. Our results indicated that thunderstorms favor forests and urban areas. This research demonstrates that advanced GIS representations and analyses for spatiotemporal events provide insights in thunderstorm climatology study.

Acknowledgements

There are a great many people who helped me during my Ph. D. study at the University of Kansas, and I would like to show my great gratitude to them.

I would like to express my sincere appreciation to my advisor, Dr. Xingong Li, who provided guidance, support, and encouragement to my research, career, and life. He always shows his patience and understanding even when I was depressed and did not do well in my study. My thanks are also extended to other committee members: Dr. Johannes Feddema, Dr. David Rahn, Dr. Stephen Egbert, and Dr. James Miller. Dr. Feddema offered me 3-year graduate research assistantship, which made my Ph. D. study possible in the United States. Dr. Feddema also provided me many good ideas and guidance in my dissertation research. Part of my dissertation research is to apply GIS methodologies and analyses in atmospheric science. Dr. Rahn helped me so much to learn domain knowledge and contribute many great ideas. Dr. Egbert taught me two classes, and offered me great support for my job search and dissertation. Dr. Miller also helped a lot on my proposal and the final version of my dissertation. I also thank Tyler Trigg who manually did the classification for single storm cells, then we could use the results for the future research.

I also extend my appreciation to Dr. David Braaten and Dr. John Paden who offered support, guidance, and help when I worked as a GRA at the Center for Remote Sensing of Ice Sheets (CReSIS).

I also thank other faculty members and staff at the Department of Geography and Atmospheric Science and CReSIS for offering me great supports and help.

Last but not least, I give my heartfelt thanks to my family and friends in China and the United States, especially to my dear parents for their selfless love and supports. Many thanks also go to my good friends, Yinling Zhang, Aaron Koop, Ashley Zung, and Chao Lan, who helped me so much in my study and life.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Problem statement

Analysis of distributions, spatial relationships, and other characteristics, such as shape and pattern from maps and charts, allows geographers to infer information about the processes that shape geographical phenomena. Recent technological improvements provide data rich environments for investigating various geographical phenomena. Especially, increasing availability of spatiotemporal data collected from satellite imagery and other remote sensors offers opportunities for advanced analysis of dynamic geographical phenomena (McIntosh and Yuan 2005a).

The objects or fields extracted from different sources of images are static. However, in many application domains, there is a growing body of work showing that studying the dynamic aspects of geographical phenomena such as infectious diseases, precipitation, hurricanes and wildfires is also essential and useful for explanatory and predictive models (Yuan 2001) that answer different geographical questions. One important characteristic of those time series of snapshot data is the constant change of variables in space and time, implying dynamic system behaviors. As a consequence, spatiotemporal fields and objects have added additional temporal attributes in object-oriented modeling. In order to fully represent dynamic geographical phenomena, the whole lifecycle of those geographical phenomena need be modeled including the initiation, development, movement, and decay. Nevertheless, traditional GIS data models have not yet

achieved this goal, and they can only represent the information at one slice in time for the area of interest. Many studies argue that the next real breakthrough in computer modeling of geographical phenomena comes when we move from an object-oriented view to an event-oriented view of the world (Peuquet and Duan 1995, Yuan 2001, Worboys 2005). Worboys (2005) also summarized the recent history of dynamic geographical data models into three stages: static GIS (stage zero), temporal snapshots (stage one), object change (stage two). He also pointed out that the future stage in this evolution is a complete treatment of change, in terms of events (stage three).

Nowadays, there are urgent needs to explore and understand how events evolve with 2- or 3D time series snapshots in many fields such as meteorology, hydrology, and transportation. Nevertheless, geographers are challenged on how to effectively identify events because of large data volume and limitations of conventional data models. Data organization and tools available in current GIS are based largely on the map metaphor and provide limited support for querying and exploring events (Goodchild et al. 2007). As a consequence, intensive human intervention is typically required when searching spatiotemporal datasets for events or processes. For many spatiotemporal data, the volume produced will quickly exceed the ability for analysts to manually explore all of the available data (MacEachren et al. 1999). Thus, there is a great need to develop automated processing methods to explore spatiotemporal data efficiently. McIntosh and Yuan (2005b) pointed out that the power and usefulness of GIS technology can be significantly enhanced by representing geographical events in GIS data models and providing functions to explore the characteristics of geographical events.

Recognizing the need for automatic methods to extract and represent dynamic geographical phenomena from large spatiotemporal datasets, this research is a continuing effort that develops

innovative and computationally efficient methods and tools for the identification, representation, and exploration of events within GIScience field. Although this dissertation chose convective storm events as a case study, the proposed methods and tools can be extended to other fields such as temperature, wildfire, and land use/land cover.

1.2 Research goals and objectives

In this research, the goal is to develop automatic approaches to facilitating the analysis of spatiotemporal datasets through the event perspective. We follow the idea that an event can be systematically delineated with an origin, a development stage, a movement stage and a potential senescence or dissolution phase. Convective storm events are chosen as an example of studying dynamic geographical phenomena. Like wildfire, precipitation varies within a rainstorm and is typically represented as a rain field, but individual rainstorms may be isolated as events that occur, evolve, and disappear in a space-time frame (Niemczynowicz 1987). My research objectives include:

a) Develop a general conceptual framework and computational methods to identify and represent events from time series of snapshot datasets and test the methods using convective storms as an example;
b) Analyze the spatial and temporal characteristics of thunderstorms in central United States, and explore the relationships between thunderstorm occurrences and land cover types.

The degree to which GIS can support spatiotemporal queries and analyses depends upon its embedded data models and representations. As discussed above, existing data models and representations in most GISs don't fully support dynamic geographical phenomena. The first research objective of this dissertation is to devise data models and representations which

incorporate the event concept and have the ability to represent both field and object characteristics of dynamic geographical phenomena in space and time. In this research, a directed spatiotemporal graph model is used to represent storm events, and graph-based algorithms are explored to further analyze the events.

The second objective is to use the storm events extracted from time series of images to answer some geographical questions. First, climatology or spatiotemporal characteristics of storm events are analyzed, which include inter-annual and intra-seasonal variations of event occurrence, duration, size, geographical location, and other characteristics such as split and merger, direction and magnitude of movement, and triggering and dissipation time. Previous studies (Pielke 2001, Matyas and Carleton 2010, Degu et al. 2011) have suggested that spatial heterogeneities in vegetation cover, water body, and other land use/land cover (LULC) alter the development of convective rainfall. The most dramatic anthropogenic LULC modification of natural environment is arguably urban development. As a consequence, in the second part, we test the hypothesis that urbans alter the occurrence of storms using the event database.

1.3 Organization of the dissertation

Chapter 2 and 3 are formatted as journal articles. Chapter 2 introduces an event tracking algorithm, sensitivity analysis of the tracking algorithm, directed spatiotemporal graph representation, graph-based algorithms for analyzing events, and a case study on convective storms. Chapter 3 covers the spatial and temporal characteristics of warm-season thunderstorm life cycles in the central United States from 2010 to 2014. It also examines the relationships between thunderstorm occurrences and land cover types. The proposed methods on assessing

similarities of convective storm events and preliminary experiments are presented in Chapter 4. A summary on this dissertation and future research are introduced in Chapter 5.

References

Degu, A.M., Hossain, F., Niyogi, D., Pielke, R., Shepherd, J.M., Voisin, N. and Chronis, T., 2011. The influence of large dams on surrounding climate and precipitation patterns. *Geophysical Research Letters*, 38 (4).

Goodchild, M.F., Yuan, M. and Cova, T.J., 2007. Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science*, 21 (3), 239-260.

MacEachren, A.M., Wachowicz, M., Edsall, R., Haug, D. and Masters, R., 1999. Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods. *International Journal of Geographical Information Science*, 13 (4), 311-334.

Matyas, C.J. and Carleton, A.M., 2010. Surface radar-derived convective rainfall associations with Midwest US land surface conditions in summer seasons 1999 and 2000. *Theoretical and applied climatology*, 99 (3-4), 315-330.

McIntosh, J. and Yuan, M., 2005a. A framework to enhance semantic flexibility for analysis of distributed phenomena. *International Journal of Geographical Information Science*, 19 (10), 999-1018.

McIntosh, J. and Yuan, M., 2005b. Assessing similarity of geographic processes and events. *Transactions in GIS*, 9 (2), 223-245.

Niemczynowicz, J., 1987. Storm tracking using rain gauge data. *Journal of Hydrology*, 93 (1), 135-152.

Peuquet, D. and Duan, N., 1995. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems*, 9 (1): 7-24.

Pielke, R.A., 2001. Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Reviews of Geophysics*, 39 (2), 151-177.

Worboys, M., 2005. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, 19 (1), 1-28.

Yuan, M., 2001. Representing complex geographic phenomena in GIS. *Cartography and Geographic Information Science*, 28 (2), 83-96.

2 Storm event representation and analysis based on a directed spatiotemporal graph model

2.1 Introduction

Environmental monitoring systems collect vast amounts of data and the time series of spatial snapshot data may be used to observe and investigate a diverse array of natural phenomena. The common characteristic of snapshot data is the constant change of variables in space and time that implies a dynamic system behavior. While organizing data as snapshots is conceptually simple and straightforward, it does not directly capture or represent the dynamic characteristics of geographic phenomena. Many scholars argue that the next real breakthrough in the modeling of geographic phenomena will come when we move from an object-oriented view to an event-oriented view (Peuquet and Duan 1995, Yuan 2001, Worboys 2005).

Many definitions on events exist in the literature. The general consensus is that events are associated with localized processes in space and time that change the attribute or state of an object or a field. Zacks and Taversky (2001) investigated the nature of events in human perception and conception and defined events as a segment of time at a given location that was conceived by an observer with a beginning and an end. In this context, when only the change of position is relevant, those objects are commonly referred to as moving objects. The trajectories of those objects can be visualized by their space-time paths (Shaw et al. 2008) and events can be identified by location-change (e.g., go-to-work and have-lunch). For this type of events, the existence and endurance of object identities are the key premise for event identification and

analysis. Similar to moving objects, naturally occurring phenomena such as convective storms are also dynamic entities with identifiable spatial and temporal variations within them. In contrast to moving objects that have pre-defined identities, the changing and clustering of attributes in space and time actually define the identities. At a more fundamental level, this type of event originates from the plenum view in the philosophy of science and particularly in modern physics where 'the spatiotemporal clusters of known attributes are the things' (Couclelis 1992). An event is defined here as an individual occurrence or episode that has a definite start and end.

There is a pressing need to explore and understand how events evolve with 2- or 3D time series of snapshots in many different fields. Nevertheless, geographers are challenged to effectively identify and depict events due to the large volume of data, the complexity of identifying events, and the limitations of conventional GIS data models. Data organization and analysis tools that are available in current GIS are largely based on the map metaphor and provide limited support for querying and exploring events. As a consequence, intensive human intervention is typically required when searching spatiotemporal datasets for specific events or processes. For many types of spatiotemporal data, the volume produced will quickly exceed the ability for analysts to manually explore all of the available data. Thus, there is a great need to develop automated processing methods and representation models to explore spatiotemporal data efficiently. As McIntosh and Yuan (2005) pointed out, the power and usefulness of GIS technology could be significantly enhanced by representing geographic events in GIS data models and providing functions to explore the characteristics of geographic events.

Langran and Chrisman (1988) are among the first who proposed the modeling concepts for temporal GIS. Galton (1995, 2000) used an instant-based model of time to describe the movement of events. Yuan and Hornsby (2007) summarized six types of spatiotemporal models

(time-stamped, change-based, event-based, movement-based, activity-based, and process-based) and emphasized that the temporal dimension should first be integrated into these models in order to capture the dynamic features of geographic phenomena. There have been many attempts to extend spatiotemporal GIS data models based on the event perspective. Peuquet and Duan (1995) proposed an event-based spatiotemporal data model (ESTDM) where an event was a change in state. The sequence of events through time, which represents the spatiotemporal manifestation of some processes, is noted via a time-line called an 'event list'. However, this model cannot directly reveal the relations between geographic entities such as a split, merger, combinatorial situation, or the filiation relations of geographic entities that belong to the same family (Thibaud et al. 2013). Claramunt and Thériault (1995) stated that events connect the geographic entities distributed across land to form independent networks. They proposed a theoretical structure that distinguishing between spatial, temporal, and thematic domains. However, the theoretical model does not directly describe the whole lifecycle of geographic events, and it mainly focuses on the changes between different time steps. Yuan (2001) made one of the first attempts at extracting states, processes, and events out of time series snapshots of precipitation data and stated that an event was a spatiotemporal aggregate of a process and that a process was a sequential change of state over space and time. McIntosh and Yuan (2005) followed Yuan's (2001) approach to organize snapshots of distributed geographic phenomena into zones, sequences, processes, and events using rainfall as an example. Although the approach was innovative, more efficient computational representations are needed to advance the extraction, exploration, and analysis of the identified events. Event representation is important not only for understanding the composition of events, but also for storing and analyzing the events.

Recognizing the need for an automated methodology to extract and represent dynamic geographic phenomena from large spatiotemporal datasets, the goal of this work is to develop event-based computational approaches to facilitate the identification, representation, and analysis of geographic events in space and time. The underlying idea is that an event can be systematically delineated with an origin, a development stage, a movement stage, and a potential cessation or dissolution phase. A directed spatiotemporal graph model is proposed to represent the dynamic characteristics of events, and graph algorithms are explored to generalize and analyze the events. The directed spatiotemporal graph model is not entirely new and it has been used to study geographic dynamics (Guo et al. 2010, Del Mondo et al. 2010, Stell et al. 2011, Thibaud et al. 2013). Del Mondo et al. (2010) used the spatiotemporal graph model to represent the spatial, spatiotemporal and filiation relations, and Thibaud et al. (2013) applied the model for marine dune dynamics analysis and representation. However, previous studies mainly focused on the visualization of dynamic geographic phenomena using the graph model. There is still a gap in applying graph algorithms to analyze the dynamics of geographic phenomena.

The method developed here will be applied to storm events inferred from weather radar reflectivity images (1 km spatial resolution, and 5-minute temporal resolution). Precipitation occurs over a large range of spatial and temporal scales, from a convective air mass thunderstorm that persists for one hour to frontal precipitation stretching across many states that can persist for days. Atmospheric conditions ultimately control the precipitation. While it is possible to use this method for any type of precipitation, our work will apply the method to a multiday convective storm outbreak. There are several reasons for this choice. Many discrete storms (events) occur during a single severe weather outbreak. Convective storms undergo many changes over their lifetime including splits and mergers. A theoretical framework exists for how these storms

behave under given atmospheric conditions. Finally, severe convective storm events are a public safety concern since they generate heavy rain, hail, and lightning strikes, which can potentially cause damage to lives and property (Han et al. 2008). Being able to objectively extract relevant information from vast amounts of radar data during severe storm outbreaks is an important step forward in constructing a better climatology of convective storms by better quantifying their lifecycles and movement characteristics (i.e., initiation, development, splitting, merging, and dissipation).

This study intends to illustrate how the graph model can be used to represent and analyze dynamic geographic phenomena since the methods and data model can be extended to other dynamic environmental events. Approaches for automatically identifying and tracking convective storm events are described in Section 2.2. Sensitivity analysis of the tracking algorithm to the reflectivity, area, and overlap threshold is discussed in Section 2.3. The directed spatiotemporal graph model for representing, storing, and analyzing storm event lifecycles is presented in Sections 2.4 and 2.5. A case study illustrates the capability of the proposed methods in Section 2.6. A summary of the research and possible future work is provided in Section 2.7.

2.2 Tracking storm events

2.2.1 Detection of storm objects

Geographic phenomena in space and time are usually identified as a field, an object, or a field-object (Goodchild et al. 2007). Different criteria such as scale, boundary, attributes, and processes (Bian 2007) were used to extract objects of interest from different data sources. Many distributed dynamic geographic phenomena like rainfall have properties that vary across space and time. The existence and delineation of these objects depend on the thresholds used to define

them (McIntosh and Yuan 2005). For a storm object that is remotely sensed by a precipitation radar, the object is a contiguous region of high radar reflectivity separated from other areas of high reflectivity (Lakshmanan et al. 2009).

The first step is the identification of storm objects on each radar image. A widely used approach is to extract a set of connected pixels that is above an intensity threshold (Dixon and Wiener 1993, Feidas and Cartalis 2001, McIntosh and Yuan 2005, Tucker and Li 2009). The connected pixels are delineated through an approach known as component labeling in digital image processing (Haralick and Shapiro 1992) and region group in raster GIS software packages. Using a single, fixed intensity threshold often works well for intense storm objects, but for initiating storm objects there may only be several pixels over the threshold (Lakshmanan et al. 2009). To help mitigate this problem, different thresholds can be applied to distinguish different types of storm objects. For example, Johnson et al. (1998) extracted storm objects using seven thresholds from 30 dBZ to 60 dBZ. The lowest 30 dBZ was used to identify storm objects and then the threshold was increased to extract more intense storm objects. Because of seasonal, regional, and climatological variability, a more general and advanced algorithm, the watershed transform algorithm, was also used in many studies (Lakshmanan et al. 2009, Zahraei et al. 2012). The lack of predefined thresholds is the biggest advantage of the watershed transform method because it tests all possible thresholds (Lakshmanan et al. 2009).

Storm object identification is simplified in this study by choosing the single threshold technique so that the focus is on the representation and analysis of the dynamic geographic phenomena. A storm object is defined as a contiguous region where the reflectivity and area are both above certain thresholds. A component-labeling algorithm with 4-connected radar reflectivity pixels was applied to extract storm objects. Since we are not interested in weak

events and focus on convective storm events, the reflectivity values should be between 30 and 40 dBZ (Dixon and Wiener 1993), and the reflectivity threshold was set to 35 dBZ in our study. The area threshold is 20 km^2 to remove noise and ground clutter and is similar to other studies (Dixon and Wiener 1993, Lakshmanan et al. 2009). We, however, investigate the sensitivity of our tracking algorithm to those two thresholds, where reflectivity was set to 30, 35, and 40 dBZ, and area was set to 20, 25, and 30 km^2. The results are discussed in Section 2.3.

2.2.2 Tracking of storm events

A critical component of a storm-tracking algorithm is to link the storm objects in one snapshot to the storm objects in the previous snapshot (Lakshmanan and Smith 2010). This linkage builds storm objects' correspondence/matching over time into an event. A large body of literature exists on tracking storms using satellite or radar data. A prominent storm matching algorithm used throughout the world is the Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN) (Dixon and Wiener 1993). In TITAN, spatial overlap and combinatorial optimization matching are combined. A storm object at t_i gets the same trajectory of the storm object at t_{i-1} that has significant spatial overlap. If there is no significant overlap of the storm objects at successive snapshots, the Hungarian algorithm would be performed, which is an optimization algorithm that considers similar characteristics (size, shape, etc.) and moving distance among the matching storm objects. Johnson et al. (1998) employed a different method that calculated the centroid distance of storm objects within a specified search radius to determine if the storm objects belong to the same trajectory.

These two major tracking methods are both centroid-based methods that first extract the separate storm objects from individual radar or satellite image and then track the storm objects

over consecutive images. These methods track individual storm objects efficiently and calculate the properties of storm objects at each temporal instant (Johnson et al. 1998). Another type of tracking method is the cross-correlation algorithm, which calculates the motion vector field and forecasts the movement of storm objects (Li et al. 1995). There are also studies using the optical flow technique to infer the velocity pattern of moving objects (Horn and Schunck 1981). In meteorological study, for example, Bowler et al. (2004, 2006) used the optical flow constraint for an improved radar echo tracking algorithm. The strength of the cross-correlation and optical flow approach is determining the direction and velocity of storm objects. However, they cannot identify and track single storm objects (Johnson et al. 1998).

We also developed a refined centroid-based algorithm that simultaneously considers the topology/spatial overlap, centroid distance of storm objects, and movement direction. Any tracking method that uses spatial overlap is ultimately dependent on the temporal sampling frequency of the dataset. There must be a high enough sampling rate to detect spatial overlap (Turdukulov et al. 2007). The radar data used in this research samples every 5 minutes. Typical convective storm motion is around 16 m/s / 58 km/h (Mohee and Miller 2010), so an average storm moves ~4.5 km between samples. Given the area threshold of 20 km^2, the temporal resolution is more than sufficient.

To outline the method, three consecutive snapshots at different time (t_1, t_2, and t_3) are used as an example (Figure 2.1). There are a total of five storm objects (a_1, b_1, c_1, d_1, e_1) at t_1. The locations of the storm objects at t_2 are predicted from t_1 and recorded as t_2'. Because the storm objects at t_1 were at the beginning of their trajectories, their velocities are initialized as zero. As a consequence, the storm objects at t_2' keep the same locations as they are at t_1. Based on the storm

movement speed and temporal resolution of the radar data, a centroid distance threshold of 10 km is used to search possible candidates that match objects at t_2 (a_2, b_2, c_2, d_2, e_2) with the

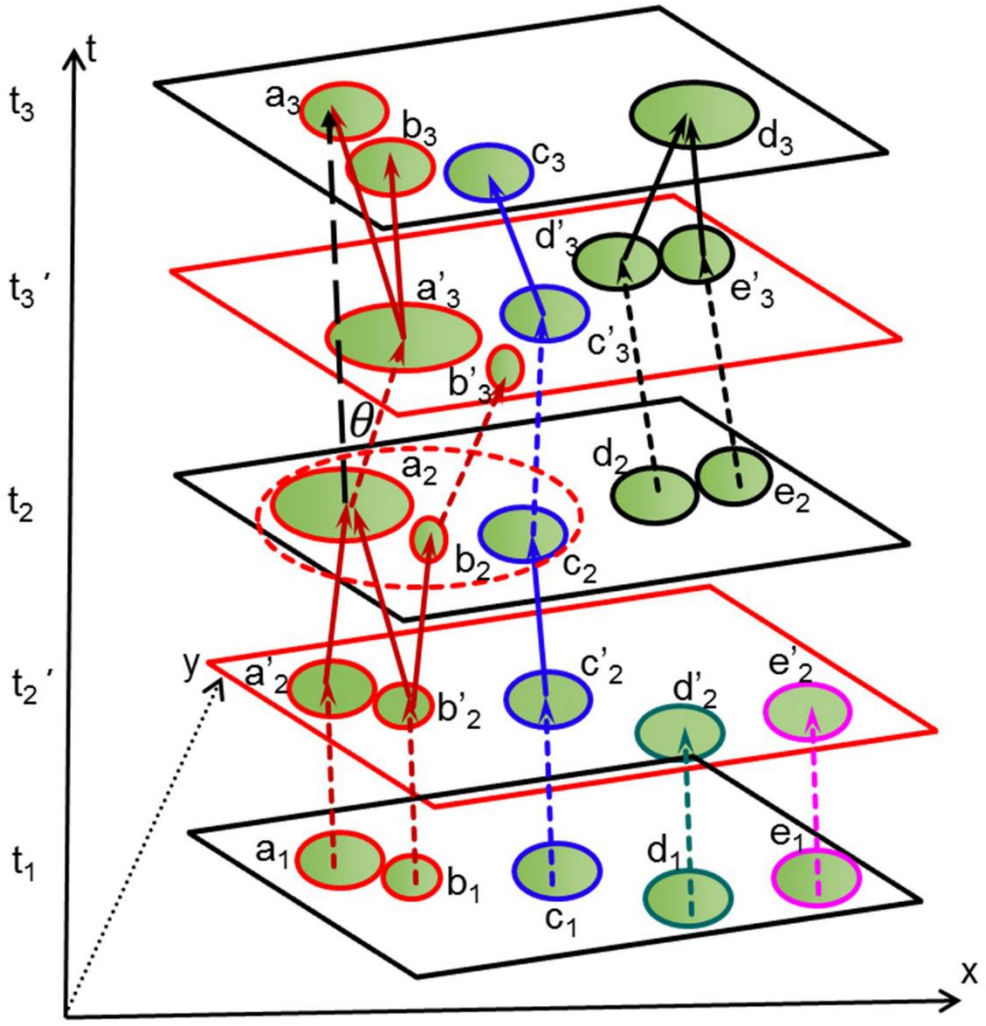


Figure 2.1 Storm event tracking examples during three time steps.

objects at t_1. For example, objects a_2, b_2, and c_2 within the red dashed circle at t_2 are matching candidates for a_1 at t_1. If the centroid of a storm object at t_2 was more than 10 km from the centroid of its nearest storm object at t_2, then this storm object is not matched with any storm objects at t_2. This indicates that the storm has dissipated and the event has ended, e.g., d_1 and e_1. Among all the matching candidates that satisfy the centroid distance threshold, a spatial overlap

function F (Equation 1) is then calculated to determine whether the candidates belong to the same trajectory of the storm objects at t_1. In this equation, $A(O)$ is the overlap area between two storm objects at the two time steps, and $A(S_1)$ and $A(S_2)$ are the area of the storm objects at the first and second time step, respectively. Based on the equation, an exact spatial overlap between two storm objects results in a value of 2 for F. No spatial overlap results in a value of 0 for F. If the value of F is above a certain threshold, the two storm objects are considered in the same trajectory, and therefore belong to the same event. In this research, the threshold of F is set to 0.6 based on the TITAN algorithm (Dixon and Wiener 1993). The sensitivity of the tracking algorithm to F is discussed in Section 2.3.

$$\mathrm{F} = \frac{A(O)}{A(S_1)} + \frac{A(O)}{A(S_2)} \tag{1}$$

After calculating the above spatial overlap function, a situation that two or more objects satisfy the threshold may arise. In our example, a_1 and b_1 both satisfy the F threshold with a_2, and b_1 could also match with b_2. The four storm objects belong to the same trajectory/event. Storm object c_2 could match with c_1. Storm objects d_2 and e_2 do not match with any objects at t_1, so they are the initial storm objects of new events which start at t_2. Storm objects at t_2 are then matched with the storm objects at t_3 using the same method. Since the initial storm velocity at t_2 is no longer zero, the predicted velocity of storm objects at t_2 is calculated as follows:

$$\mathbf{V}(S) = \frac{1}{2\sum_{i=1}^{n} A(S_i)} \left[\frac{1}{\Delta t} \sum_{i=1}^{n} \left(A(S_i) \times \boldsymbol{S_i S} \right) + \sum_{i=1}^{n} \left(A(S_i) \times \boldsymbol{V}(S_i) \right) \right] \tag{2}$$

where $\mathbf{V}(S)$ is the velocity of a storm object S at t_2. S_i represents the storm objects at t_1 that have the same trajectory as S. For example, if S is a_2 at t_2, S_i represents a_1 and b_1 at t_1. $A(S_i)$ is the area

of the corresponding storm object at t_1, $\boldsymbol{S_iS}$ is the movement vector between the centroids of mass of S_i and S, and $\boldsymbol{V}(S_i)$ is the velocity of S_i. If S is a newly generated storm object at t_2, its velocity is the same as its nearest storm object. However, if the centroid distance of the two storm objects is more than 10 km, then $\mathbf{V}(S)$ is considered to be zero (Morel and Senesi 2002). The area of storm objects at t_1 is used as a weighting factor to better predict the velocity of storm objects at t_2. The centroid distance and spatial overlap thresholds are also performed to match storm objects at t_2 and t_3.

When the size of a storm rapidly expands or contracts, its centroid could change significantly between consecutive snapshots. This may produce unrealistic storm movement. To include only realistic storm movement, the change of movement direction is also checked when deciding whether two storm objects belong to the same event. The angle θ between the predicted movement direction of a storm object at t_2 and the direction from its centroid to the centroid of a matching storm object at t_3 is calculated. Only the matching storm objects with an angle less than 90 degrees are considered in the same event.

With our extraction algorithm, there are six filiation relationships: generation, continuity, split, merger, combinatorial, and dissipation (Zahraei et al. 2012) between storm objects in an event. In Figure 2.1, five storm events are extracted (shown with different colors in Figure 2.1) from three consecutive snapshots. Storm object d_2 is newly generated because it is not associated with any storm objects in the previous time step. Storm object c_2 is a continuation of storm object c_1 from t_1 to t_2. Split means a storm object at time t_{i-1} is associated with two or more storm objects at time t_i (Morel and Senesi 2002). In Figure 2.1, storm object a_2 splits into a_3 and b_3. When a merger occurs, two or more storm objects at time t_{i-1} can be linked to a storm object at time t_i. For example, d_2 and e_2 merged into d_3 from t_2 to t_3. Split and merger occur

simultaneously in the combinatorial situation. For example, one part splitting from b_1 merges with a_1 to form a_2, and the other part that splits from b_1 contributes to the initiation of b_2. Dissipation occurs when a storm object is not matched with any objects at the next snapshot such as storm object b_2 at t_2 that disappears at t_3.

Previous studies (Dixon and Wiener 1993, Morel and Senesi 2002, Han et al. 2009, Zahraei et al. 2012) dealt with merging and splitting cases by extending a maximum of one trajectory and terminating the remainder for a merger case or by extending a maximum of one trajectory and generating new trajectories to the remainder for a split case. This method is comparatively easy, but it cannot capture the complete lifecycle and the interactions among storm objects. In contrast, our method records the filiation relationships among all the storm objects in the time series snapshots which satisfy the matching criteria (overlapping area, centroid distance, and movement direction).

2.3 Sensitivity analysis

One of the key aspects of any method that uses thresholds is how sensitive the results are to changes in the threshold value. There are three values that must be set in this method: the minimum reflectivity, area, and overlap. The choice of any particular threshold is dictated by the application. For instance, the focus of this application is identifying and tracking convective storm events that are at least of moderate strength, so the minimum reflectivity is set to 35 dBZ, the minimum area is set to 20 km^2, and the overlap is 0.6. It is important to understand how sensitive the results are to the choice of threshold, especially if the ultimate goal of this method is to construct a climatology. Sensitivity is obtained by varying the reflectivity to 30, 35, and 40 dBZ, the area to 20, 25, and 30 km^2, and the overlap was incremented by 0.2 from 0 to 2.

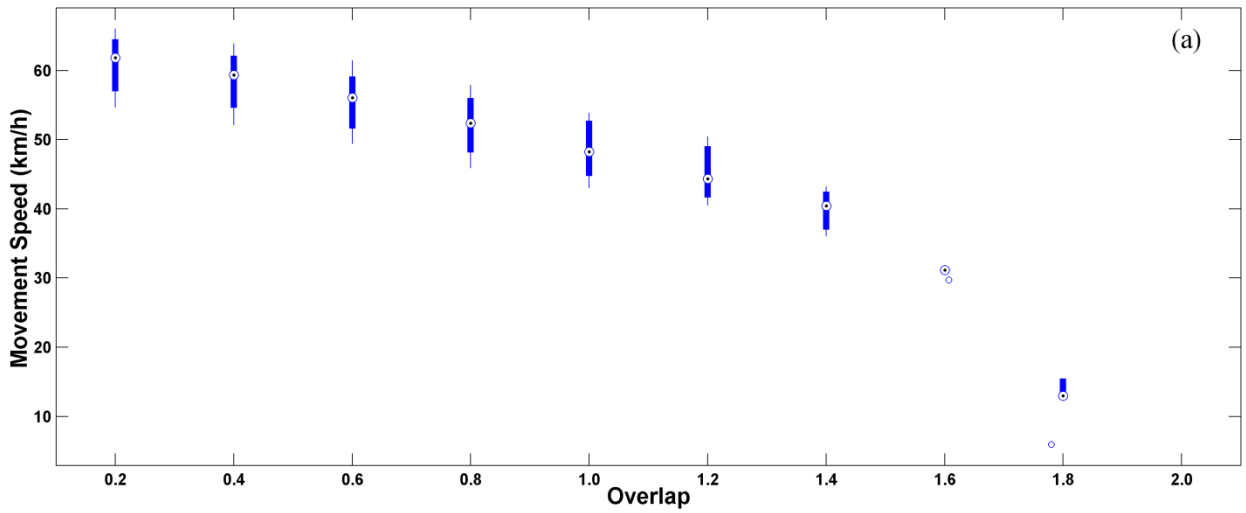
Varying all of these leads to 90 combinations. To save computational time, the test area is chosen 32.5°N-40.5°N and 93.5°W-103.5°W (the blue dashed boundary in Figure 2.6), and to the storm outbreak occurred 23-29 April 2011.

If the thresholds are more restrictive (e.g., higher minimum reflectivity threshold or minimum area), then there will of course be a fewer number of total storms. A more useful metric for sensitivity would be changes to physical parameters such as movement speed. There is some information on climate statistics of storm motion, but these studies tend to be spatially and temporally very limited. For example, a simple climatology of storms that does not include any information on storm splitting or merging was created for a five year period over just North Dakota (Mohee and Miller, 2010). They found an average movement speed of 16.4 m/s (59 km/h). Even though this is not the same study area, it at least provides some benchmark that can be used to compare results from the more sophisticated method developed here.

Box plots in Figure 2.2 demonstrate the sensitivity of movement speed to changes in the three thresholds. In Figure 2.2a, there are nine combinations of reflectivity and area threshold for each overlap threshold. It is easily to see that the movement speed is very sensitive to the overlap threshold, and the overall movement speed decreases with the increase of overlap threshold. The explanation is that the lower the storm events' speed, the more overlap between two consecutive images. Using 59 km/h as an expected value, the movement speed is likely greatly underestimated when overlap threshold is greater than 1.4. Using thresholds of 0.4, 0.6, and 0.8 are much closer to the expected value and are a more appropriate choice. Sensitivity of movement to the reflectivity and area thresholds is much less, which is expected since movement of any developing storm, regardless of size or intensity, depends primarily on the steering wind. Figure 2.2b suggests that an increase of reflectivity is weakly correlated to an increase in the

movement speed. This may actually not be an artifact of the methodology, but a real phenomenon. Particularly intense storms develop their own internal structure (pressure perturbations) that accelerates the movement of the storm complex. Of course, these findings will have to be examined in more rigor and detail and is the next step after developing this method.

Finally, Figure 2.2c suggests that the movement speed is not sensitive to the area threshold. The above analysis reveals that the storm tracks are relatively insensitive to the reflectivity and area thresholds in the storm identification step, but the storm tracks are the most sensitive to the overlap threshold. We stress that proper choice of thresholds depends on the research question, and this method allows for easy sensitivity analysis by simply changing thresholds in the storm identification step. For our analysis, we choose 35 dBZ, 20 km^2, and 0.6 overlap. We are not recommending that these are the correct or only thresholds for convective storms research. Researchers could adjust these thresholds based on the sensitivity analysis and their own needs.



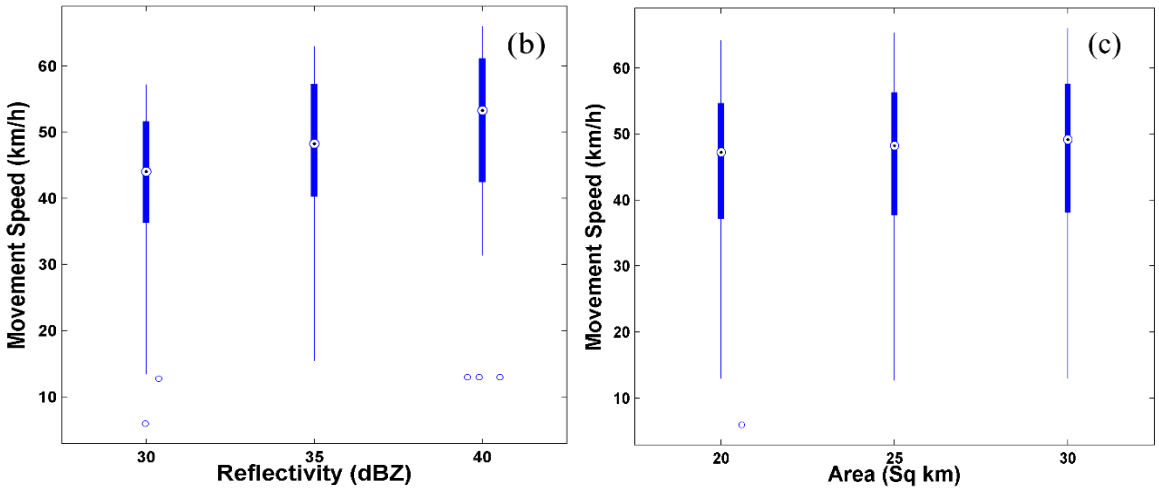


Figure 2.2 (a) Overlap vs Movement speed. (b) Reflectivity vs Movement speed. (c) Area vs Movement speed.

2.4 A directed spatiotemporal graph model

It is natural to use a directed spatiotemporal graph model (Figure 2.3a) to depict the evolution, change, and interaction of storm events. The nodes in the graph, *V (G)*, represent the spatially contiguous storm objects at each time step. The directed edges in the graph, *E (G)*, denote the spatial and temporal linkages (i.e., filiation relations) among storm objects at two adjacent snapshots where direction indicates the time sequence. As a result, this graph model contains spatial, temporal, and filiation relations. The vector polygonal footprint represents the geometric shape of a single storm object. The approximate ellipse in Figure 2.3a is used as a simplified representation of the storm object geometry, which works fairly well for distinct storm objects. Figure 2.3a is the three-dimension view (x, y, t) of the directed spatiotemporal graph derived from Figure 2.1. To facilitate the analysis of the storm events, the spatiotemporal graph is also projected in two dimensions (x, y) in Figure 2.3b. The nodes are further simplified using the reflectivity-weighted centroid of a storm object to capture the core of the most intense precipitation. This graph model describes the evolution of a storm event at three consecutive

snapshots over the entire lifespan of the storm and includes the generation, continuity, split, merger, combinatorial, and dissipation filiation relationships.

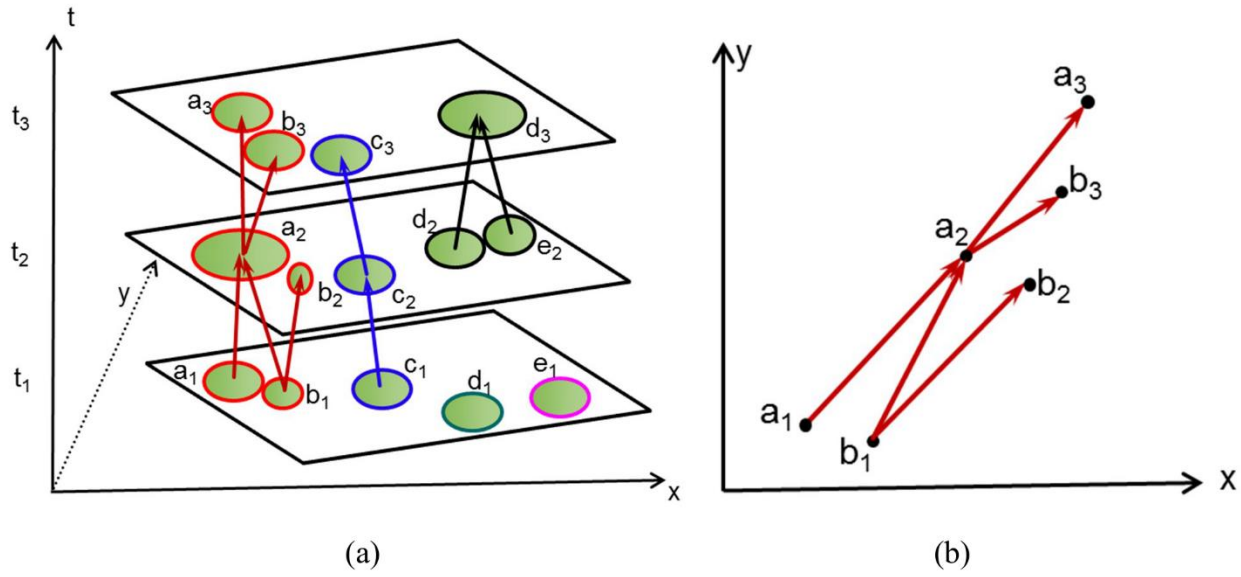


Figure 2.3 (a) The storm events and their directed spatiotemporal graph models identified from Figure 2.1. (b) The projection of the first storm graph on the x-y plane.

A number of node-level, edge-level, and event-level spatial and non-spatial attributes are stored in a graph database to represent the storm events (Figure 2.4). A node object stores the information of a single storm object, and an edge object describes the filiation relationships among storm objects. An event graph object describes different features of a storm event, such as the number of nodes and edges, duration, movement speed and direction, and other attributes. All the nodes and edges within the same storm event are linked to the event graph with a many-to-one relationship. As a consequence, the GraphID is used as a foreign key in the node and edge object.

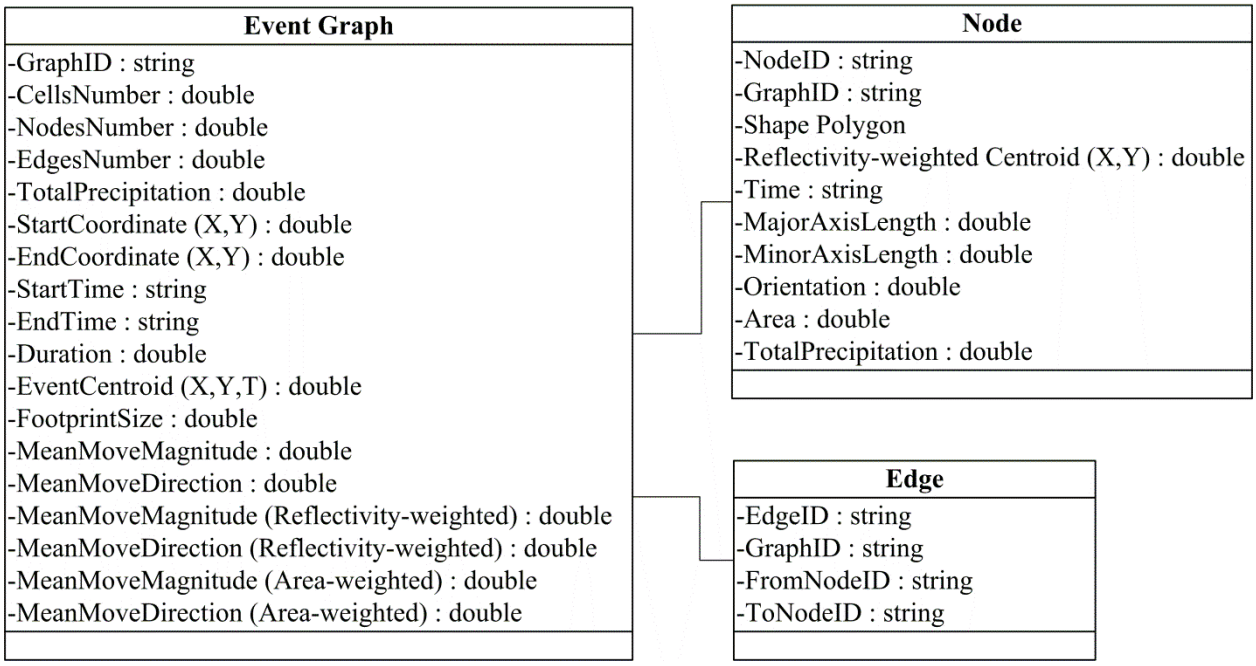


Figure 2.4 Components and their primary properties of the directed graph model for storm events.

2.5 Graph-based storm event analysis

The full advantage of our graph model is realized when the graph properties of storm events are investigated. The remainder of this section will concentrate on how to use graph theory/algorithms to generalize and assess storm events' interactions.

2.5.1 Generalization of Storm Events

Complex or detailed geographic phenomena often require simplification or generalization to understand (Guo et al. 2010). Any generalization must capture the key properties of the original geographic phenomena. The structures and relationships of storm objects on consecutive images can be quite complicated. As a consequence, the graph representation could be too detailed and complex, especially when the storm events have many small storm objects. The distribution of reflectivity/precipitation is one of the most important characteristics of storm events. A

generalization of an event, i.e., the skeleton of the event, can be achieved by using the maximum reflectivity path between the starting and ending storm objects of the event (Figure 2.5). The maximum reflectivity path can be found by first calculating the reverse reflectivity of the i^{th} node as:

$$W_i = \max(R) - R_i + \min(R) \quad (3)$$

where max(R) and min(R) are the maximum and minimum reflectivity of all the storm nodes in a graph, and R_i is the reflectivity of the i^{th} node. The reverse reflectivity of the i^{th} node is used as the weight of all the in-edges of the i^{th} node in the directed graph. The lesser W_i as weight for a node, the higher reflectivity. The maximum reflectivity path can then be obtained using the classic Dijkstra's shortest path algorithm on the graph. Specifically, the maximum reflectivity paths between each starting and ending storm nodes are first calculated and the maximum reflectivity path is then identified for the storm event. The generalization method can be applied recursively to any branches that are connected to the generalized event path.

2.5.2 Interaction among storm objects

It is common for storm objects to interact over consecutive images. In a storm event, two independent storm objects may merge into one object or one object may split into several smaller objects. These interactions can be depicted as the in-degree/out-degree of a node in a directed graph. The in-degree of a node is the number of edges directed into that node. The out-degree of a node is the number of edges directed out of that node. The number of in-degree/out-degree could reveal the interactions or filiation relationships among storm objects (Table 2.1). In our case study, we will further examine the split and merger that occur on the left and right side along the main movement direction of an event (Figure 2.5).

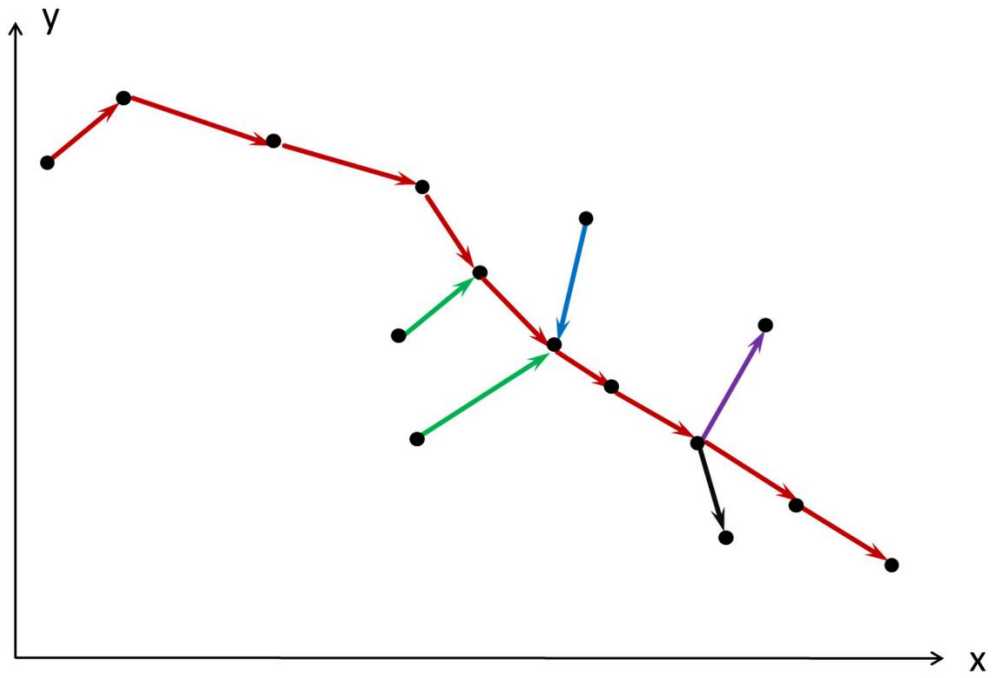


Figure 2.5 A storm event graph showing the maximum accumulative reflectivity path (red line), left split (purple line), right split (black line), left merger (blue line) and right merger (green line) along its movement direction.

Table 2.1 The relationship between node in-degree/out-degree number and storm object filiations.

Number of in-degree	**Filiation Relation**	**Number of out-degree**	**Filiation Relation**
0	Start	0	End
1	Continuation	1	Continuation
>= 2	Merge	>= 2	Split

2.6 A Case Study

A convective storm outbreak that occurred 23-29 April 2011 is used to illustrate both how the directed spatiotemporal graph model is constructed and also what kind of information can be obtained about the spatiotemporal characteristics of storm events (Figure 2.6). The storm outbreak contained many discrete cells and there were 355 tornados confirmed by the National

Weather Service. This was a high impact event that caused substantial property damage and fatalities.

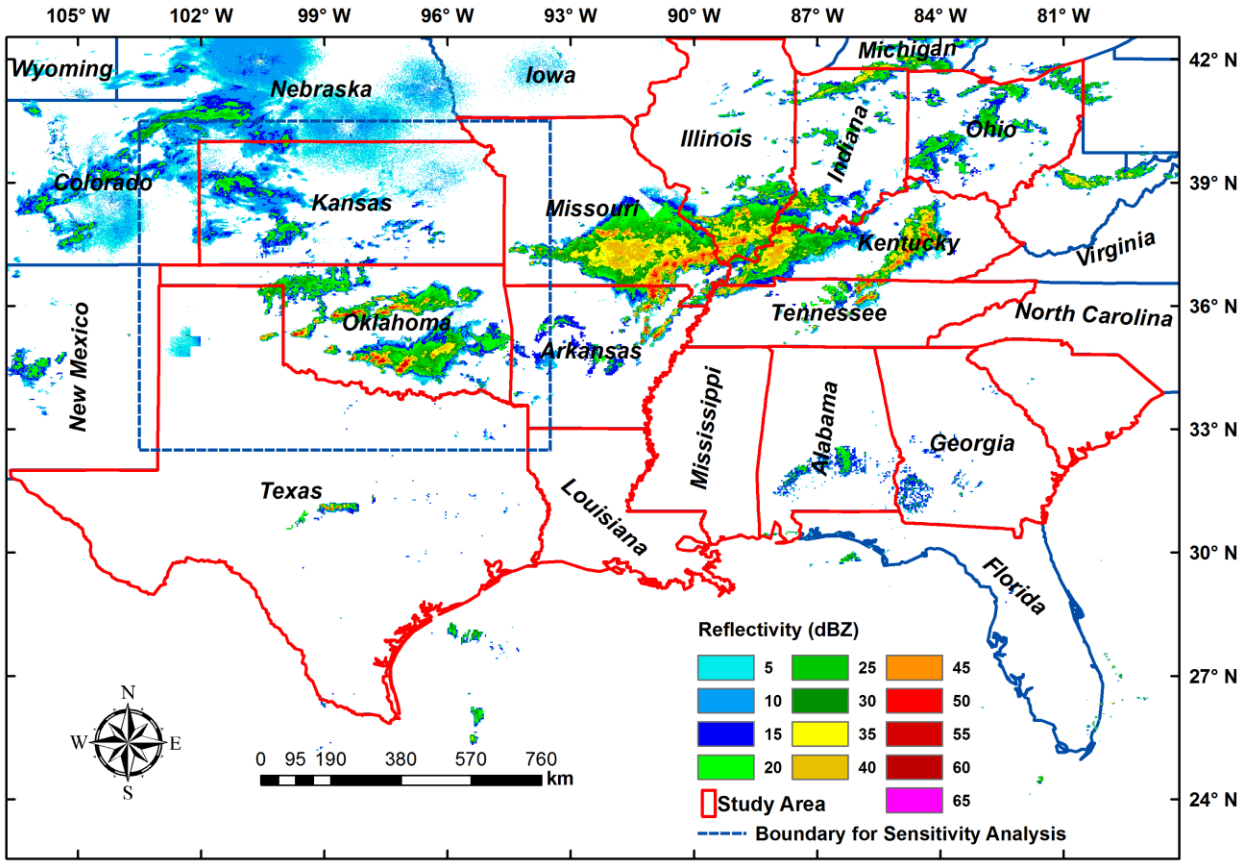


Figure 2.6 Case study area which covers 15 states (red) and an example radar image showing the reflectivity values at 8:00, 24 April 2011 UTC.

The data used in this case study are the final reflectivity product (N0R) provided by Iowa Environmental Mesonet (IEM). The United States National Weather Service operates the NEXRAD (Next Generation Radar) program that monitors precipitation over almost all of the country. This is a network of S-band (10 cm) Weather Surveillance 1988 Doppler radars (WSR-88D, Choi et al. 2009), which has been recently upgraded to dual-polarization radar. The IEM receives, processes, and archives the NEXRAD level III products with a 5-minute temporal resolution (http://mesonet.agron.iastate.edu/docs/nexrad_composites/). The radar reflectivity

images are stored in the PNG format with a WGS84 spatial reference system (EPSG: 4326) that has a spatial resolution of equivalent to 1 km.

Convection can be classified into several archetypes including squall lines, multicellular, bow echoes, derechos, super cells, disorganized, and other subcategories (Smith et al. 2012). The specific dynamics of an individual storm depends on many factors including wind shear, magnitude of instability, interaction with neighboring cells, and so on. One of the most studied type of storm is the supercell. Theory, numerical simulations, and observed case studies have shown that if a supercell splits when there is a veering wind profile, then the right-moving cell is likely to strengthen while the left-moving cell is likely to weaken. Conversely, if there is a backing wind profile than the left-moving cell is likely to strengthen while the right-moving cell is likely to weaken. Veering wind profiles are far more common than backing wind profiles so that any climatology of supercells that split should reveal a preference for right-moving cells while the left-movers would dissipate. The distinct advantage of this method over past techniques is the inclusion of the splitting and merger information that can be used to create a preliminary climatology of the behavior of convective precipitation.

While we recognize that convection in this case study is not all strictly one mode (e.g., supercells), we examine all convection as a single group when identifying the characteristics of this particular outbreak. Separating convective mode is not a trivial task and often requires subjective methods (e.g., Smith et al. 2012). The primary goal of this work is to begin to objectively construct and apply a directed spatiotemporal graph model to identify, store, represent, link, and track storm objects. This initial step must also efficiently deal with spatiotemporal data management because of the high volume of radar data. Work is already underway to refine storm classification and interaction, but that is beyond the scope of this paper.

The method developed here is meant to be the foundation for continued study of these complex systems.

2.6.1 Implementation

A prototype system was developed using MATLAB to process the reflectivity data, delineate storm objects, track storm events, visualize, verify, and analyze the events. There are four primary components in the work flow (Figure 2.7): the spatiotemporal database generation from raw NEXRAD snapshots, storm object identification, storm event tracking and event graph generation, and storm event visualization and analyses based on graph theory/algorithms.

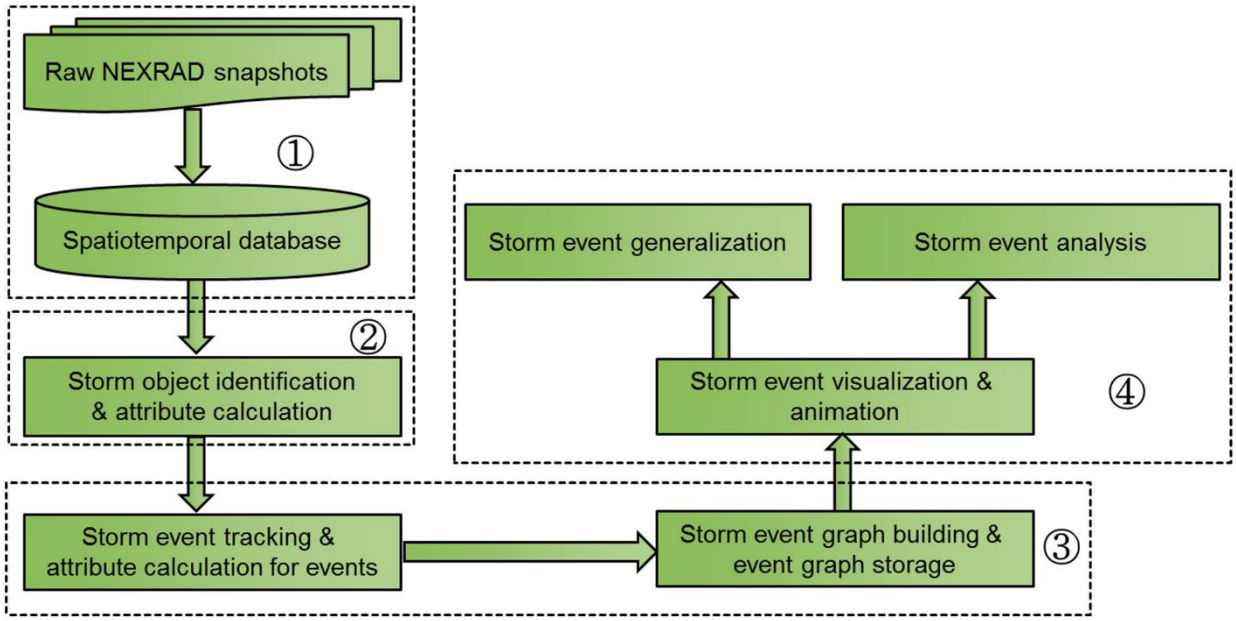


Figure 2.7 The general workflow of storm event identification and analysis developed using MATLAB.

Raw radar snapshots of PNG files are converted into the MATLAB file format (.mat files) to build a spatiotemporal database that is used as the input for the next step. After applying the threshold to a radar image, a component labeling algorithm delineates spatially contiguous storm objects and their properties are calculated and saved with the objects (component 2 in Figure 2.7).

The tracking process outlined in section 2.2 is then applied to track the storm events based on the storm objects delineated on two consecutive images. During the tracking process, the lineage or filiation relations are recorded to build event graphs. When all the nodes in an event graph at t_i cannot find any matching storm objects at t_{i+1}, the complete event graph (i.e., the whole lifecycle) of the storm event has been identified. The event graph, including its nodes, edges, and attributes at node/edge/graph levels are saved and the event is deleted from memory. The above process is applied sequentially to all the images to identify the events in the database (component 3 in Figure 2.7). The last component takes the event graphs as input and provides functions for event visualization, animation, generalization, and interaction assessment. The visualization and animation interface, shown in Figure 2.8, could be used to browse event graph attributes, the maximum reflectivity path, and the visualization and animation of event graphs. In Figure 2.8, a storm event is shown with the red skeleton and one right split with black line.

2.6.2 Characteristics of storm events

The program was run on a UNIX server machine with 2 six-core AMD Opteron 2435 processors at 2.6 GHz with 64 GB of RAM, and the total running time is about ten hours. In the study area there were a total of 7,297 storm events that have a duration of at least 15 minutes, meaning they span at least three consecutive radar images. Figure 2.9 shows the distribution and tracks of the storm events with their maximum reflectivity paths. The storm events are mainly concentrated in a broad swath from Texas to Ohio. Figure 2.10a shows that the bulk of storm paths (86.9%) are to the northeast. This movement is consistent with the mid-level steering wind during the event. The histogram of the storm event movement speed (Figure 2.10b) depicts a positively skewed Gaussian distribution. The average and median movement speeds of the storm events are 61 km/h and 63 km/h, respectively. This is around the same value found in the climatology analysis

done by Mohee and Miller (2010). The histogram of the storm event duration (Figure 2.10c) depicts an exponential decay distribution similar to Novo et al. (2013). The mean duration of the storm events is 36.5 minutes and 1,141 out of 7,297 storm events lasted more than one hour (15.6%). Most of storm events (44.4%) had a duration from 15 to 20 minutes.

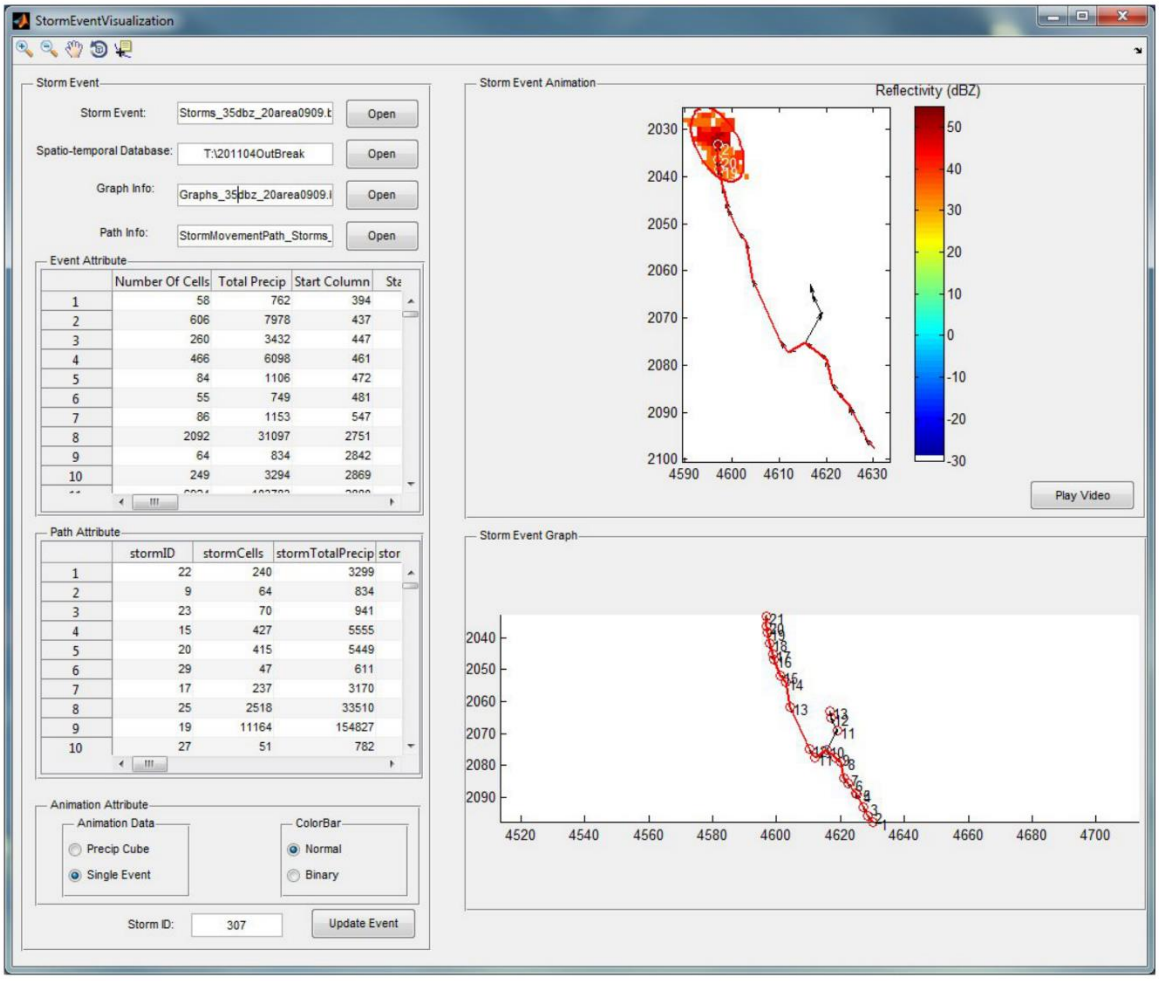


Figure 2.8 The user interface for event visualization and animation.

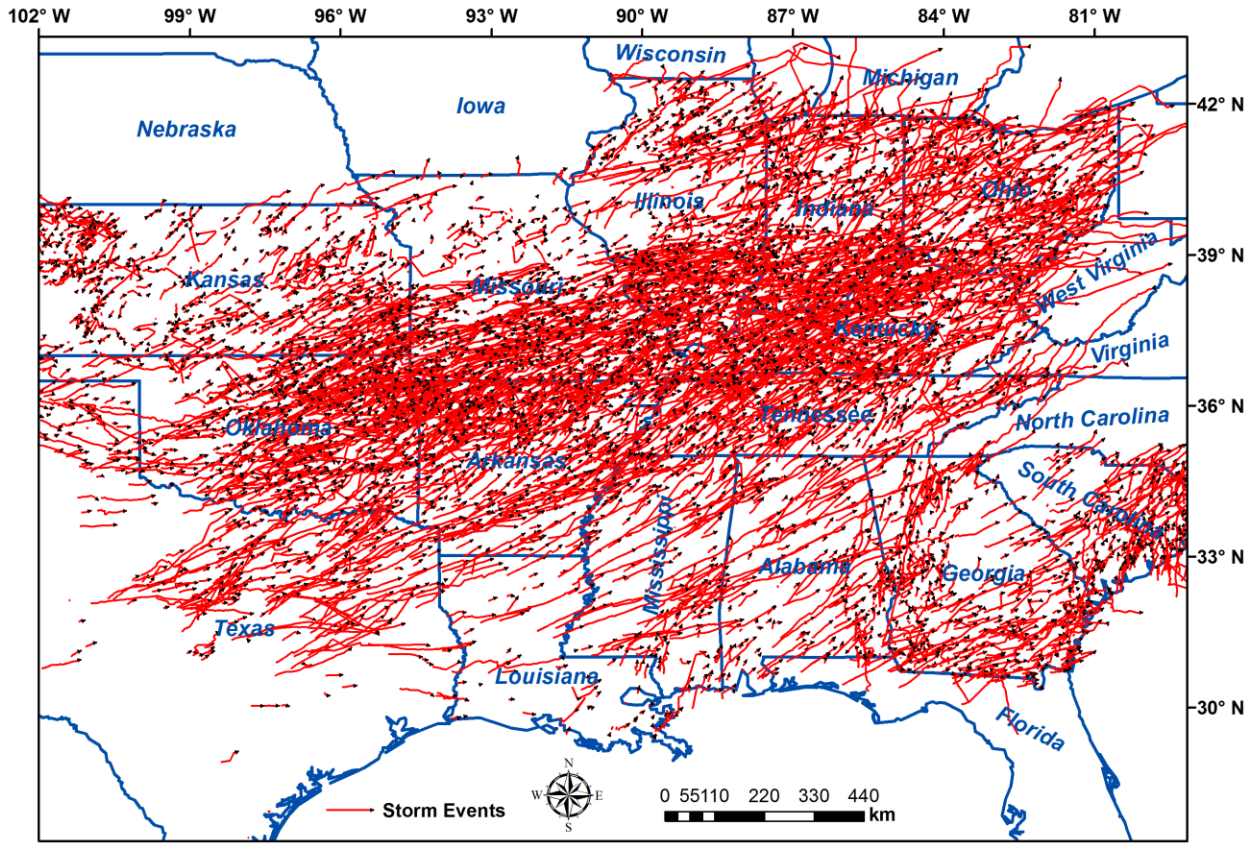


Figure 2.9 Map of generalized tracks of the storm events identified from the radar data between 04/23/2011 and 04/29/2011 in the study area.

Among the 7,297 storm events, just over a quarter of storm events (1,928) have either splits or mergers ($mean_{split}$ = 0.17, $mean_{merger}$ = 0.38). There are 863 storm events that split and most of them (796) have one or two splits during their lifespans. There are 1,529 storm events that merged and most of them (1,274) have only one or two mergers during their lifespans.

One advantage of our method is the inclusion of additional details such as storm object split and merger. In this case study, the ambient vertical wind profile was strongly veering. Thus, a reasonable hypothesis would be that given all of the convection occurring, there should be a

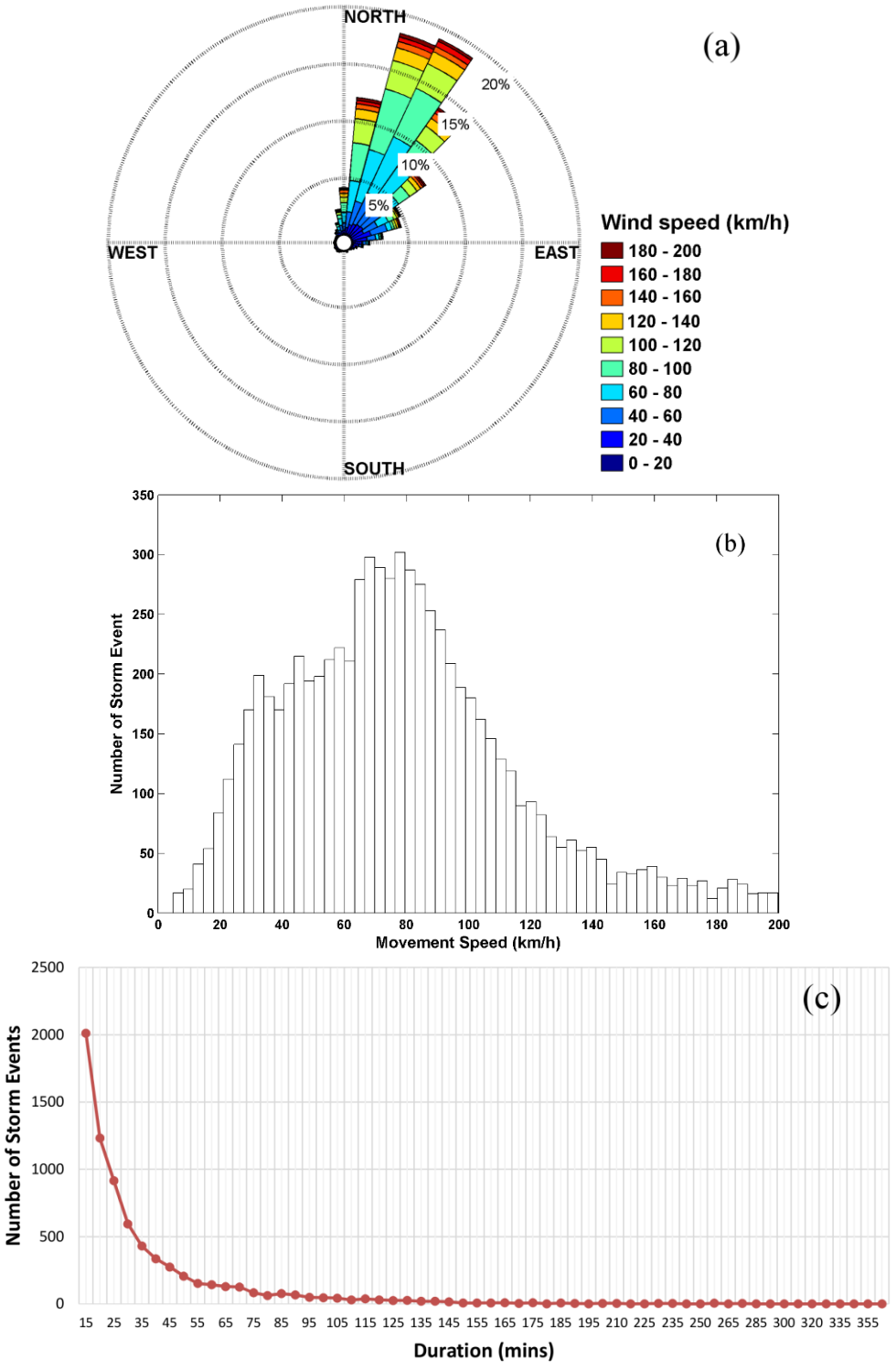


Figure 2.10 Characteristics of the storm events: (a) rose diagram of event velocity; (b) histogram of event speed; (c) histogram of event duration.

preference for right-splitting storms to survive while left-splitting storms dissipate. Again, we do not discriminate between storm modes, but the group as a whole should still show this.

The splitting and merging within the storm objects are separated into whether they are on the left or right side of the primary storm track as defined by the maximum radar reflectivity path. Examples of these splits and mergers are given in Figure 2.5. For these events, the locations of left and right split/merger are shown in Figure 2.11 and Figure 2.12. Table 2.2 gives the distribution of the left and right splits and mergers of the storm events. Among the storm events that have splits or mergers, these splits and mergers only occur once or twice during their lifespans. There are only 29 (12) events that have more than two left (right) splits. There are typically more mergers that occur during a storm event than storm splitting. There are 104 (87) storm events that have three or more left (right) mergers. It is rare for a storm event to have more than six splits or mergers.

The difference between the number of left and right splits and mergers are shown in Figure 2.11and Figure 2.12, respectively. The left split and merger appear more than the right split and merger. To test whether there is a preference for the side of splitting or merging, a *t*-test is performed on the mean of the difference between the number of left and right splitting and merging. The null hypothesis is that the mean is equal to 0, which implies that the chances of having a left or right split or merger is the same for the events. The results in Table 2.3 show that *p* values for both split and mergers are less than 0.05 so that the null hypothesis is rejected. We should accept that the left split/merger appear more than right split/merger. This is consistent with what would be expected under conditions with a veering wind profile. Again, this should really be refined by storm mode, but this is a promising result that stresses the potential of using this objective technique to better quantify the properties of convective storms.

Table 2.2 The summary of split and merger appeared per storm event. LS and RS represent left split and right split. LM and RM represent left merger and right merger.

Total number of LS per storm event	**Number of storm events**	**Total number of RS per storm event**	**Number of storm events**	**Total number of LM per storm event**	**Number of storm events**	**Total number of RM per storm event**	**Number of storm events**
1	427	**1**	367	**1**	740	**1**	600
2	76	**2**	55	**2**	191	**2**	128
3	14	**3**	7	**3**	60	**3**	40
4	7	**4**	3	**4**	22	**4**	20
5	6	**5**	1	**5**	12	**5**	13
6	1	**6**	1	**6**	6	**6**	8
8	1			**7**	2	**7**	2
				8	2	**8**	1
						9	1
						10	1
						13	1

Table 2.3 The t-test result for mean of difference between left split/merger and right split/merger.

Level of significance	**p-value for split**	**p-value for merger**
0.05	$3.768 * 10^{-6}$	$1.4726 * 10^{-7}$

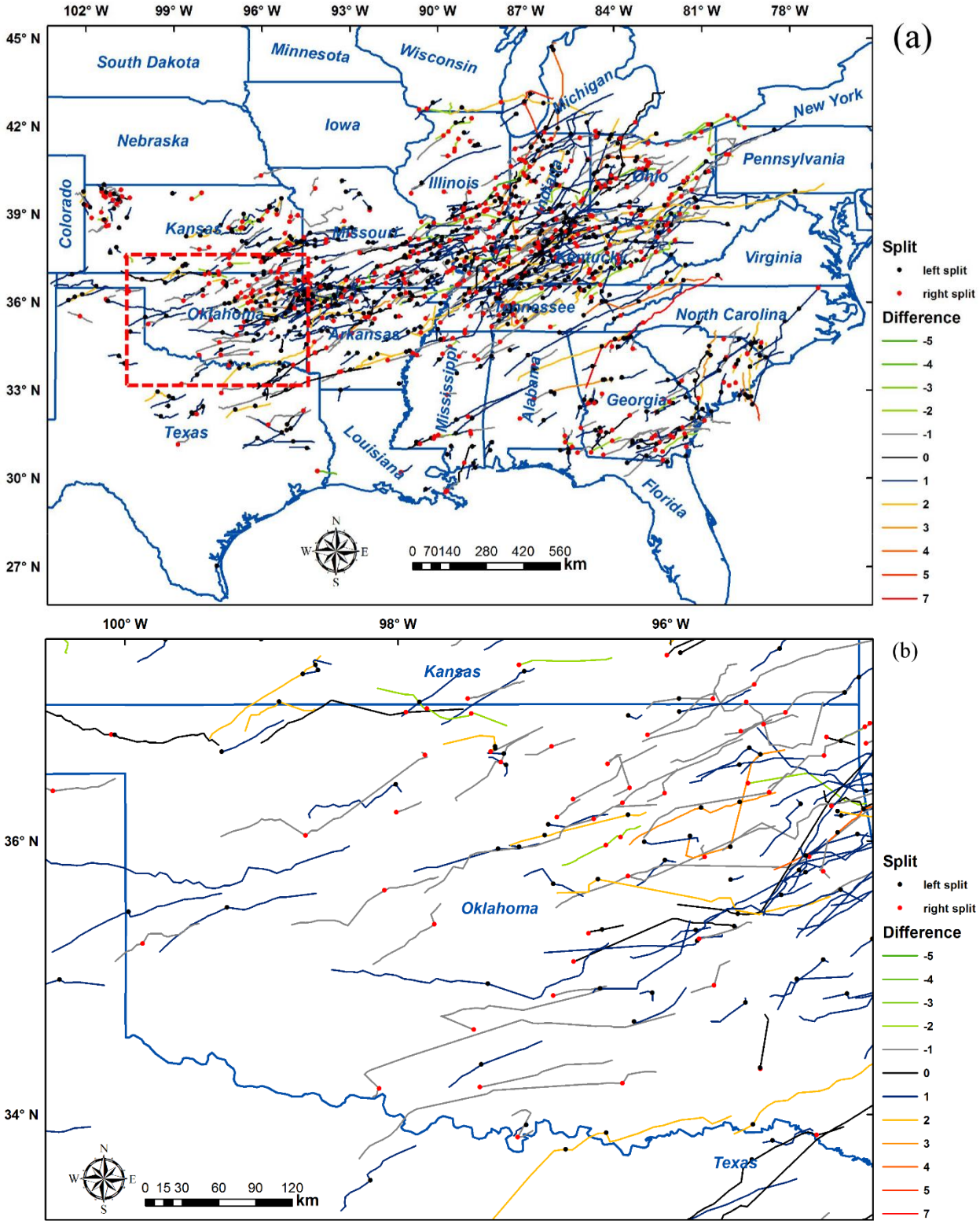


Figure 2.11 (a) The sides (left or right relative to storm movement direction) of the splits occurred in storm events, and the difference between the number of left splits and right splits. (b) A detailed view in the dashed box in (a).

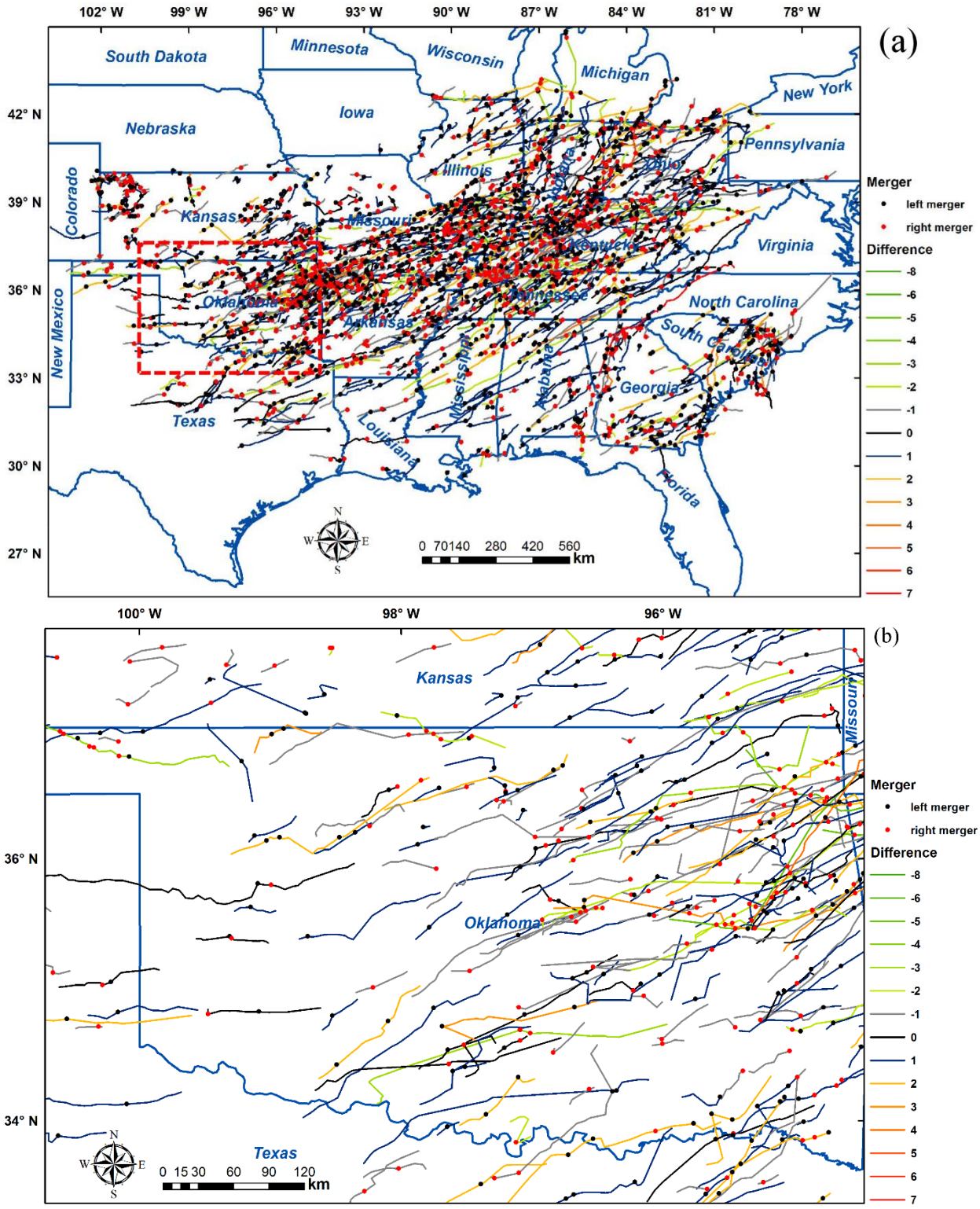


Figure 2.12 (a) The sides (left or right relative to storm event movement directions) of the mergers occurred in storm events, and the difference between the number of left mergers and right mergers. (b) A detailed view in the dashed box in (a).

2.7 Conclusions

This work presented the event tracking method and graph model to identify, represent, and analyze dynamic spatiotemporal phenomena from time series of snapshot data. Storm events were chosen in this research in part because of the multiple interactions among storm objects during an event's lifespan. A directed spatiotemporal graph model was used to represent the evolution and the filiation relationships among storm objects within an event. In the model, storm objects identified from radar reflectivity images were denoted as graph nodes. The interactions among storm objects were tracked using improved algorithms, and these relationships were denoted as graph edges. General event characteristics and side preference of split and merger in the storm events were analyzed based on graph algorithms.

The 23-29 April 2011 storm event outbreak spanning in a wide swath from Texas to Ohio was utilized to demonstrate the development and application of our event tracking method and the directed graph model. The tracking algorithm identified all the storm objects belonging to the same storm event. Various attributes at the node, edge, and event/graph level were calculated and stored. Storm events were generalized using the shortest-path graph algorithm where the cost at a graph node is radar reflectivity. Several basic properties of the convection storms that occurred during this outbreak were obtained and were characteristically consistent with previous studies. As a check on the physical consistency of this representation of the storms, the side of the split and merger was also examined. Given that the vertical wind shear was veering, it was expected that if a storm split, then the cell that moves to the right would be stronger and last longer than the cell that moves to the left. Even though the data was not separated into storm mode, there was a significant preference for the left splitting storms to be weaker and dissipate while the right splitting storms were stronger and lasted longer.

Our research demonstrated that the event approach provided an extension to GIS to represent and analyze dynamic behaviors existing in time series of spatial snapshot data. As the amount of spatiotemporal data have been constantly collected or generated with better spatial and temporal resolutions, science has entered into an era where discovery of new knowledge can be obtained through the analysis and mining of 'big data'. The event approach represents a transformation of heterogeneous spatiotemporal data into events that could be compared and integrated across time and location to support the study of the interactions and dynamic behavior of environmental systems. In the future, we plan to study the spatiotemporal characteristics of storm events in the central U.S. using long term radar reflectivity data. We are also interested in modifying and using the graph edit distance, a graph matching algorithm, to assess the similarities in geometry and movement dynamics of storm events, which may provide useful information for storm forecasting. In addition, the spatiotemporal graph model will be further validated in other application domains, such as ocean eddies, wildfire, and urban heat islands.

References

Bian, L., 2007. Object-oriented representation of environmental phenomena: is everything best represented as an object? *Annals of the Association of American Geographers,* 97 (2), 267-281.

Bowler, N. E., Pierce, C. E., and Seed, A., 2004. Development of a precipitation nowcasting algorithm based upon optical flow techniques. *Journal of Hydrology,* 288 (1), 74-91.

Bowler, N. E., Pierce, C. E., and Seed, A., 2006. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quarterly Journal of the Royal Meteorological Society,* 132 (620), 2127-2155.

Choi, J., Olivera, F. and Socolofsky, S. A., 2009. Storm identification and tracking algorithm for modeling of rainfall fields using 1-h NEXRAD rainfall data in Texas. *Journal of Hydrologic Engineering,* 14 (7), 721-730.

Claramunt, C. and Thériault, M., 1995. Managing time in GIS an event-oriented approach. *In:* J. Clifford and A. Tuzhilin, ed. *Recent Advances in Temporal Databases,* London: Springer 23-42.

Couclelis, H., 1992. People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS. *In*: A.U. Frank, I. Campari and U. Formentini, ed. *Theories and methods of spatio-temporal reasoning in geographic space.* Berlin: Springer Verlag, 65-77.

Dixon, M. and Wiener, G., 1993. TITAN: thunderstorm identification, tracking, analysis, and nowcasting-A radar-based methodology. *Journal of Atmospheric and Oceanic technology,* 10 (6), 785-797.

Feidas, H. and Cartalis, C., 2001. Monitoring mesoscale convective cloud systems associated with heavy storms using Meteosat imagery. *Journal of Applied Meteorology,* 40 (3), 491-512.

Galton, A., 1995. Towards a qualitative theory of movement. *In*: A. U. Frank and W. Kuhn, ed. *Spatial Information Theory A Theoretical Basis for GIS,* Berlin: Springer Verlag, 377-396.

Galton, A., 2000. *Qualitative Spatial Change.* New York: Oxford University Press.

Goodchild, M.F., Yuan, M. and Cova, T.J., 2007. Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science,* 21 (3), 239-260.

Guo, D., Liu, S. and Jin, H., 2010. A graph-based approach to vehicle trajectory analysis. *Journal of Location Based Services,* 4 (3-4), 183-199.

Han, L., Fu, S., Yang, G., Wang, H., Zheng, Y. and Lin Y., 2008. A stochastic method for convective storm identification, tracking and nowcasting. *Progress in Natural Science,* 18 (12), 1557-1563.

Han, L., Fu, S., Zhao, L., Zheng, Y., Wang, H. and Lin Y., 2009. 3D convective storm identification, tracking, and forecasting-An enhanced TITAN algorithm. *Journal of Atmospheric and Oceanic Technology,* 26 (4), 719-732.

Haralick, R.M. and Shapiro, L.G., 1992. *Computer and Robot Vision*. London: Addison-Wesley Longman.

Horn, B.K. and Schunck, B.G., 1981. Determining optical flow. *Artificial Intelligence,* 17 (1-3), 185-203.

Johnson, J., MacKeen, P.L., Witt, A., Mitchell, E.D., Stumpf, G.J., Eilts, M.D. and Thomas, K.W., 1998. The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Weather and Forecasting*, 13 (2), 263-276.

Lakshmanan, V., Hondl, K. and Rabin R., 2009. An efficient, general-purpose technique for identifying storm cells in geospatial images. *Journal of Atmospheric and Oceanic Technology*, 26 (3), 523-537.

Lakshmanan, V. and Smith, T., 2010. An objective method of evaluating and devising storm-tracking algorithms. *Weather and Forecasting*, 25 (2), 701-709.

Langran, G. and Chrisman, N.R., 1988. A framework for temporal geographic information. *Cartographica* 25:1–14.

Li, L., Schmid, W. and Joss, J., 1995. Nowcasting of motion and growth of precipitation with radar over a complex orography. *Journal of Applied Meteorology*, 34 (6), 1286-1300.

McIntosh, J. and Yuan, M., 2005. A framework to enhance semantic flexibility for analysis of distributed phenomena. *International Journal of Geographical Information Science*, 19 (10), 999-1018.

Mohee, F.M. and Miller, C., 2010. Climatology of thunderstorms for North Dakota, 2002–06. *Journal of Applied Meteorology and Climatology*, 49 (9), 1881–1890.

Morel, C. and Senesi, S., 2002. A climatology of mesoscale convective systems over Europe using satellite infrared imagery. I: Methodology. *Quarterly Journal of the Royal Meteorological Society*, 128 (584), 1953-1971.

Novo, S., Martinez, D. and Puentes, O., 2013. Tracking, analysis, and nowcasting of Cuban convective cells as seen by radar. *Meteorological Applications*, 21 (3), 585-595.

Peuquet, D. and Duan, N., 1995. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems*, 9 (1): 7-24.

Del Mondo, G., Stell, G.S., Claramunt, C. and Thibaud, R., 2010. A graph model for spatio-temporal evolution. *Journal of Universal Computer Science*, 16 (11), 1452-1477.

Shaw, S.L., Yu, H. and Bombom, L.S., 2008. A space-time GIS approach to exploring large individual-based spatiotemporal datasets. *Transactions in GIS*, 12 (4), 425-441.

Smith, B.T., Thompson, R.L., Grams, J.S. and Broyles, C., 2012. Convective modes for significant severe thunderstorms in the contiguous United States. Part I: storm classification and climatology. *Weather Forecasting*, 27 (5), 1114-1135.

Stell, J., et al., 2011. Spatio-temporal evolution as bigraph dynamics. *In*: M.J. Egenhofer, et al., ed. *COSIT 2011*. Belfast, ME: Springer-Verlag, 148-167.

Thibaud, R., Mondo G.D., Garlan T., Mascret, A. and Carpentier, C., 2013. A spatio-temporal graph model for marine dune dynamics analysis and representation. *Transactions in GIS*, 17 (5), 742-762.

Tucker, D.F. and Li, X., 2009. Characteristics of warm season precipitating storms in the Arkansas–Red River basin. *Journal of Geophysical Research*, 114 (D13).

Turdukulov, U.D., Kraak, M.J. and Blok C.A., 2007. Designing a visual environment for exploration of time series of remote sensing data: In search for convective clouds. *Computers & Graphics*, 31 (3), 370-379.

Worboys, M., 2005. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, 19 (1), 1-28.

Yuan, M., 2001. Representing complex geographic phenomena in GIS. *Cartography and Geographic Information Science*, 28 (2), 83-96.

Yuan, M., and Hornsby, S. K., 2007. *Computation and Visualization for Understanding Dynamics in Geographic Domains: A Research Agenda*. Boca Raton: CRC Press.

Zacks, J.M. and Tversky, B., 2001. Event structure in perception and conception. *Psychological Bulletin*, 127 (1), 3-21.

Zahraei, A., Hsu, K., Sorooshian, S., Gourley, J.J., Hong, Y. and Behrangi, A., 2012. Short-term quantitative precipitation forecasting using an object-based approach. *Journal of Hydrology*, 483, 1-15.

3 Life cycle characteristics of warm-season thunderstorms in central United States from 2010 to 2014

3.1 Introduction

Meteorologists have great interests on the climatology of thunderstorms across the world because severe thunderstorms could cause heavy rain, large hail, lightning strikes, and strong winds, which can potentially create damage to lives and property (Han et al. 2008, Hocker et al. 2008a). Weather monitoring systems, such as the Doppler radar, collect data with increasing spatial and temporal resolutions and provide great opportunities for researchers to study convective weather events (Feidas and Cartalis 2005, Hocker and Basara 2008b). For example, a number of researchers have studied the life cycle characteristics of mesoscale convective systems (MCS) using meteorological satellite products. Machado et al. (1998) used GOES-7 ISCCP-B3 satellite data to track the life cycle of deep convective systems (CS) across the United States at both tropical and middle latitudes during 1987 - 88. They mainly used areal overlap to extract the evolution of CS using images of 3-h temporal resolution. Mathon and Laurent (2001) provided an eight-year (June - September, 1989 - 98) climatology of Sahelian MCS using the METEOSAT infrared images with 0.5-h temporal resolution and 5 km spatial resolution. They used both forward and backward areal overlapping on consecutive images to construct whole life cycles of MCS. Moreover, they illustrated dynamic changes and interactions among the life cycles including generation, development, dissipation, merger, split, and combinatorial (merger

and split occurring simultaneously). Morel and Senesi (2002a, 2002b) studied the climatology of MCS life cycles in western European using an automatic cloud-tracking algorithm considering three factors: temperature, area, and size of areal overlap. In the tracking algorithm, they estimated the velocity of cloud systems in order to detect clouds with the size of 1,000 km^2 efficiently. Their algorithm has some advantages over previous tracking algorithms, which only use areal overlapping on consecutive images and are difficult to capture small or fast moving clouds because of low temporal sampling frequency without velocity prediction.

The above climatology studies are all based on meteorological satellite images where the systems are on the scale of more than 5,000 km^2. To track the mesoscale (down to ~20 km^2) life cycle of storm scale, the data must have a much higher spatial and temporal resolution than the above studies, as well as an improved tracking algorithm. As a consequence, a number of researchers have utilized radar-based algorithms to extract, represent, analyze, and predict the life cycles of storm events.

The critical component of a storm tracking algorithm is how to associate the storm cells that are identified over consecutive radar images (Lakshmanan and Smith 2010). There are two major categories of storm tracking algorithms: centroid-based tracking algorithms (Dixon and Wiener 1993, Johnson et al. 1998, Meyer et al. 2013, Zahraei et al. 2013, Liu et al. 2016) and cross-correlation tracking algorithms (Tuttle and Foote 1990, Li et al. 1995). Both tracking algorithms have advantages and disadvantages. While centroid-based tracking algorithms delineate and track single storm cells and provide attributes of storm cells, cross-correlation tracking algorithms can provide more accurate movement speed and direction (Johnson et al. 1998, Wilson et al. 1998, Wilson et al. 2004).

Among the existing centroid-based tracking algorithms, the Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN) (Dixon and Wiener 1993) system developed at the National Center for Atmospheric Research (NCAR) is widely used throughout the world. The TITAN algorithm firstly delineated a single storm cell as a contiguous region where the reflectivity and volume both exceed certain thresholds. Spatial overlap and the Hungarian optimization algorithm were combined to determine whether the storm cells on consecutive images belong to the same storm. However, the previous studies only allow one trajectory when they deal with merger and split situations. In the TITAN algorithm (Dixon and Wiener 1993), when two or more storm cells merge into a single storm cell, only one trajectory is kept and the remainder would be ended. When a single storm cell splits into two or more small storm cells, only one trajectory is kept and the rest would be new storms. While the above treatment of merger and split is relatively easy, it does not represent the complete life cycle and interactions among storm cells (Liu et al. 2016) since it does not include storms that split or merge.

A number of studies have been done on the spatiotemporal characteristics of thunderstorms across the central United States (Changnon 1988a, Changnon 1988b, Tucker and Li 2009). However, very few have focused on the whole life cycle of storm events. The analyses of spatial and temporal characteristics of storm events through the United States are mainly from the National Weather Service (NWS) storm reports contained in the storm data of National Oceanic and Atmospheric Administration (NOAA). However, the data are point features which do not fully represent storm initiation, development, termination, and geographic extent.

Recognizing the need for an automated methodology to extract thunderstorms from large spatiotemporal datasets and analyzing their spatiotemporal characteristics, the goals of this work are to identify thunderstorm life cycles over central United States, reveal seasonal, diurnal, and

spatial patterns of thunderstorms, and test the association between land cover properties and thunderstorm features especially in urban and rural areas. Geographic Information Systems (GIS) has been widely applied to meteorological research (Yuan 2001, McIntosh and Yuan 2005, Liu et al. 2016). Because of the spatial focus of this study, GIS is used to identify, represent, query, and analyze thunderstorm life cycles. The first task is to develop a thunderstorm GIS database storing their whole life cycles where directed graph representation and algorithms are explored to characterize the thunderstorms. The second task is to quantify the spatiotemporal patterns of the thunderstorms using GIS query, spatial analyses, and spatial statistics.

This study intends to illustrate how innovative GIS representations and analyses can be used to characterize the spatial and temporal patterns of thunderstorm life cycles. The three major datasets, methodologies, and GIS representations are described in Section 3.2. A number of spatial and temporal analyses on thunderstorm life cycles are presented in Section 3.3. Summary of the research and possible future work are provided in Section 3.4.

3.2 Data and methodology

3.2.1 Radar reflectivity data

The United States NWS maintains the Next Generation Radar (NEXRAD) program covering almost the whole country to monitor precipitation and other meteorological and hydrological phenomena (Tucker and Li 2009). This is a network of S-band (10 cm) Weather Surveillance 1988 Doppler radars (Choi et al. 2009), which has been recently upgraded to dual-polarization radar. The radar data used in this study are the final reflectivity product (N0R) from Iowa Environmental Mesonet (IEM). The IEM interpolates the base reflectivity to a 1 km grid every 5 minutes, and the archived datasets could be accessed via the IEM Geographic Information

System (GIS) data service (http://mesonet.agron.iastate.edu/docs/nexrad_composites/) in PNG or GeoTIFF format. Precipitation can be estimated from radar returns based on a Z-R relationship, and the radar reflectivity data have been applied to a number of different applications including storm identification and nowcasting (Dixon and Wiener 1993, Johnson et al. 1998, Han et al. 2008, Han et al. 2009), climatology (Chen et al. 2012, Lock and Houston 2015), and urbanization impacts on precipitation (Ashley et al. 2012, Perryman and Dixon 2013).

3.2.2 Lightning data

The second dataset used in this study is cloud-to-ground lightning data from the United States National Lightning Detection Network (NLDN, Cummins and Murphy 2009), which is produced by the Vaisala Corporation (http://www.vaisala.com). Cloud-to-ground lightning point data are used to determine whether the precipitation is associated with a thunderstorm. The precipitation cluster is considered a thunderstorm if at least a lightning strike occurs during its life cycle.

The radar reflectivity data and cloud-to-ground lightning data used in this research cover 32.5°N – 40.5°N and 93.5°W – 103.5°W (the red dashed boundary in Figure 3.1). The study domain mainly covers the states of Kansas, Oklahoma, and northern Texas. The data span a period of five years from April 1, 2010 to September 30, 2014, and only include warm seasons (April – September) each year. Severe thunderstorms are common features in the study area, so understanding their life cycle characteristics is significant for weather forecasting, disaster management, and hydrological management (Whitehall et al. 2015).

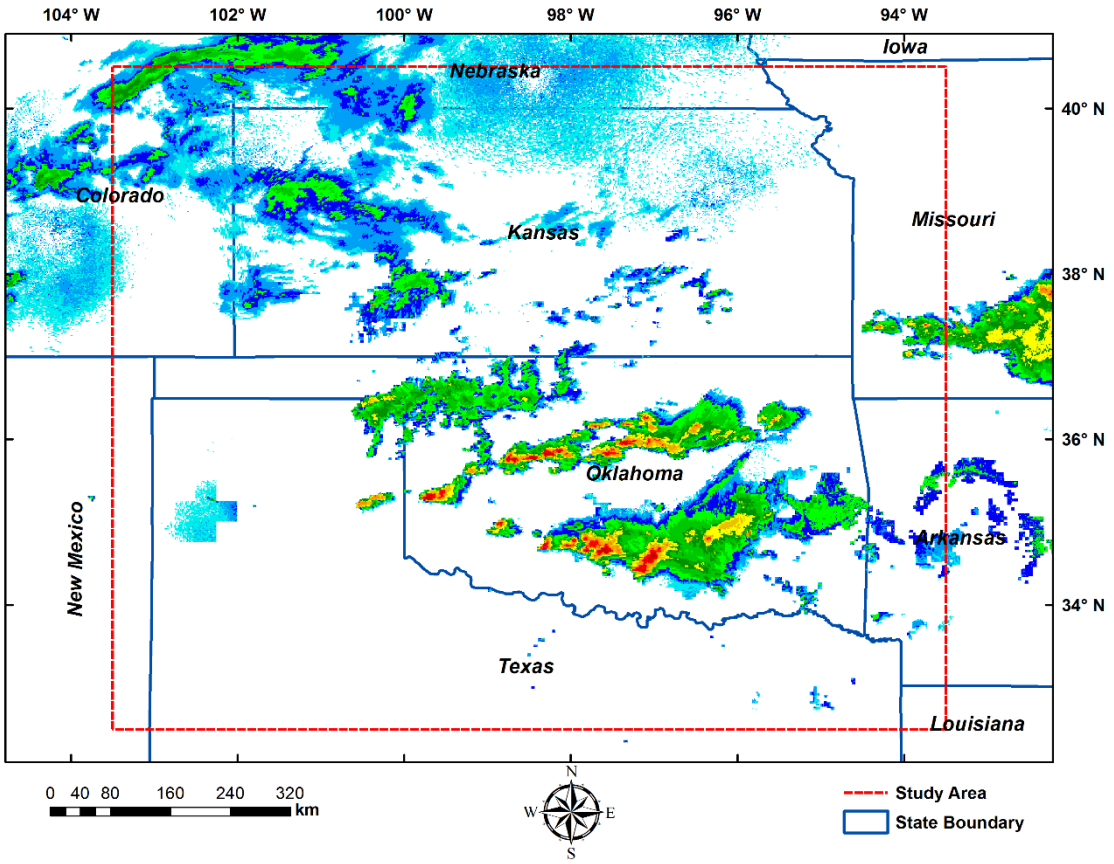


Figure 3.1 The study area (highlighted in red dashed rectangle) mainly covers the state of Kansas, Oklahoma, and northern Texas. A radar reflectivity image at 8:00, 24 April 2011 UTC is shown as an example.

3.2.3 Land cover data

The latest 2011 version of the National Land Cover Dataset (NLCD) is used to study the relationship between land cover types and thunderstorm occurrence. NLCD 2011 is primarily based on the unsupervised classification of 2011 Landsat satellite data, which provides information on water bodies, vegetation, and developed lands at a spatial resolution of 30 meters. To link thunderstorm occurrence to the main land cover types in the study area, the original NLCD land cover classes are generalized and reclassified into seven major types including water (1.2%), barren (0.2%), grasses (42.4%), wetlands (1.0%), urban areas (3.2%), forests (11.5%),

and crops (40.4%) (Figure 3.2). The study area is mainly covered by grasses, crops, and forests. The NLCD data is reprojected to the same coordinate system (GCS_North_American_1983 spatial reference system) as the radar reflectivity data, and is resampled to 1 km spatial resolution using a majority resampling method to match the radar data.

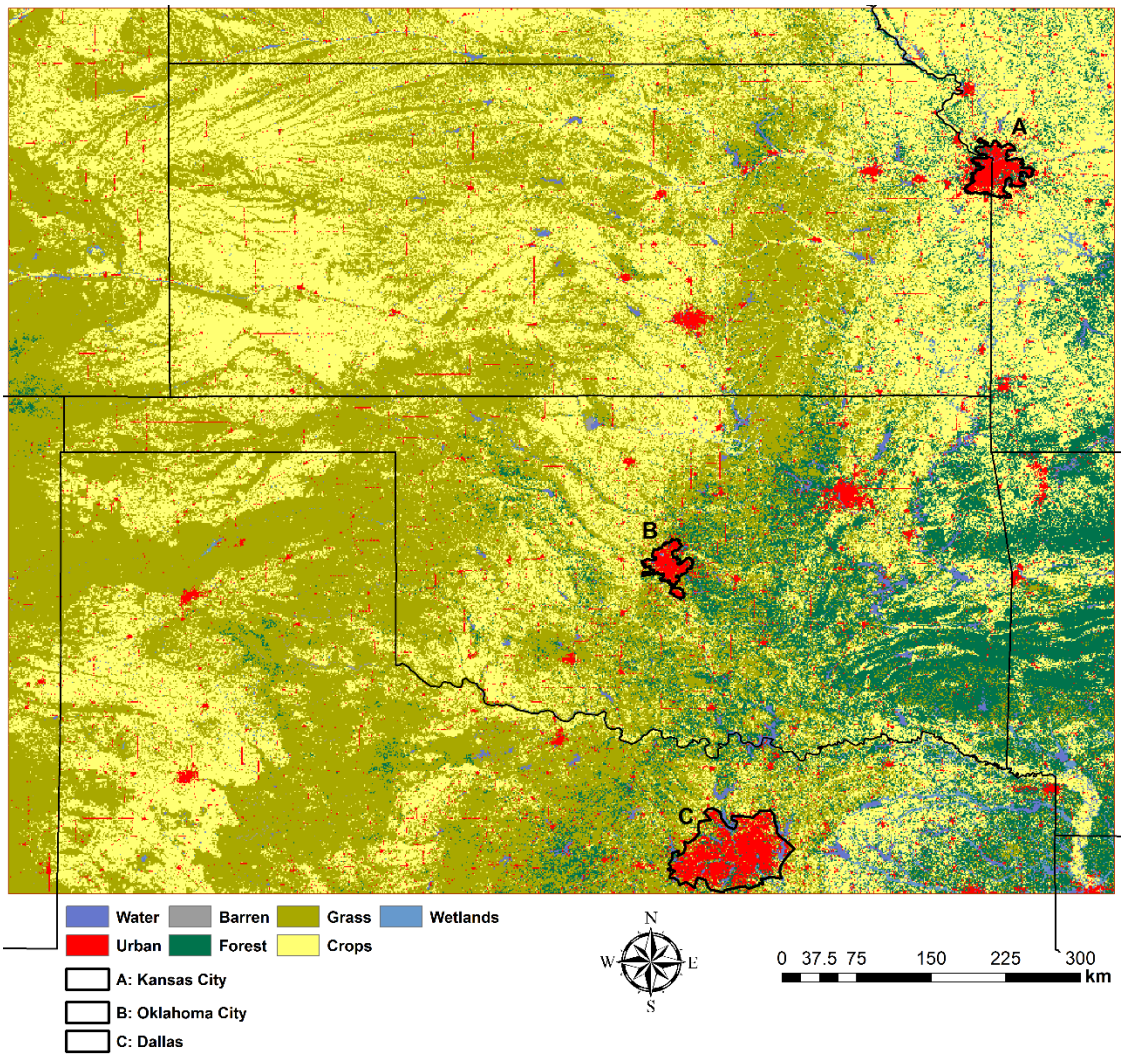


Figure 3.2 The land cover types in the study area after reclassifying 2011 NLCD into seven main types.

3.2.4 Extraction of thunderstorm life cycles

First, a storm cell in a radar image is delineated as a contiguous region where both the reflectivity and area are greater than or equal to certain thresholds. Based on the sensitivity analysis in Liu et al. (2016), the reflectivity range should be between 30 and 40 dBZ, and the area range should be between 20 and 30 km^2 for convective storms. Because the focus is on substantial thunderstorms, the reflectivity and area thresholds were chosen as 35 dBZ and 20 km^2, respectively. A component-labeling algorithm (Haralock and Shapiro 1991) was used to extract storm cells from individual radar images.

After extracting storm cells in each radar image, the critical and challenging step is to associate the storm cells on consecutive radar images to extract the whole life cycle of a storm. In this study, storm life cycle extraction used an improved centroid-based storm tracking algorithm developed by Liu et al. (2016) which considered the spatial overlap, centroid distance, and movement direction of storm cells simultaneously. A storm centroid is the reflectivity-weighted mean position of the radar pixels in the storm cell. If a storm cell on the current image and a storm cell on the next image have sufficient spatial overlap and are within a reasonable distance and movement direction, the two storm cells are considered in the same storm trajectory. The sensitivity analysis (Liu et al. 2016) on spatial overlap indicates a reasonable range of spatial overlap is between 0.4 and 0.8. In this study, we set 0.6 as the threshold of spatial overlap. Based on the average storm movement speed and sampling frequency of the radar images, the threshold for centroid distance was set to 10 km. Finally, the angle between the predicted movement direction of a storm cell and the direction from the centroid of the storm cell on the current radar image to the centroid of a possible matching storm cell on the next radar image should be less than 90° in order to keep only realistic storm movement (Liu et al. 2016).

Cloud-to-ground lightning data determines whether extracted storms are thunderstorms. The lightning point data were overlaid with storm cell polygons. If lightning occurs in any storm cells of a storm, the storm is counted as a thunderstorm. This method, however, may omit some thunderstorms because not all thunderstorms generate cloud-to-ground lightning (Lock and Houston 2015).

An application developed via MATLAB was utilized to process, extract, represent, and analyze the thunderstorms. A flow chart illustrating the key steps is shown in Figure 3.2. The radar reflectivity images are preprocessed by creating a subset of radar data over the study area and converting the data format. Afterwards, three most important tasks are carried out and include storm cell delineation, storm life cycle extraction, and thunderstorm identification. A number of properties are calculated for the thunderstorms at both cell level and life cycle level. At the cell level, we calculated the total reflectivity and area. Thunderstorm cells are approximated by ellipses, so we also calculated the orientation, major and minor axis of the ellipses. At the life cycle level, we calculated the duration, mean movement speed and direction during the life span of a thunderstorm. For the thunderstorm climatology, we also calculate statistics of these properties.

3.2.5 Directed graph representation of thunderstorms in GIS

A directed spatiotemporal graph model (Figure 3.4) is used to represent the life cycle of a thunderstorm including its initiation, development, interactions among storm cells, and termination. Nodes and directed edges are the two basic components of our directed graph model. Reflectivity-weighted centroid, which captures the most intense precipitation in a storm cell, is used as a node to represent the cell. Edges in the graph model represent the linkages among thunderstorm cells at two consecutive radar images and direction denotes time sequence.

Figure 3.4 shows an example of a thunderstorm extracted from five radar images from time $\mathbf{t_1}$ to $\mathbf{t_5}$ with a duration of 20 minutes. We could see there are a total of eight thunderstorm cells in its life cycle. There are five filiation relationships (initiation, continuity, split, merger, and termination) in the figure. The thunderstorm initiates from thunderstorm cell **a** at time $\mathbf{t_1}$, then develops into cell **b** at time $\mathbf{t_2}$. Cell **b** and **c** merge into cell **d** at time $\mathbf{t_3}$. At time $\mathbf{t_4}$, cell **d** splits into three cells **e**, **f**, and **g**, and finally this thunderstorm terminates at time $\mathbf{t_5}$ as cell **h**. Using the directed spatiotemporal graph model, the initiation, termination, split, and mergers can be quantified, which will be discussed in Section 3.3.

Many details are contained in the directed graph representation. To begin with the simple storm statistics such as storm duration and the mean speed and direction of storm movement, the maximum reflectivity path (Liu et al. 2016), which is based on the classic Dijkstra graph shortest path algorithm, is applied to the database to produce a polyline (the red line in Figure 3.4). When the density of thunderstorm trajectories is calculated, the maximum reflectivity path is extracted to calculate the density using the polylines. Compared to previous studies (Dixon and Wiener 1993, Johnson et al. 1998, Lakshmanan and Smith 2010, Zahraei et al. 2013), our approaches study both the interactions among thunderstorm cells within its life cycle (the directed graph representation) and generalized trajectory from a simplified life cycle (the polyline representation).

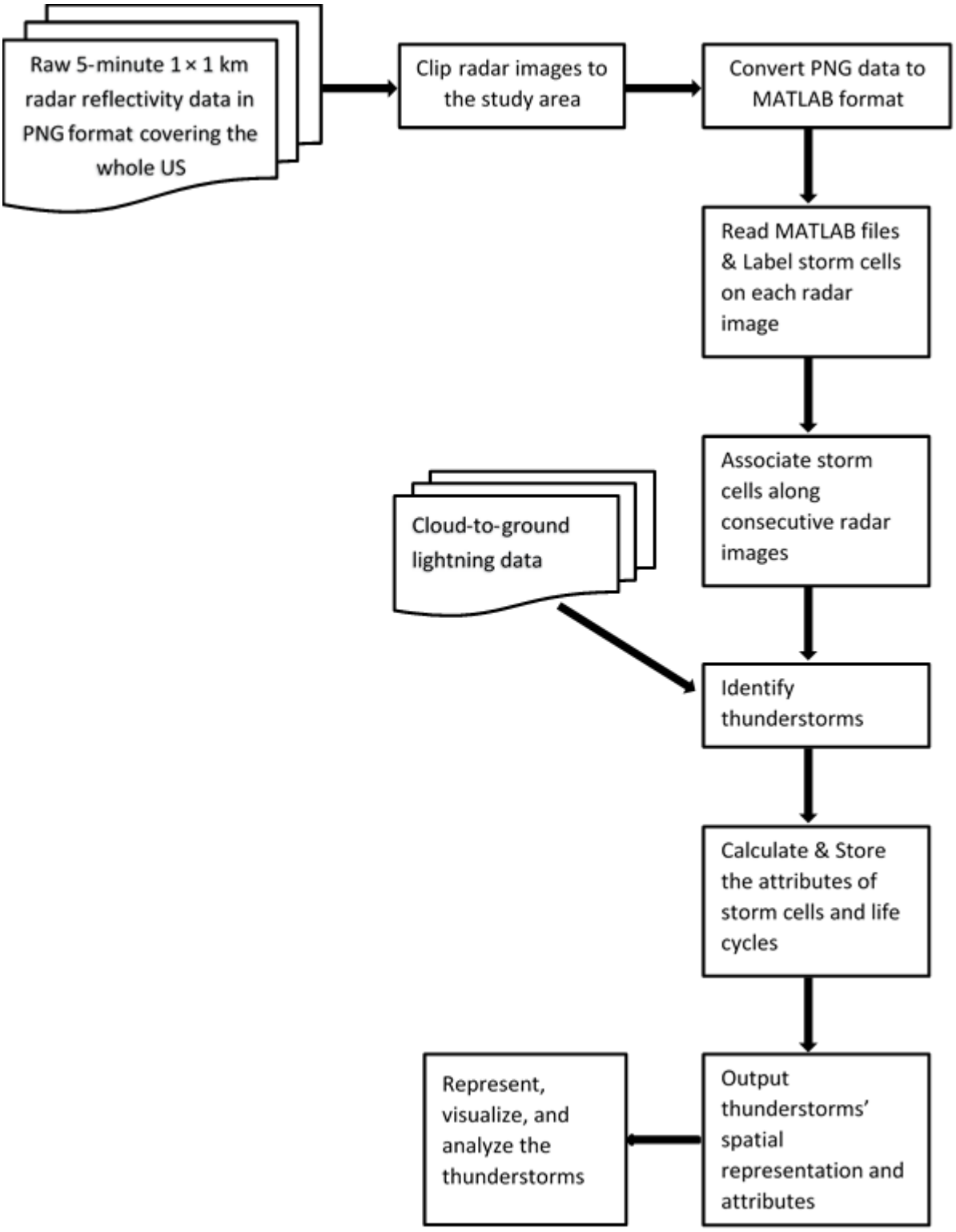


Figure 3.3 The workflow of identifying, representing, and analyzing thunderstorms.

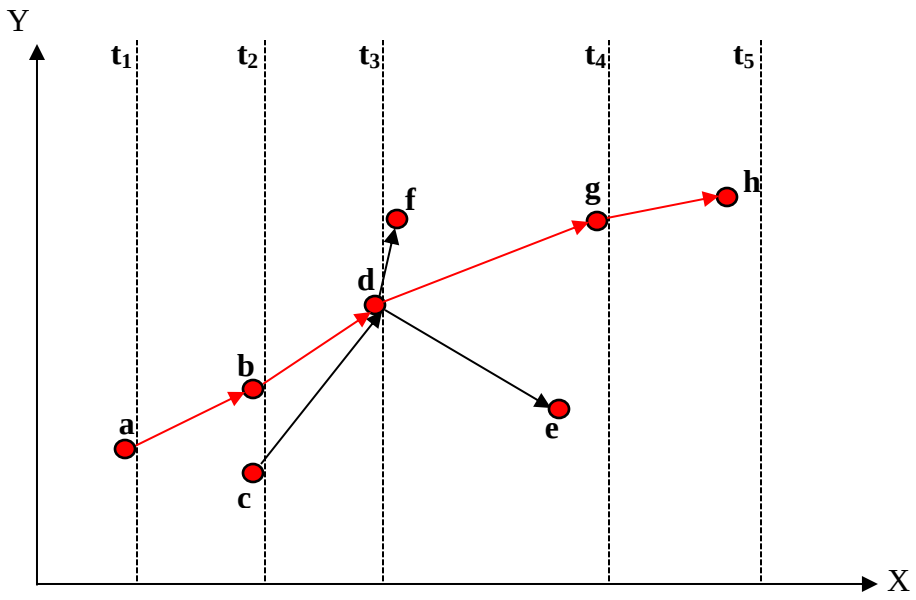


Figure 3.4 A graph representation showing the life cycle of a thunderstorm. The red polyline is the simplification of the graph representation using the maximum reflectivity path.

3.3 Results and discussion

The results of the thunderstorm climatology highlight a number of different spatial and temporal characteristics found during the warm seasons from 2010 to 2014. An important emphasis is to illustrate how GIS representations and analyses can be used for climatological research in meteorological communities.

3.3.1 Temporal characteristics

A total of 130,097 thunderstorms were identified across the study area during the five-year time period. Annual and monthly numbers of thunderstorms are shown in Figure 3.5a and Figure 3.5b. On average there are 26,019 per year and the annual number of thunderstorms does not vary much with a coefficient of variation of 5.4% (Figure 3.5a). While 2010 has the most number of thunderstorms (27,824) during the study period, 2011 is the most inactive year (23,942). 2010 is 16% higher than the number in 2011. The number of thunderstorms in 2012 (26,164), 2013

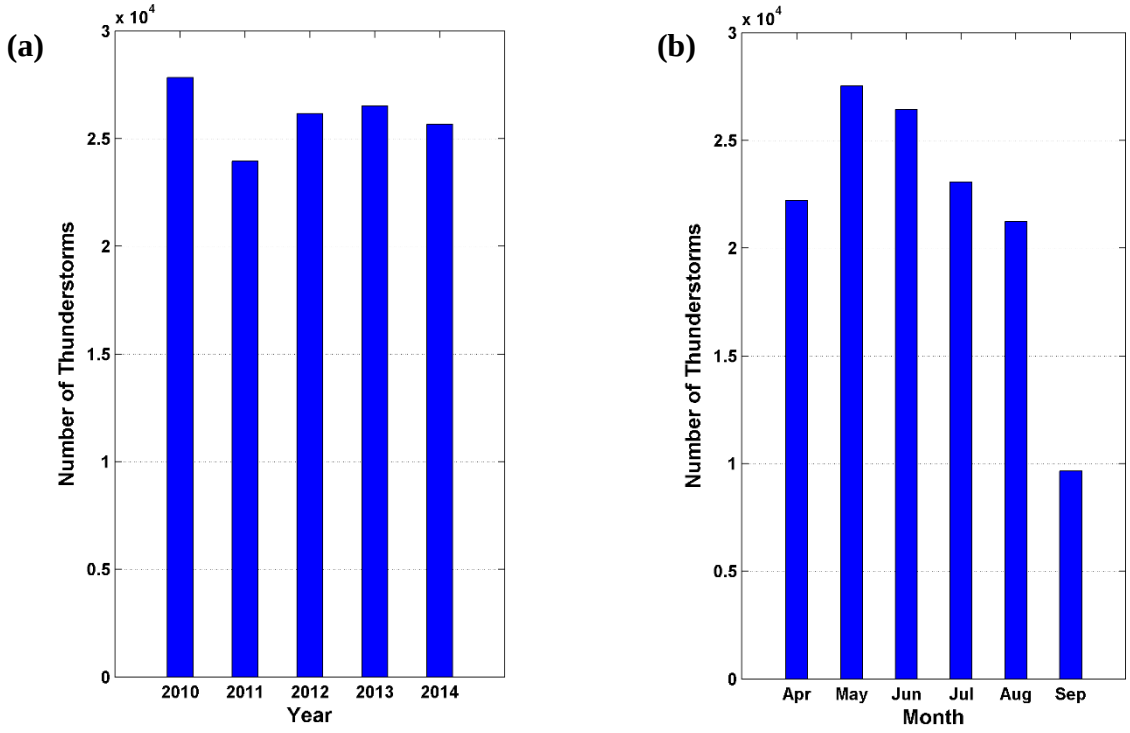


Figure 3.5 (a) Annual and (b) monthly thunderstorm occurrences in the study area during warm seasons (April to September) from 2010 to 2014.

(26,518), 2014 (25,649) does not have a great difference, which is around the mean number of thunderstorms (26,019).

Figure 3.5b shows dramatic variation in monthly numbers of thunderstorms. The number of thunderstorms increases dramatically from April (22,209) to the peak in May (27,528), a 23% increase. Then thunderstorms decrease quickly from June to September. May to July accounts for approximately 59.2% of the total thunderstorms during the warm seasons, and they are the three most active thunderstorm months.

The histogram of thunderstorm durations (Figure 3.6a) shows an exponential decay distribution consistent with Hocker and Basara (2008b) and Novo et al. (2013). The average duration of the thunderstorms is 23.1 mins, and 65.8% of the thunderstorms have a duration between 5 and 20 minutes. 8,914 out of 130,097 thunderstorms (6.9%) lasts more than one hour.

If a thunderstorm extends outside of the study area, its duration time would be shortened. As a result, thunderstorm durations are underestimated for those that end outside of the domain. Figure 3.6b shows the average thunderstorm durations by month (April: 21.5 mins, May: 23.4 mins, June: 22.1 mins, July: 23.7 mins, August: 24.4 mins, September: 24.6 mins). Moreover, very long lived thunderstorms, which last over 3 hours, are more frequent between July and September than in other months. Robinson and Easterling (1988) pointed out that thunderstorms had a longer lifespan in summer than in spring in central United States. In our study area, the average duration in summer (June to August) is 23.2 mins, while the average duration in spring months (April and May) is 22.7 mins, which verifies that long lived thunderstorms do favor summer months though long lived thunderstorms consist of only a small portion of total thunderstorms (Figure 3.6a).

Figure 3.7 demonstrates the histograms of thunderstorm initiation and termination time. The most frequent time for thunderstorm initiation occurs from 2100 to 0000 UTC which is in the early evening at local time and favors single cell and multicellular thunderstorms (Tucker and Li 2009). Thunderstorm termination is most common from 2100 to 0300 UTC.

Figure 3.8 shows the monthly rose diagrams of thunderstorm movement speed and direction. The monthly average movement speeds are 63.2 km/h (April), 53.8 km/h (May), 48.1 km/h (June), 38.1 km/h (July), 42.5 km/h (August), and 47 km/h (September), which are similar to the climatological analysis in North Dakota in Mohee and Miller (2010). For movement direction, thunderstorms are most common between 45 and 90 degrees with an average thunderstorm vector from the southwest to northeast during the warm seasons from 2010 to 2014.

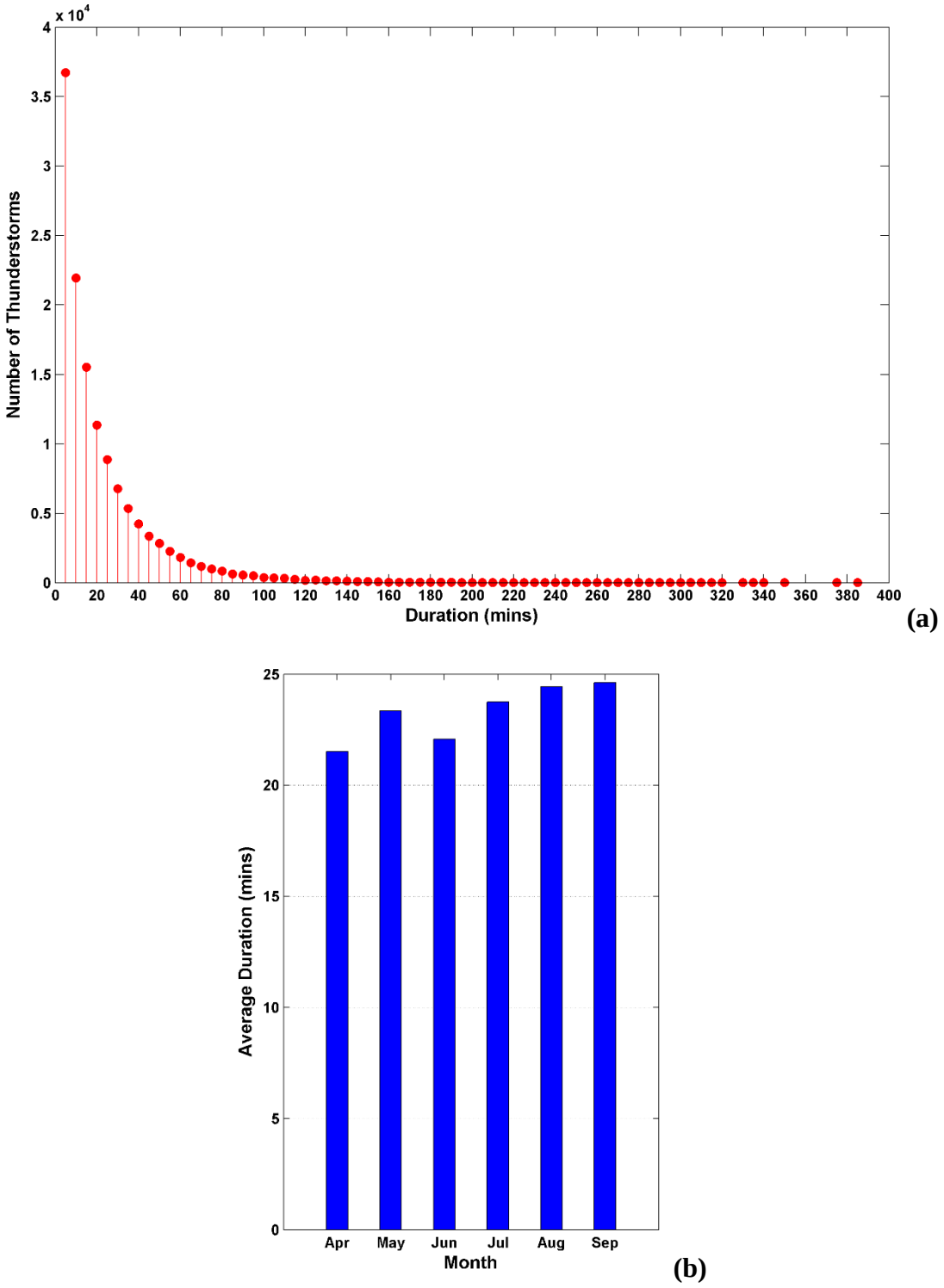


Figure 3.6 (a) Histogram of thunderstorm durations; (b) average duration of thunderstorms by month.

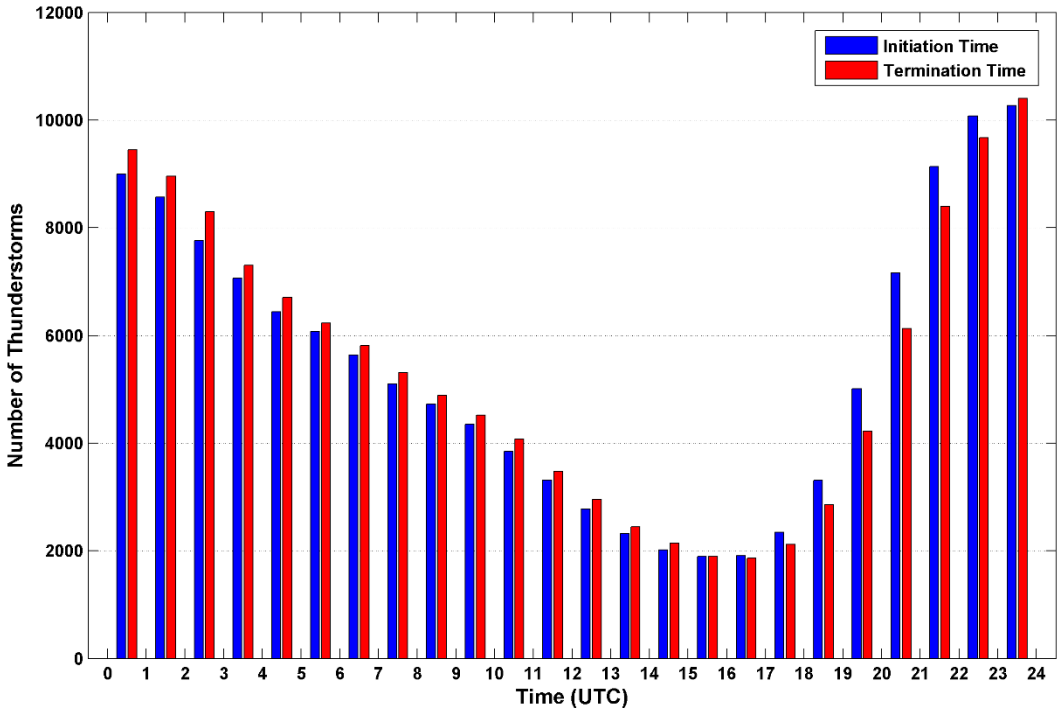
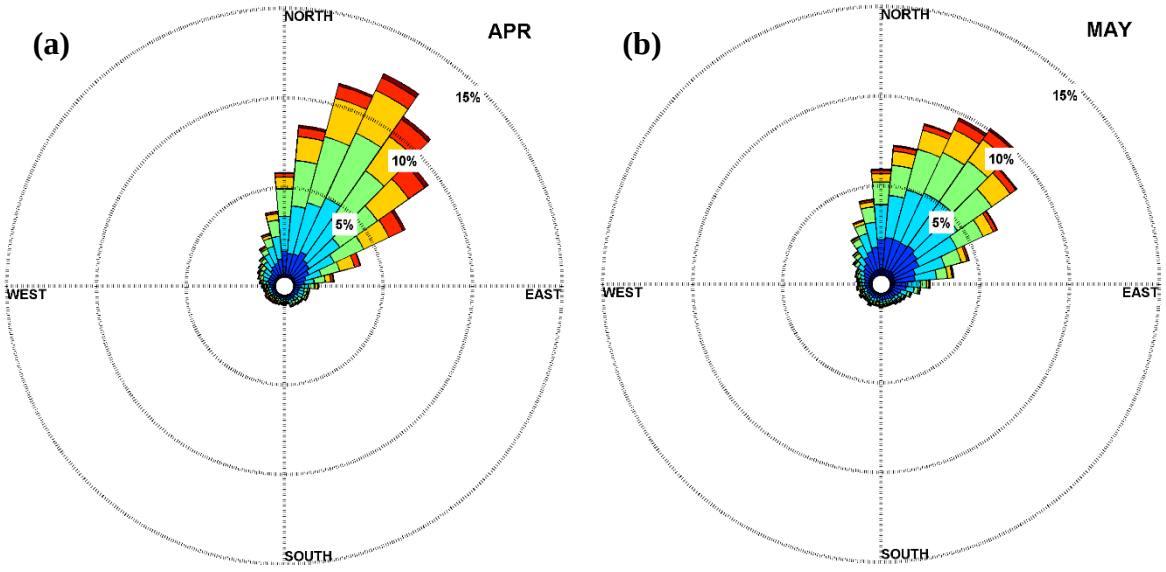


Figure 3.7 Histogram of thunderstorm initiation (blue) and termination (red) time during a day.



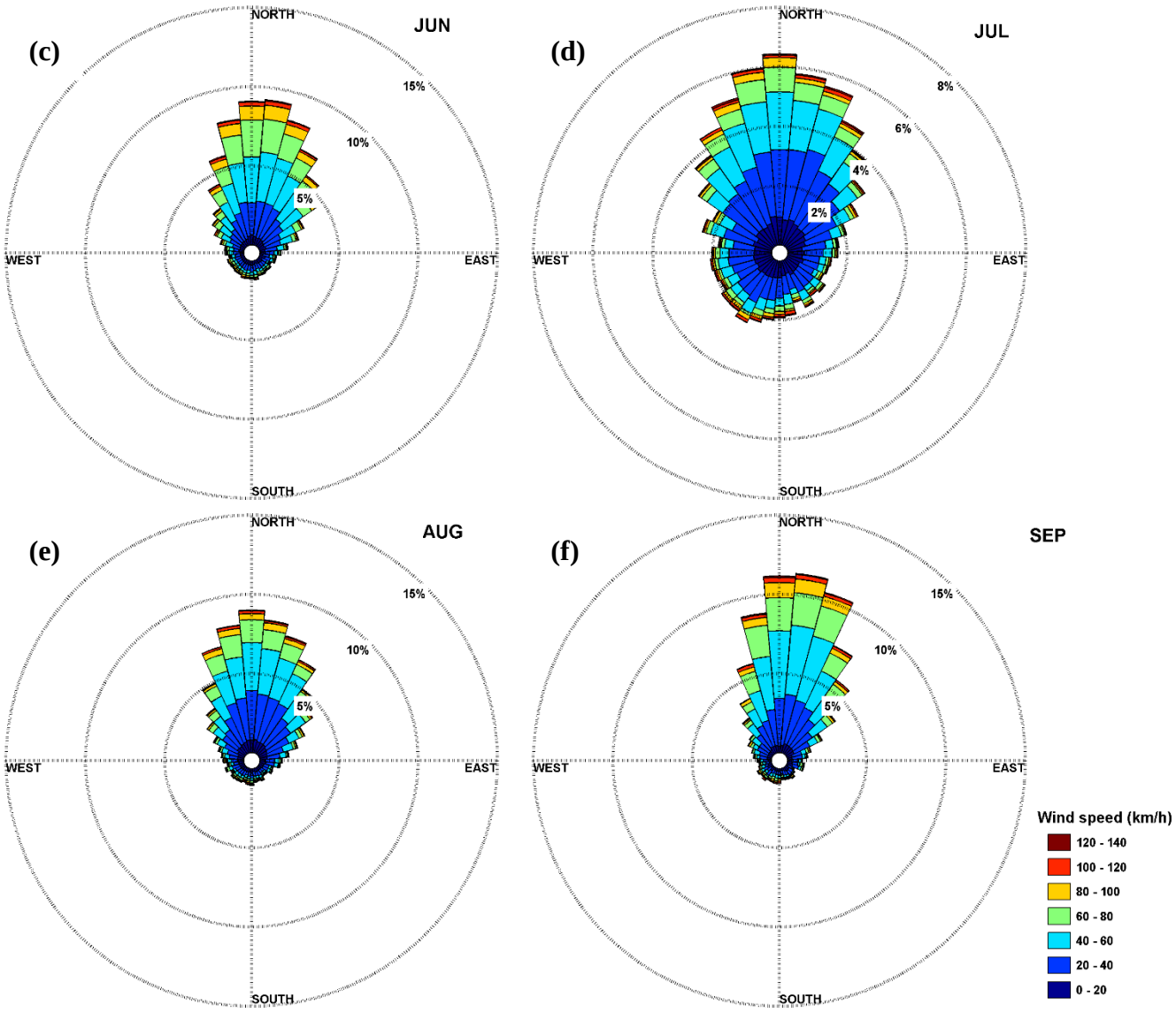


Figure 3.8 Rose diagrams of thunderstorm movement distributions for (a) April through (f) September.

3.3.2 Spatial characteristics

Using maximum reflectivity path, a density raster grid of thunderstorm trajectories with a cell size of 0.01 degrees (~1 km) was generated with a search radius of 0.5 degrees. The thunderstorms were divided into cumulative month periods to quantify the spatial and temporal variability of thunderstorm track from 2010 to 2014 (Figure 3.9). The spatial frequency analyses could highlight a number of hot spots across the study area during the limited 5-year period.

Figure 3.9 shows thunderstorm density in each month from April to September, and all the months during the five-year period.

In April (Figure 3.9a), thunderstorm hot spots are mainly located in the east 1/3 of the study area, from central Kansas and Oklahoma to the east boundary. May has the peak of thunderstorm occurrences and the thunderstorms mainly concentrate on the east half of the study area, in southeastern Kansas, north-central Oklahoma and northeast Texas (Figure 3.9b). In June, there is a pronounced density decrease in thunderstorm occurrence (Figure 3.9c). Thunderstorms mainly occur in the western and northern quarters of the study area, which is different from those in April and May.

Thunderstorm occurrence further decreases gradually through July, August, and September (Figure 3.9d-f). In July, thunderstorm hot spots are scattered throughout the study area with some concentration in the state of Kansas and Oklahoma. In August, thunderstorms concentrate in the middle 1/3 of the study area covering southern Kansas, northern Oklahoma, and the most northern part of Texas. Thunderstorms concentrate along the border region between Kansas and Missouri and central Oklahoma in September.

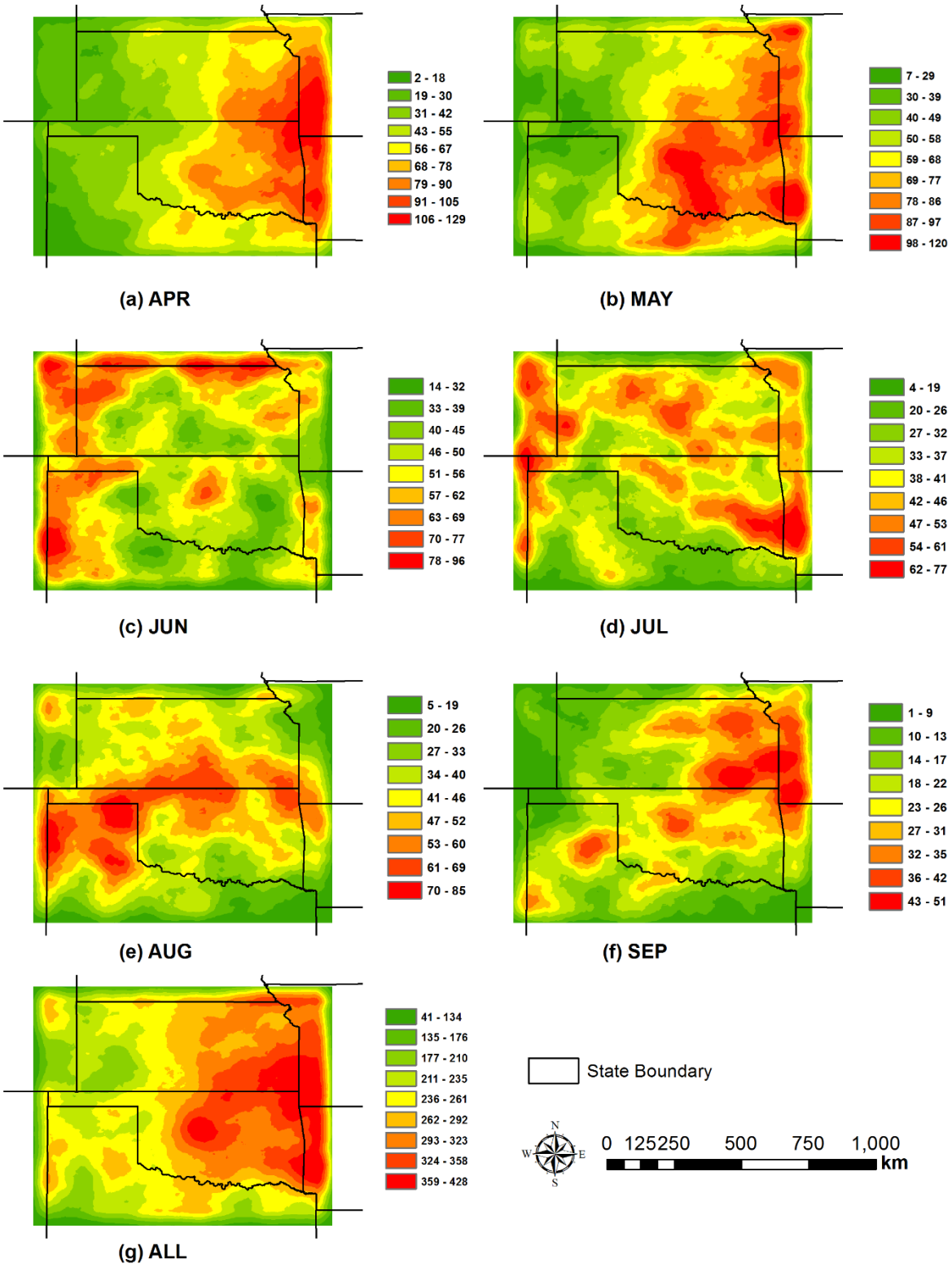


Figure 3.9 Thunderstorm track density (km/km^2) from April (a) to September (f), and all the months (g).

Cumulative density in the six months (Figure 3.9g) shows that major thunderstorm activity occurs in the east half of the study area, especially centered at the border among Kansas, Missouri, Oklahoma, and Arkansas. For the three major states (Kansas, Oklahoma, and northern Texas) in the study area, the thunderstorm activities are mainly in Kansas (central to the most east) and Oklahoma (north-central, northeastern and southeastern). Based on the land cover types in Figure 3.2, the land cover type of highest thunderstorm density is forest covering from central Oklahoma to the state of Arkansas in the study area. The second land cover type with active thunderstorms is crops, which are the main land cover stretching from central Kansas to the east boundary of the study area. The area in and around Oklahoma City is a hot spot for thunderstorm activity. The relationship between thunderstorm track and land cover types is discussed in Section 3.3.4, and a comparison between thunderstorm track in urban and rural areas is provided in Section 3.3.5.

For the locations of thunderstorm initiation and termination, we used the starting and ending points of a thunderstorm's trajectory, i.e. the maximum reflectivity path. The initiation and termination point density maps are shown in Figure 3.10 for the 5-year thunderstorms to quantify active areas of thunderstorm occurrences. Figure 3.10 demonstrates the kernel point density analyses determining the concentration of points within a search radius of 0.5 degrees of each initiation and termination point. The spatial distributions of initiation (Figure 3.10a) and termination density (Figure 3.10b) are quite similar with the greatest density on both maps found along the border between Oklahoma and Arkansas, and in south part of the border between Kansas and Missouri. In Kansas, initiation and termination hot spots are mainly located in the southeast corner of Kansas. In Oklahoma, thunderstorm initiations and terminations are most concentrated in the east half of Oklahoma. In Texas, initiations and terminations are concentrated

mainly in northern Texas and at the border between northern Texas and New Mexico. By overlaying the initiation and termination density maps with the land cover map (Figure 3.2), we find that forest and crop land cover types contain most initiation and termination hot spots located in the border among Kansas, Missouri, Oklahoma, and Arkansas. Urban areas also have higher initiation and termination density.

3.3.3 Thunderstorm cell split and merger

Thunderstorm cells interact with each other in the process of storm development and the interaction is a significant factor influencing storm evolution (Lee et al. 2006a, Lee et al. 2006b). For example, two or more small thunderstorm cells may merge into a large thunderstorm cell, and a large thunderstorm cell may split into a number of small thunderstorm cells. The directed graph representation of thunderstorm life cycles provides an opportunity to study their split and merger characteristics. In a directed graph, if the number of in-degree of a node (a thunderstorm cell) is greater than 1, a merger occurs to the thunderstorm cell. For example, the in-degree of thunderstorm cell **d** is 2, because cell **b** and **c** merge into cell **d** (Figure 3.4). If the number of out-degree of a node is greater than 1, then a split occurs to the thunderstorm cell. For example, the out-degree of thunderstorm cell **d** is 3, because it splits into cell **f**, **g**, and **e** (Figure 3.4). Based on the in-degree and out-degree calculated at each thunderstorm cell, the split and merger density maps (a raster grid of cell size of 0.01 degrees) with a search distance of 0.5 degrees are shown in Figure 3.11.

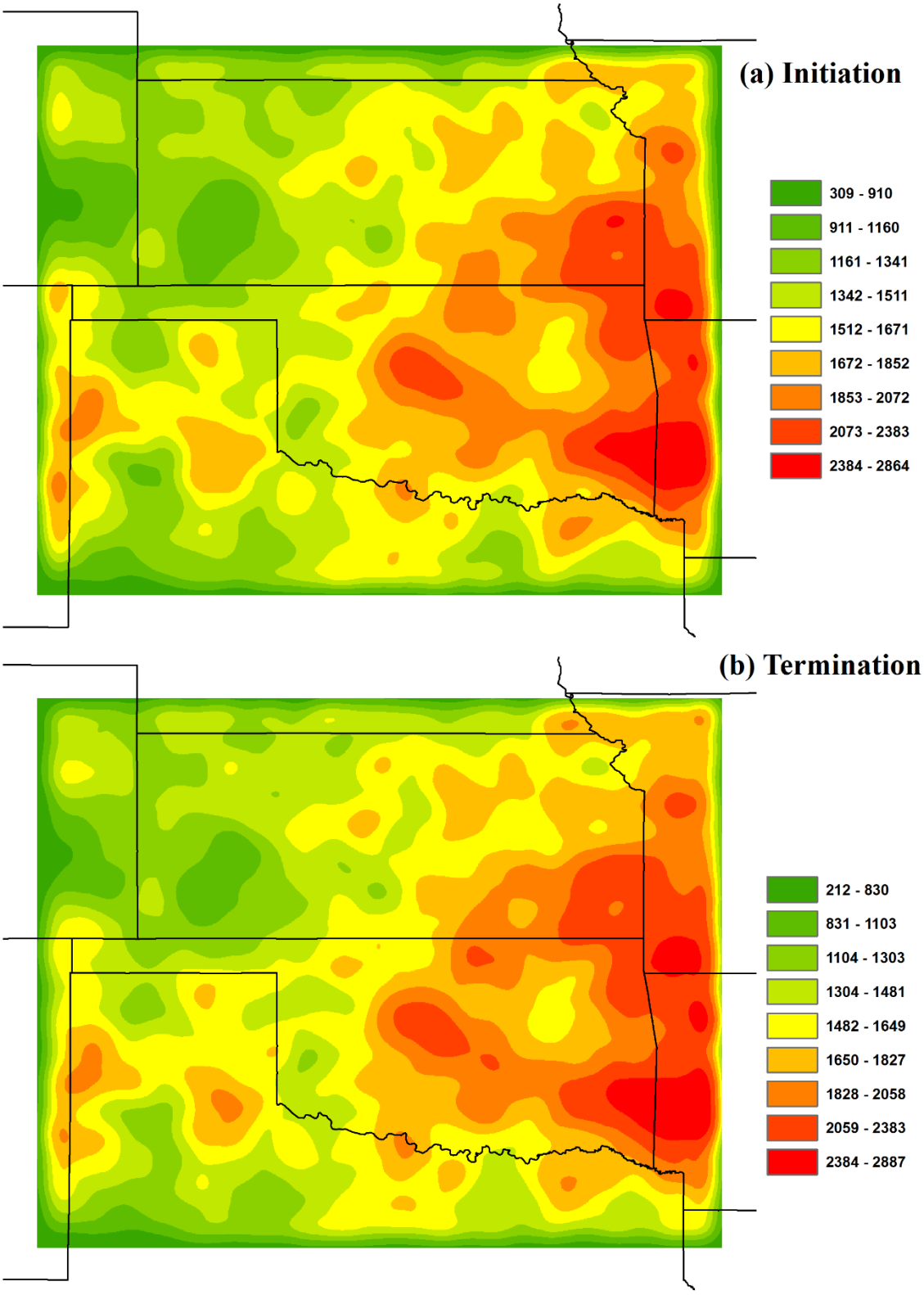


Figure 3.10 Thunderstorms initiation (a) and termination (b) density (count/km^2) maps.

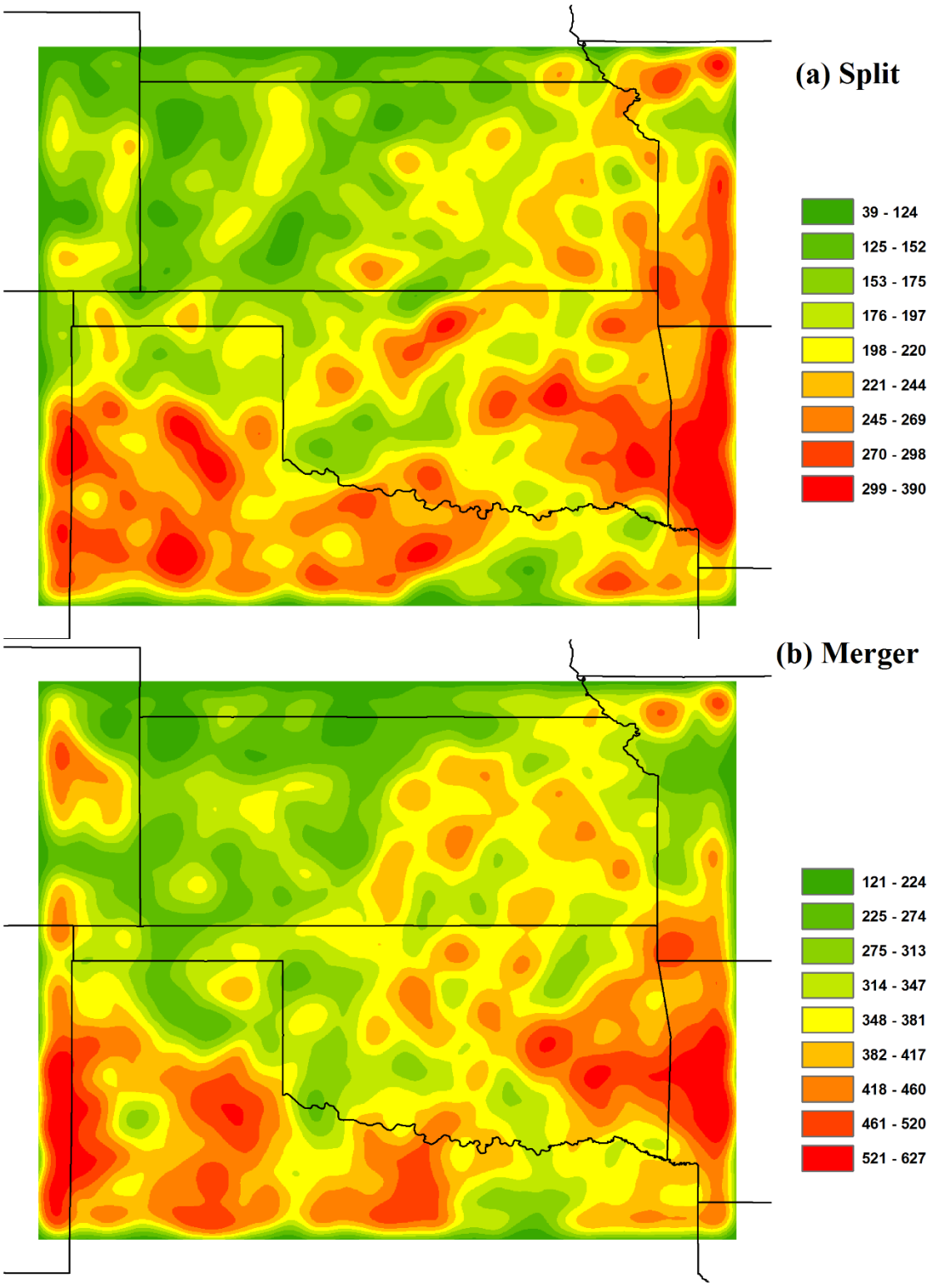


Figure 3.11 Thunderstorm cell split (a) and merger (b) density (count/km^2) maps.

Among all the thunderstorms (130,097), about 13% (17,295) of the thunderstorms have splits, and 22% (29,837) of the thunderstorms have mergers in their life cycles. Compared to splits, there are more mergers existing in the life spans of thunderstorm events. For splits, most of thunderstorm cells only split into two or three smaller thunderstorm cells. The largest split density is around 400 count/km^2, but the greatest merger density reaches around 600 count/km^2. Most thunderstorm cells have only two or three thunderstorm cells merge into a large thunderstorm cell. The average size of thunderstorm cells with splits is 226 km^2, and the average size of thunderstorm cells after mergers is 247 km^2.

As for the spatial distribution of splits and mergers (Figure 3.11), their hot spots have some similarities. They are mainly located along the border between Oklahoma and Arkansas and the border between Kansas and Missouri, in central Oklahoma, and northern Texas. When comparing Figure 3.10 and Figure 3.11, the biggest difference is in the panhandle area in northern Texas. There are not many initiation and termination hot spots in the area. But splits and mergers occur a lot in the region. This means the thunderstorms do not initiate or terminate often in the northern Texas compared to other states in the study area. However, there are also a lot of splits and mergers occurring in Texas during the thunderstorm life cycles. In Kansas and Oklahoma, split and merger hot spots have similar spatial patterns seen in initiation and termination in the eastern part of the study area.

For land cover types (Figure 3.2), we see that splits and mergers mainly occur in forest areas where they have the highest density. This is the case in the states of Oklahoma, Arkansas, and Missouri. For Texas and Kansas, splits and mergers mainly occur in crop and grass areas. Overall, thunderstorm initiation, termination, split, and merger favor the forest and crop land

cover types in the study area. In the next section, we present a statistical analysis on the relationship between thunderstorm tracks and land cover types.

3.3.4 Relationship between thunderstorm occurrences and land cover types

Many factors could affect the locations of thunderstorms' initiation, termination, split, and merger such as the dry line from western Oklahoma northward through western Kansas over the Great Plains in this study area (Owen 1966) and topographic effects (Chen et al. 2012, Wang et al. 2014). In this study, the relationships between thunderstorm track, initiation, termination, split, merger density and seven major land cover types is examined quantitatively (Table 3.1). The densities were obtained by dividing the results in Figure 3.9g, Figure 3.10, and Figure 3.11 by the area of each land cover type. From Table 3.1, thunderstorm track, initiation, and termination have the same rankings over the seven major land cover types in the study area. Forests, urban areas, and crops are the top three land cover types favoring thunderstorm events, and grasses have the lowest ranking. The highest thunderstorm tracks in forest land cover may be because dense forests have greater ability to store and release moisture, which is likely to increase aerodynamic roughness values (Gambill and Mecikalski 2011, Wang et al. 2014). Most of the dense forests are in the mountainous areas of the Ozark National Forest and Ouachita National Forest in the study area. Meanwhile, there are also thunderstorm hot spots in Wichita Mountains and Ouachita Mountains located in Oklahoma. Tucker and Li (2009) also found that mountainous areas had more storms than flatter areas. For urban areas, a strong urban heat island can influence vertical mixing, raise planetary boundary layer height, and weaken the capped inversion intensity, which are conductive to the development of convection (Shepherd et al. 2002, Shepherd 2005, Ashley et al. 2012). Niyogi et al. (2011) also used the radar data to verify that

urban areas alter the initiation and intensity of thunderstorms due to land surface heterogeneity which favors the convective initiation or preconvection (Holt et al. 2006).

The split and merger densities have different ranking over the seven major land cover types. Forests, crops, and grasses are the favorite land cover types triggering splits and mergers during thunderstorm life cycles. The split and merger densities in urban areas are not very high.

Table 3.1 The densities of thunderstorm track (km/km^2), initiation (count/km^2), termination (count/km^2), split (count/km^2), and merger (count/km^2) over seven major land cover types.

Land cover	Track	Initiation	Termination	Split	Merger
Forests	303	1,867	1,904	242	404
Urban	282	1,679	1,689	208	350
Crops	271	1,634	1,629	228	390
Water	266	1,624	1,622	215	358
Barren	264	1,603	1,610	201	342
Wetlands	259	1,549	1,553	202	356
Grass	250	1,504	1,478	220	368

3.3.5 Comparison of thunderstorm occurrences between urban and rural areas

A number of studies indicated that urban areas could change local climate because of higher heat content, increased surface roughness, and boundary layer instability associated with urbans (Chase et al. 2000, Feddema et al. 2005, Shepherd 2005). In this section, our research examines whether urban areas augment warm-season thunderstorm activities by comparing thunderstorm tracks between urban and rural areas. Three big cities, Kansas City, Oklahoma City, and Dallas (Figure 3.2) are in the study area were chosen. We identified urban areas that contain four NLCD land cover classifications including Developed, Open Space (21), Developed, Low Intensity (22), Developed, Medium Intensity (23), and Developed, High Intensity (24). This produces large urban polygons for each of the metropolitan areas. Those urban polygons also include some non-urban land cover types such as grasses and water due to their containment in the larger urban

polygon. After delineating the three urban polygons, rural areas were delineated as the buffers of 10 km (Kansas City), 10 km (Oklahoma City), and 20 km (Dallas) surrounding the urban polygons. Table 3.2 shows the statistical significance for the mean differences of thunderstorm tracks (Figure 3.9g) and initiations (Figure 3.10a) between the urban and rural areas around the three cities using the *t*-test at a 5% significance level.

From Table 3.2, we see that Kansas City and Oklahoma City have statistically significant increases in warm-season thunderstorm occurrences and initiations in comparison to their rural counterparts because the p value is less than 0.05 and t value is positive. However, Dallas urban and rural areas have no significant differences (p value is greater than 0.05) in both thunderstorm tracks and initiations indicating that the Dallas urban area may not favor thunderstorms or it is masked by other circulations and convergence mechanisms induced by non-urban LULC.

Table 3.2 T-test on difference of means for thunderstorm tracks **(**km**)** and initiation (count) between urban and rural areas at three cities. No significance is shaded gray.

	Track	**Occurrence**	**Track**	**Initiation**
	t	**p**	**t**	**p**
Kansas City	11.79	0.000	26.47	0.000
Oklahoma City	17.21	0.000	32.51	0.000
Dallas	-6.5	1.000	-0.51	0.547

3.4 Conclusions

This research studies spatial and temporal characteristics of thunderstorm life cycles in central United States mainly covering Kansas, Oklahoma, and northern Texas during the warm seasons from 2010 to 2014. An improved centroid-based thunderstorm tracking algorithm was utilized to identify thunderstorm life cycles from radar reflectivity data and cloud-to-ground lightning data. The recorded life cycle of a thunderstorm includes initiation, development, termination, merger, and split. A directed graph model was used to represent the life cycles and to study the

interactions of thunderstorm cells (split and merger), and the maximum reflectivity path as a polyline was used to generalize the life cycle of a thunderstorm. Thunderstorm life cycles and their attributes were stored in a GIS database and GIS was used to visualize, query, and analyze thunderstorm life cycles.

Our climatological analyses indicate a strong peak of thunderstorm occurrences in May. Most of thunderstorms (65.8%) have a duration from 5 to 20 minutes. Thunderstorm initiation is most frequent from 2100 to 0000 UTC, and the thunderstorm termination is most common from 2100 to 0300 UTC.

Major thunderstorm activities are in the eastern part of the study area, especially at the border among Kansas, Missouri, Oklahoma, and Arkansas. We found initiation and termination hot spots along the border between Oklahoma and Arkansas and the south end at the border between Missouri and Kansas. Based on the directed graph representation, we found that splits and mergers are mainly located along the border between Oklahoma and Arkansas and the border between Kansas and Missouri, in central Oklahoma, and in central and northern part of Texas.

We also linked thunderstorms to land cover types, and found that thunderstorms favor forests and urban areas. Forests, crops, and grasses may trigger splits and mergers during the life cycle of a thunderstorm. Statistical analyses demonstrated that the urban areas in Kansas City and Oklahoma City had statistically significant thunderstorm occurrences than the surrounding rural areas, though the Dallas urban area did not show this feature.

The methods and analyses presented in this work demonstrate how to apply GIS representations and spatial analyses to meteorological study. Atmospheric science has many potentials to incorporate GIS due to spatiotemporal nature in atmospheric systems. For example, it is also interesting to represent and analyze other meteorological phenomena such as hurricanes

and heat wave. We would also like to enlarge the spatiotemporal coverage of the radar and lightning data to study thunderstorm characteristics for the entire United States, and explore the relationships between thunderstorms and other factors such as terrain in the future.

References

Ashley, W.S., M.L. Bentley, and J. A. Stallins, 2012. Urban-induced thunderstorm modification in the Southeast United States. *Climatic Change*, 113 (2), 481-498.

Changnon, S.A., 1988a. Climatography of thunder events in the conterminous United States. Part I: temporal aspects. *Journal of Climate*, 1(4), 389-398.

Changnon, S.A., 1988b. Climatography of thunder events in the conterminous United States. Part II: spatial aspects. *Journal of Climate*, 1(4), 399-405.

Chase, T.N., Pielke Sr, R.A., Kittel, T.G.F., Nemani, R.R. and Running, S.W., 2000. Simulated impacts of historical land cover changes on global climate in northern winter. *Climate Dynamics*, 16 (2-3), 93-105.

Chen, M., Y. Wang, F. Gao, and X. Xiao, 2012. Diurnal variations in convective storm activity over contiguous North China during the warm season based on radar mosaic climatology. *Journal of Geophysical Research: Atmospheres*, 117, D20.

Choi, J., Olivera, F. and Socolofsky, S. A., 2009. Storm identification and tracking algorithm for modeling of rainfall fields using 1-h NEXRAD rainfall data in Texas. *Journal of Hydrologic Engineering*, 14 (7), 721-730.

Cummins, K.L. and Murphy, M.J., 2009. An overview of lightning locating systems: history, techniques, and data uses, with an in-depth look at the US NLDN. *Electromagnetic Compatibility, IEEE Transactions on*, 51 (3), 499-518.

Dixon, M. and Wiener, G., 1993. TITAN: thunderstorm identification, tracking, analysis, and nowcasting-A radar-based methodology. *Journal of Atmospheric and Oceanic technology*, 10 (6), 785-797.

Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A. and Washington, W.M., 2005. The importance of land-cover change in simulating future climates. *Science*, 310 (5754), 1674-1678.

Feidas, H. and Cartalis, C., 2001. Monitoring mesoscale convective cloud systems associated with heavy storms using Meteosat imagery. *Journal of Applied Meteorology*, 40 (3), 491-512.

Feidas, H. and Cartalis, C., 2005. Application of an automated cloud-tracking algorithm on satellite imagery for tracking and monitoring small mesoscale convective cloud systems. *International journal of remote sensing*, 26 (8), 1677-1698.

Gambill, L.D. and Mecikalski, J.R., 2011. A satellite-based summer convective cloud frequency analysis over the southeastern United States. *Journal of Applied Meteorology and Climatology*, 50 (8), 1756-1769.

Han, L., Fu, S., Yang, G., Wang, H., Zheng, Y. and Lin Y., 2008. A stochastic method for convective storm identification, tracking and nowcasting. *Progress in Natural Science*, 18 (12), 1557-1563.

Han, L., Fu, S., Zhao, L., Zheng, Y., Wang, H. and Lin Y., 2009. 3D convective storm identification, tracking, and forecasting-An enhanced TITAN algorithm. *Journal of Atmospheric and Oceanic Technology*, 26 (4), 719-732.

Haralock, R.M. and Shapiro, L.G., 1991. *Computer and Robot Vision*. London: Addison-Wesley Longman.

Hocker, J.E. and Basara, J.B., 2008a. A geographic information systems-based analysis of supercells across Oklahoma from 1994 to 2003. *Journal of Applied Meteorology and Climatology*, 47 (5), 1518-1538.

Hocker, J.E. and Basara, J.B., 2008b. A 10-year spatial climatology of squall line storms across Oklahoma. *International Journal of Climatology*, 28 (6), 765-775.

Holt, T.R., Niyogi, D., Chen, F., Manning, K., LeMone, M.A. and Qureshi, A., 2006. Effect of land-atmosphere interactions on the IHOP 24-25 May 2002 convection case. *Monthly Weather Review*, 134 (1), 113-133.

Johnson, J., MacKeen, P.L., Witt, A., Mitchell, E.D., Stumpf, G.J., Eilts, M.D. and Thomas, K.W., 1998. The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Weather and Forecasting*, 13 (2), 263-276.

Lakshmanan, V. and Smith, T., 2010. An objective method of evaluating and devising storm-tracking algorithms. *Weather and Forecasting*, 25 (2), 701-709.

Lee, B.D., Jewett, B.F. and Wilhelmson, R.B., 2006a. The 19 April 1996 Illinois tornado outbreak. Part I: Cell evolution and supercell isolation. *Weather and forecasting*, 21 (4), 433-448.

Lee, B.D., Jewett, B.F. and Wilhelmson, R.B., 2006b. The 19 April 1996 Illinois tornado outbreak. Part II: Cell mergers and associated tornado incidence. *Weather and forecasting*, 21 (4), 449-464.

Li, L., Schmid, W. and Joss, J., 1995. Nowcasting of motion and growth of precipitation with radar over a complex orography. *Journal of Applied Meteorology*, 34 (6), 1286-1300.

Liu, W., Li, X. and Rahn, D.A., 2016. Storm event representation and analysis based on a directed spatiotemporal graph model. *International Journal of Geographical Information Science*, 30 (5), 948-969.

Lock, N.A. and Houston, A.L., 2015. Spatiotemporal distribution of thunderstorm initiation in the US Great Plains from 2005 to 2007. *International Journal of Climatology*, 35 (13), 4047-4056.

Machado, L.A.T., Rossow, W.B., Guedes, R.L. and Walker, A.W., 1998. Life cycle variations of mesoscale convective systems over the Americas. *Monthly Weather Review*, 126 (6), 1630-1654.

Mathon, V. and Laurent, H., 2001. Life cycle of Sahelian mesoscale convective cloud systems. *Quarterly Journal of the Royal Meteorological Society*, 127 (572), 377-406.

McIntosh, J. and Yuan, M., 2005. A framework to enhance semantic flexibility for analysis of distributed phenomena. *International Journal of Geographical Information Science*, 19 (10), 999-1018.

Meyer, V.K., Höller, H. and Betz, H.D., 2013. Automated thunderstorm tracking: utilization of three-dimensional lightning and radar data. *Atmospheric Chemistry and Physics*, 13 (10), 5137-5150.

Mohee, F.M. and Miller, C., 2010. Climatology of thunderstorms for North Dakota, 2002–06. *Journal of Applied Meteorology and Climatology*, 49 (9), 1881–1890.

Morel, C. and Senesi, S., 2002a. A climatology of mesoscale convective systems over Europe using satellite infrared imagery. I: Methodology. *Quarterly Journal of the Royal Meteorological Society*, 128 (584), 1953-1971.

Morel, C. and Senesi, S., 2002b. A climatology of mesoscale convective systems over Europe using satellite infrared imagery. II: Characteristics of European mesoscale convective systems. *Quarterly Journal of the Royal Meteorological Society*, 128 (584), 1973-1995.

Niyogi, D., Pyle, P., Lei, M., Arya, S.P., Kishtawal, C.M., Shepherd, M., Chen, F. and Wolfe, B., 2011. Urban modification of thunderstorms: an observational storm climatology and model case study for the Indianapolis urban region. *Journal of Applied Meteorology and Climatology*, 50 (5), 1129-1144.

Novo, S., Martinez, D. and Puentes, O., 2013. Tracking, analysis, and nowcasting of Cuban convective cells as seen by radar. *Meteorological Applications*, 21 (3), 585-595.

Owen, J., 1966. A study of thunderstorm formation along dry lines. *Journal of Applied Meteorology*, 5 (1), 58-63.

Perryman, N. and P.G. Dixon, 2013. A radar analysis of urban snowfall modification in Minneapolis–St. Paul. *Journal of Applied Meteorology and Climatology*, 52 (7), 1632-1644.

Robinson, P.J. and Easterling, D.R., 1988. The frequency distribution of thunderstorm durations. *Journal of Applied Meteorology*, 27 (1), 77-82.

Shepherd, J.M., 2005. A review of current investigations of urban-induced rainfall and recommendations for the future. *Earth Interactions*, 9 (12), 1-27.

Shepherd, J.M., Pierce, H. and Negri, A.J., 2002. Rainfall modification by major urban areas: Observations from spaceborne rain radar on the TRMM satellite. *Journal of Applied Meteorology*, 41 (7), 689-701.

Tucker, D.F. and Li, X., 2009. Characteristics of warm season precipitating storms in the Arkansas–Red River basin. *Journal of Geophysical Research*, 114 (D13).

Tuttle, J.D. and Foote, G.B., 1990. Determination of the boundary layer airflow from a single Doppler radar. *Journal of Atmospheric and oceanic Technology*, 7 (2), 218-232.

Wang, Y., Han, L. and Wang, H., 2014. Statistical characteristics of convective initiation in the Beijing-Tianjin region revealed by six-year radar data. *Journal of Meteorological Research*, 28, 1127-1136.

Whitehall, K., Mattmann, C.A., Jenkins, G., Rwebangira, M., Demoz, B., Waliser, D., Kim, J., Goodale, C., Hart, A., Ramirez, P. and Joyce, M.J., 2015. Exploring a graph theory based algorithm for automated identification and characterization of large mesoscale convective systems in satellite datasets. *Earth Science Informatics*, 8 (3), 663-675.

Wilson, J.W., Crook, N.A., Mueller, C.K., Sun, J. and Dixon, M., 1998. Nowcasting thunderstorms: A status report. *Bulletin of the American Meteorological Society*, 79 (10), 2079-2099.

Wilson, J.W., Ebert, E.E., Saxen, T.R., Roberts, R.D., Mueller, C.K., Sleigh, M., Pierce, C.E. and Seed, A., 2004. Sydney 2000 forecast demonstration project: convective storm nowcasting. *Weather and forecasting*, 19 (1), 131-150.

Yuan, M., 2001. Representing complex geographic phenomena in GIS. *Cartography and Geographic Information Science*, 28 (2), 83-96.

Zahraei, A., Hsu, K., Sorooshian, S., Gourley, J.J., Hong, Y. and Behrangi, A., 2013. Short-term quantitative precipitation forecasting using an object-based approach. *Journal of Hydrology*, 483, 1-15.

4 Storm event similarity analysis

4.1 Introduction

Modeling movement patterns is an important research topic in GIS. A number of researchers (McIntosh and Yuan 2005, Xia et al. 2011, Dodge et al. 2012, Yuan and Raubal 2014) have studied inter-event similarity which supports event clustering, generalization, and pattern recognition (Dodge et al. 2009, Miller 2012, Buchin et al. 2014).

Similarity analysis answers the question: 'How similar are the movement paths of two or more events?' (Buchin et al. 2014). In GIS, similarity measure is quantified as the cost or distance transforming one event to another (Faloutsos et al. 1997, Dodge et al. 2012). There are two basic similarity analyses: spatial similarity which only considers event geometric shape and spatiotemporal similarity which considers the spatial and temporal characteristics of events (Tiakas et al. 2009, Abraham and Lal 2012, Dodge et al. 2012, Buchin et al. 2014). A number of distance measurements such as Euclidean distance, edit distance, Hausdorff distance, and dynamic time warping (Ding et al. 2008, Alt 2009, Yuan and Raubal 2014) are used to assess event similarity. Ranacher and Tzavella (2014) reviewed physical movement similarity measures from spatial, temporal, and spatiotemporal perspectives.

McIntosh and Yuan (2005) assessed storm similarity using six indices to capture static and dynamic features of storms such as elongation, orientation, distribution, growth, granularity of change, and relative movement. They then applied the dynamic time warping method to the sequences of the indices to determine the similarity among storm events. However, they did not

consider specific types of storm cells such as single storm cell, multi-cell clusters, and supercells. Dodge et al. (2012) compared the similarity among hurricane trajectories via three major steps. The first one is to decompose a trajectory into a number of segments with homogeneous characteristics such as speed, acceleration, or direction. Then the trajectory was transformed into a sequence of labels/letters. Finally, normalized weighted edit distance was used to assess the similarities among different letter sequences. The advantage of this method is to convert complex trajectories to sequences of letters and apply the string edit distance (Levenshtein 1966, Wagner and Fischer 1974) to measure similarity. However, they only applied one movement parameter such as speed or direction to segment trajectories and did not deal with split or merger interactions. Yuan and Raubal (2014) developed spatiotemporal edit distance to determine the similarity among people trajectories using phone detailed records. They improved traditional string edit distance algorithm by considering both spatial and temporal information in their cost functions.

In our research, the major objective is to assess the similarity among storm events. Convective storm events are represented using a spatiotemporal directed graph model. The advantage of graph-based data model is that there are a number of graph algorithms we could explore and extend. Since graph edit distance (GED) measures the similarity among graphs (Conte et al. 2004, Gao et al. 2010, Ferrer and Bunke 2012, Cheung et al. 2015) we would like to use the GED to assess the similarities of convective storm events. If we use the maximum reflectivity path to generalize storm events as single trajectories, we could also use the string edit distance algorithm to assess their similarity. The data and methodologies are introduced in Section 4.2, and preliminary experiments are presented in Section 4.3.

4.2 Data and methodology

The data used in this study is the 5-years' (2010 - 2014) warm season thunderstorm events identified in Chapter 3. The primary methods for measuring event similarity include string edit distance and graph edit distance.

4.2.1 Classification of single storm cells

Graph nodes, representing people and vehicles, could find the exact identities between the two graphs. However, there are no clear identities for convective storms, and we used reflectivity and area thresholds to delineate convective storm cells (Section 2.2.1). As a consequence, we firstly used attributes, such as area, geometric shape, and intensity, to differentiate convective storm cells as single cell, multi-cell clusters, and supercells (Figure 4.1a). In this study, there are a total of twenty classes for single storm cells based on previous studies (Johnson 2004, Gallus et al. 2008, Smith et al. 2012). Then we manually classified about 600 storm events which have about 3,000 single storm cells (Table 4.1) in 2014. These storm cells are on the maximum reflectivity paths. We used the first twenty letters (A through T) to represent the storm cell classes and transformed the sequence of storm cells into a string. We could then use string edit distance algorithm (Section 4.2.2) to assess the similarity among storm events. Figure 4.1b is the string representations for the ten example storm events in Table 4.1. For example, storm #2 has three storm cells, and their types are isolated thunderstorm (I), cluster (Q), and cluster (Q) respectively.

We cannot manually classify all the single storm cells because of the high volume, so we will do some regression models to automatically relate the types of storm cells with their attributes including area and geometric shape. After classifying all the storm cells in the whole graph, the graph edit distance algorithm (Section 4.2.3) could be applied to assess the similarity among storm event graphs.

(a)

1. Quasi-Linear Convective System
   - 1.1 No Stratiform (A)
   - 1.2 Trailing Stratiform (B)
   - 1.3 Leading Stratiform (C)
   - 1.4 Parallel Stratiform (D)
   - 1.5 Bow Echo (E)
   - 1.6 Broken Line (F)
   - 1.7 Back Building (G)
2. Discrete
   - 2.1 Isolated Supercell (H)
   - 2.2 General Isolated Thunderstorm (I)
   - 2.3 Multiple Discrete Cells (J)
   - 2.4 Splitting Discrete Cells (K)
3. Synoptic
   - 3.1 Stratiform (L)
   - 3.2 MCS (M)
4. Clusters
   - 4.1 Clustered Supercell (N)
   - 4.2 Clustered General Thunderstorms (O)
   - 4.3 Multicell Cluster (P)
5. Disorganized
   - 5.1 Cluster (Q)
   - 5.2 Independent Storm (R)
   - 5.3 Line (S)
6. Other
   - 6.1 Unusual (T)

(b)

1. **IN**
2. **IQQ**
3. **LL**
4. **DDF**
5. **QIQNN**
6. **HHHHH**
7. **LLIL**
8. **FFFCC**
9. **CDD**
10. **HHH**

Figure 4.1 (a) Classes of single storm cells; (b) String representations for the ten storm events in Table 4.1.

Table 4.1 Properties and single storm cell types of ten example storm events.

ID	# of Cells	Total Precip	Footprint Size	Footprint CV	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5
1	416	5806	333	0.35453	I	N			
2	392	5309	283	0.41642	I	Q	Q		
3	106	1403	87	0.33764	L	L			
4	1095	15256	620	0.43553	D	D	F		
5	1061	14528	524	0.46996	Q	I	Q	N	N
6	531	7358	209	0.48902	H	H	H	H	H
7	209	2805	126	0.44599	L	L	I	L	

8	3629	51979	1411	0.52396	F	F	F	C	C
9	672	9611	373	0.45253	C	D	D		
10	599	8181	397	0.44988	H	H	H		

4.2.2 String edit distance

Wagner and Fischer (1974) proposed edit distance, which quantifies how similar two strings are using the minimum number of operations required to transform one string to another. There are three edit operations including insertion, deletion, and substitution. For example, the edit distance between IN and IQQ is 2. IN to IQ (substitution of Q for N), and IQ to IQQ (insertion of Q at the end). The edit distance between IN and LL is also 2. IN to LN (substitution of L for I), and LN to LL (substitution of L for N). However, there are no the same storm cell types between the two storm events, so there should be no similarity between the two storm events.

4.2.3 Graph edit distance

Graph edit distance (GED) is a metric using the minimum number of editing nodes and edges including insertion, deletion, and substitution (Riesen and Bunke 2009, Cheung et al. 2015, Fischer et al. 2015) when changing a graph into another. An example GED between graph G1 and G2 is shown in Figure 4.2. When graph G1 is transformed into G2, the following six operations are needed: removal of node C, removal of edge CB, addition of node E, addition of edge BE, addition of node F, and addition of edge DF. If the above six operations have equal cost/distance, such as 1, the final edit distance between the two graphs is 6. The smaller the edit distance is, the more similar the two graphs are. Two graphs are the most similar if they have the smallest GED.

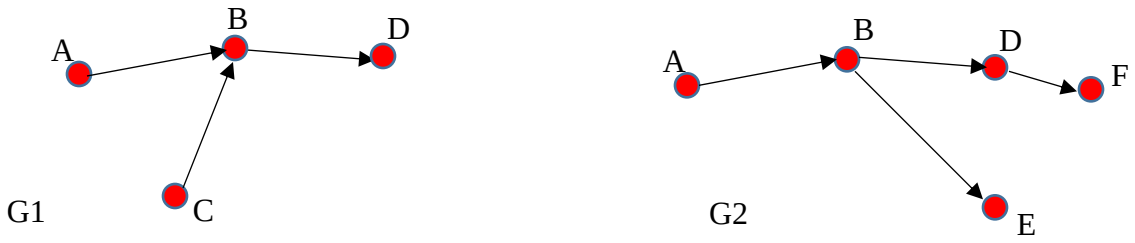


Figure 4.2. Two graphs for an example GED.

After classifying all the storm cells using the regression model based on the manual classifications, GED algorithm will be used to assess the similarity among different storm graphs. A graph matching tool (Figure 4.3) developed by Riesen et al. (2013) will be explored. From component ② in Figure 4.3, there are several different GED algorithms we could use. After assessing the similarities of convective storm events, we will evaluate the similarity algorithm in storm event clustering.

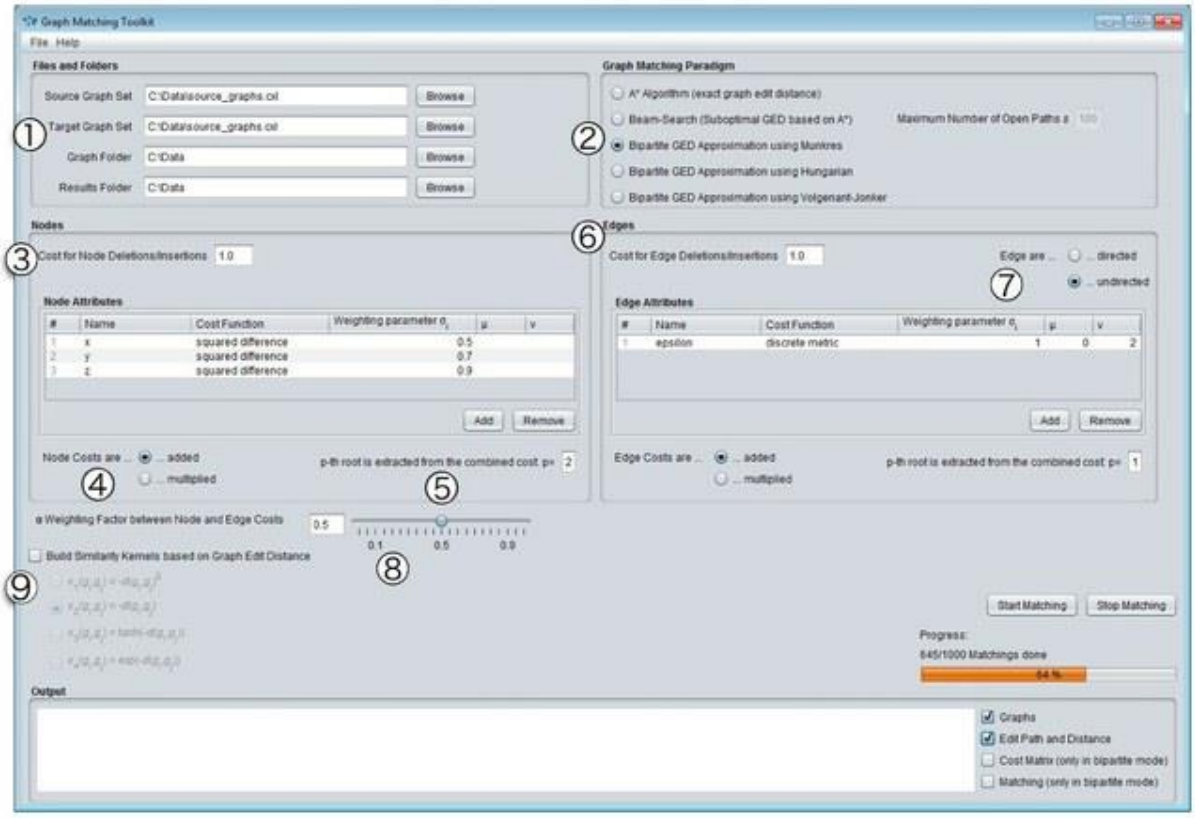


Figure 4.3 The main window of graph matching tool. Adopted from Riesen et al. (2013).

4.3 Preliminary results

We calculated string edit distance on the storm events that we have manually classified their single storm cells. The edit distance matrix for the ten storm events in Table 4.1 is shown in

Table 4.2. Based on the edit distance, we could identify the most similar storm event for any storm events in the table. However, even if the edit distance is the smallest, two storm events may not have any similarity. In this case, we assign a big number such as 1000 to the edit distance to indicate no similarity if there are no the same storm cell types between two storm events. In the future, we will explore the graph edit distance introduced in previous section.

Table 4.2 Edit distance matrix for ten example storm events in Table 4.1. Yellow color indicates the most similar storm event for the storm event in a column.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1		2	2	3	3	5	3	5	3	3
S2			3	3	3	5	4	5	3	3
S3				3	5	5	2	5	3	3
S4					5	5	4	4	2	3
S5						5	5	5	5	5
S6							5	5	5	2
S7								5	4	4
S8									5	5
S9										3
S10										

References

Abraham, S. and Lal, P.S., 2012. Spatio-temporal similarity of network-constrained moving object trajectories using sequence alignment of travel locations. *Transportation research part C: emerging technologies*, 23, 109-123.

Alt, H., 2009. The computational geometry of comparing shapes. *In*: S. Albers, H. Alt, and S. Näher, eds. *Efficient Algorithms*. Springer Berlin Heidelberg, 235-248.

Buchin, M., Dodge, S. and Speckmann, B., 2014. Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science,* 2014 (9), 101-124.

Cheung, A.K.L., O'sullivan, D. and Brierley, G., 2015. Graph-assisted landscape monitoring. *International Journal of Geographical Information Science*, 29 (4), 580-605.

Conte, D., Foggia, P., Sansone, C. and Vento, M., 2004. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18 (03), 265-298.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X. and Keogh, E., 2008. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment*, 1 (2), 1542-1552.

Dodge, S., Laube, P. and Weibel, R., 2012. Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science,* 26 (9), 1563-1588.

Dodge, S., Weibel, R. and Forootan, E., 2009. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33 (6), 419-434.

Faloutsos, C., Jagadish, H.V., Mendelzon, A.O. and Milo, T., 1997. A signature technique for similarity-based queries. *In*: *Compression and Complexity of Sequences 1997.* Washington, DC: IEEE Computer Society, 2-20.

Ferrer, M. and Bunke, H., 2012. Graph edit distance – theory, algorithms, and applications. *In*: O. Lezoray and L. Grady, eds. *Image processing and analysis with graphs: theory and practice*. London: CRC Press, 383-422.

Fischer, A., Suen, C.Y., Frinken, V., Riesen, K. and Bunke, H., 2015. Approximation of graph edit distance based on Hausdorff matching. *Pattern Recognition,* 48 (2), 331-343.

Gallus Jr, W.A., Snook, N.A. and Johnson, E.V., 2008. Spring and summer severe weather reports over the Midwest as a function of convective mode: A preliminary study. *Weather and Forecasting*, 23 (1), 101-113.

Gao, X., Xiao, B., Tao, D. and Li, X., 2010. A survey of graph edit distance. *Pattern Analysis and applications*, 13 (1), 113-129.

Johnson, R.H., 2004. Organization of Mesoscale Convective Systems. *In*: *Abstracts for the 6th International GAME Conference*.

Levenshtein, V.I., 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet physics doklady*, 10, 707-710.

McIntosh, J. and Yuan, M., 2005. Assessing similarity of geographic processes and events. *Transactions in GIS*, 9 (2), 223-245.

Miller, J.A., 2012. Using spatially explicit simulated data to analyze animal interactions: a case study with brown hyenas in northern Botswana. *Transactions in GIS,* 16 (3), 271-291.

Ranacher, P. and Tzavella, K., 2014. How to compare movement? A review of physical movement similarity measures in geographic information science and beyond. *Cartography and geographic information science*, 41 (3), 286-307.

Riesen, K. and Bunke, H., 2009. Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision computing*, 27 (7), 950-959.

Riesen, K., Emmenegger, S. and Bunke, H., 2013. A novel software toolkit for graph edit distance computation. *In*: *International Workshop on Graph-Based Representations in Pattern Recognition*. Springer Berlin Heidelberg, 142-151.

Smith, B.T., Thompson, R.L., Grams, J.S., Broyles, C. and Brooks, H.E., 2012. Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Storm classification and climatology. *Weather and Forecasting*, 27 (5), 1114-1135.

Tiakas, E., Papadopoulos, A.N., Nanopoulos, A., Manolopoulos, Y., Stojanovic, D. and Djordjevic-Kajan, S., 2009. Searching for similar trajectories in spatial networks. *Journal of Systems and Software*, 82 (5), 772-788.

Wagner, R.A. and Fischer, M.J., 1974. The string-to-string correction problem. *Journal of the ACM*, 21 (1), 168-173.

Xia, Y., Wang, G.Y., Zhang, X., Kim, G.B. and Bae, H.Y., 2011. Spatio-temporal similarity measure for network constrained trajectory data. *International Journal of Computational Intelligence Systems*, 4 (5), 1070-1079.

Yuan, Y. and Raubal, M., 2014. Measuring similarity of mobile phone user trajectories–a Spatio-temporal Edit Distance method. *International Journal of Geographical Information Science*, 28 (3), 496-520.

5 Conclusions and Future Research

5.1 Summary of findings and contributions

Large amounts of spatiotemporal data have been collected or generated for the monitoring and modeling of environmental systems. Those time series of data also provide the opportunity to study the movements and dynamics of many different natural phenomena. While the snapshot organization is simple and straightforward, it does not directly capture or represent the dynamic characteristics of geographic phenomena. By defining the clusters of attributes and their changes in space and time as spatiotemporal events and by making the events explicit from spatiotemporal snapshot data, we develop innovative and computational efficient methods and tools for the identification, representation, and characterization of events based on a directed spatiotemporal graph model using convective storm events as an example.

Chapter 2 of this dissertation addresses how to identify the life cycle of convective storm events from consecutive radar images, how to use a directed spatiotemporal graph model to represent the convective storms, and how to use graph algorithms to analyze the storm events. The major findings and contributions can be listed as follows:

1) An event-based theory was applied to dynamic geographical phenomena, which can be systematically delineated with an origin, a development stage, a movement stage, and a potential cessation or dissolution phase.

2) An improved centroid-based storm tracking algorithm was developed, which simultaneously considers the topology/spatial overlap, centroid distance of storm objects, and movement direction.
3) Sensitivity analysis was implemented to test the accuracy of the storm tracking algorithm to three thresholds including area, reflectivity, and area overlap.
4) A directed spatiotemporal graph model was proposed to represent the convective storm events. In this model, graph nodes represent the storm cells, and directed graph edges represent the filiation relationships among storm cells. Using this graph model, the following filiations can be well visualized and represented: generation, continuity, split, merger, combinatorial, and dissipation.
5) Graph-based algorithms were explored to analyze the convective storm events. The first one is the maximum reflectivity path using the classic Dijkstra's shortest path algorithm on the graph to generalize the original storm event. The second one is in-degree/out-degree to model the interactions among storm cells including split and merger.
6) A MATLAB program was developed to realize the storm tracking algorithm, directed spatiotemporal graph representation, graph-based algorithms, visualization of convective storms using plots and videos.
7) The methodologies was applied to one of the deadliest storm outbreaks that impacted 15 states of southeastern U.S. between April 23 and 29, 2011. Several statistical analyses was done including duration, movement speed and direction, which confirmed theory, numerical simulations, and other observed case studies.

Chapter 3 of this dissertation addresses how to apply the identification and representation methodologies to thunderstorm spatiotemporal characteristics in central United States from 2010 to 2014. The major findings and contributions can be listed as follows:

1) The storm tracking algorithms were applied to five years' warm season radar reflectivity data, and the cloud-to-ground lightning data were used to extract the thunderstorm events in central United States mainly covering the whole Kansas, Oklahoma, and northern Texas.
2) A number of spatial and temporal characteristics of thunderstorms were studied. Most of thunderstorms (65.8%) have a duration from 5 to 20 minutes. Thunderstorm initiation is most frequent from 2100 to 0000 UTC, and the thunderstorm termination is most common from 2100 to 0300 UTC.
3) Major thunderstorm activities are in the eastern part of the study area, especially at the border among Kansas, Missouri, Oklahoma, and Arkansas.
4) Thunderstorm initiation and termination hot spots are along the border between Oklahoma and Arkansas and the south end at the border between Missouri and Kansas.
5) Splits and mergers are mainly located along the border between Oklahoma and Arkansas and the border between Kansas and Missouri, in central Oklahoma, and in central and northern part of Texas.
6) As for the relationship between thunderstorms and land cover types, thunderstorms favor forests and urban areas. Forests, crops, and grasses may trigger splits and mergers during the life cycle of a thunderstorm.

7) Statistical analyses demonstrated that the urban areas in Kansas City and Oklahoma City had statistically significant thunderstorm occurrences than the surrounding rural areas, though the Dallas urban area did not show this feature.

In conclusion, this dissertation tries to propose a general workflow to identify, represent, and analyze dynamic geographical phenomena especially in the big data era. The identification algorithm should be application dependent, but the representational framework (directed spatiotemporal graph model) can be general like the two basic models (vector model and raster model) in the GIScience field. Based on this general representational model, we could further analyze the dynamic geographical phenomena. The biggest advantage of graph-based data model is that there are a number of existing graph algorithms we could explore and extend. Although this dissertation chose convective storm events as an example, the automatic processing methods and data model could be extended to other fields such as temperature, wildfire, land use/land cover and hurricane.

5.2 Future research

The possible future research is listed as follows:

1) We will continue the similarity analysis based on the proposed methodologies in Chapter 4. First, the storm cell types will be automatically classified, and then use the graph edit distance algorithm to assess the similarity. Storm event clustering will be used to evaluate the algorithms. There are also some problems we should deal with such as different cost/distance among different storm cell types' substitution to improve the current cost 1.
2) The identification, representation, and analysis framework is intended to apply to other geographical phenomena such as hurricanes and urban heat. The workflow may be

domain dependent, so there should be some modifications and improvements to the current methods.

3) I would like exploring high performance computation such as cloud, distributed computation in order to improve computational efficiency. With respect to my research of storm event data from 2010 to 2014, finding ways to shorten computational time is critical and I would therefore like to parallelize the event tracking algorithm on high performance computational infrastructure.
4) I am interested in exploring additional graph-based algorithms and methodologies to analyze the patterns, similarities, semantics, and dynamics of geographic phenomena. Although I have tried several graph algorithms, there are several promising algorithms I could look into in the future.