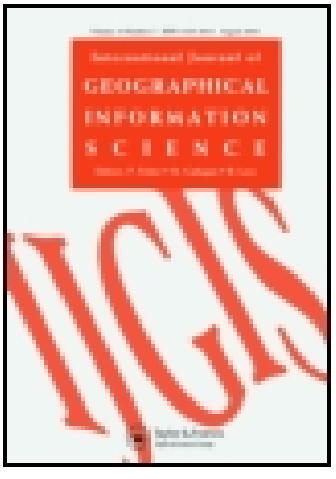

International Journal of Geographical Information Science

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

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

Weibo Liu[a], Xingong Li[ab] & David A. Rahn[c]

[a] Department of Geography, University of Kansas, Lawrence, KS, USA

[b] State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China

[c] Atmospheric Science Program, Department of Geography, University of Kansas, Lawrence, KS, USA

Published online: 28 Aug 2015.

To cite this article: Weibo Liu, Xingong Li & David A. Rahn (2015): Storm event representation and analysis based on a directed spatiotemporal graph model, International Journal of Geographical Information Science

To link to this article: http://dx.doi.org/10.1080/13658816.2015.1081910

International Journal of Geographical Information Science, 2015
http://dx.doi.org/10.1080/13658816.2015.1081910

Taylor & Francis
Taylor & Francis Group

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

Weibo Liu[a], Xingong Li[a,b]* and David A. Rahn[c]

[a]*Department of Geography, University of Kansas, Lawrence, KS, USA;* [b]*State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China;* [c]*Atmospheric Science Program, Department of Geography, University of Kansas, Lawrence, KS, USA*

(*Received 31 January 2015; accepted 29 July 2015*)

Large amounts of time series of spatial snapshot 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 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 used to analyze the dynamics of the events. The method is applied to time series of high-resolution radar reflectivity images during one of the deadliest storm outbreaks that impacted 15 states of southeastern United States between 23 and 29 April 2011. As shown in this case study, convective storm events identified using our methods are consistent with previous studies, and our analysis confirms 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. While this study does not differentiate between storm modes, the method shows potential for capturing a more detailed climatology of precipitation characteristics.

Keywords: storm event; graph model; split and merger; movement

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, which 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 of 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

*Corresponding author. Email: lixi@ku.edu

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 predefined 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, 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 2D 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 data sets 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 life cycle 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 data sets, 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 1 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 life cycles 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. Sensitivity analysis of the tracking algorithm to the reflectivity, area, and overlap threshold is discussed in Section 3. The directed spatiotemporal graph model for representing, storing, and analyzing storm event life cycles is presented in Sections 4 and 5. A case study illustrates the

capability of the proposed methods in Section 6. A summary of the research and possible future work is provided in Section 7.

2. Tracking storm events

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 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*. 2013). 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 3.

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 studies, 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 data set. 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, i.e., 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 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 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

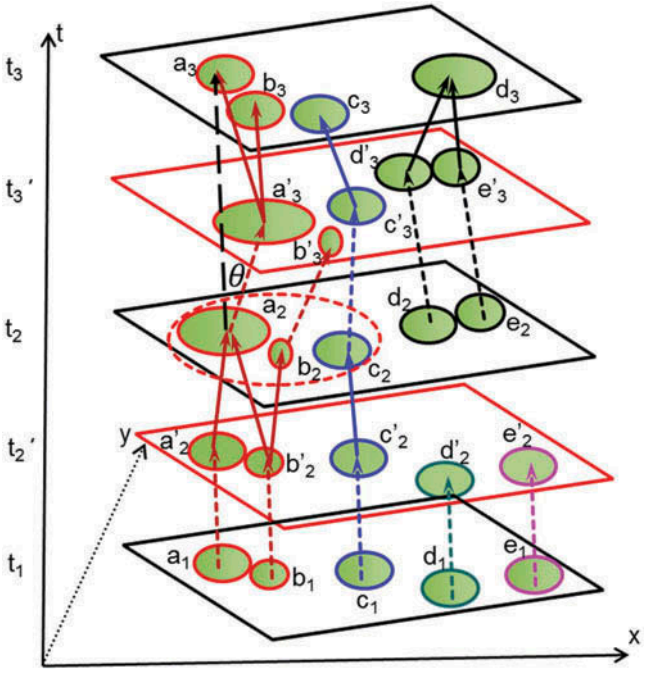

Figure 1. Storm event tracking examples during three time steps.

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 3.

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

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:

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

where $\boldsymbol{V}(S)$ 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, $\boldsymbol{S}_i\boldsymbol{S}$ 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 $\boldsymbol{V}(S)$ 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° 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*. 2013) between storm objects in an event. In Figure 1, five storm events are extracted (shown with different colors in Figure 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 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*. 2013) 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 life cycle and the interactions among storm objects. In contrast, our method records the filiation relationships among all the storm objects in the time series snapshots that satisfy the matching criteria (overlapping area, centroid distance, and movement direction).

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 chosen is 32.5°N–40.5°N and 93.5°W–103.5°W (the blue dashed boundary in Figure 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 demonstrate the sensitivity of movement speed to changes in the three thresholds. In Figure 2a, there are nine combinations of reflectivity and area threshold for each overlap threshold. It is easy 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

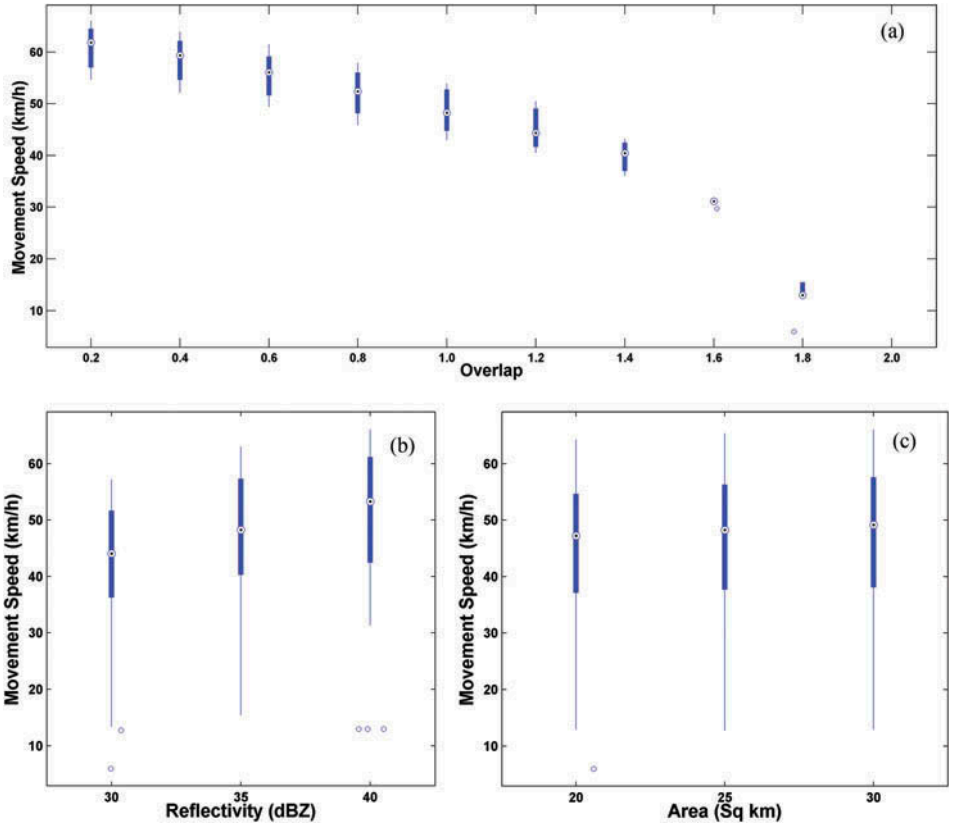

Figure 2. (a) Overlap versus movement speed. (b) Reflectivity versus movement speed. (c) Area versus movement speed.

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 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 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.

4. A directed spatiotemporal graph model

It is natural to use a directed spatiotemporal graph model (Figure 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 3a is used as a simplified representation of the storm object geometry, which works fairly well for distinct storm objects. Figure 3a is the three-dimensional view (x, y, t) of the directed spatiotemporal graph derived from Figure 1. To facilitate the analysis of the storm events, the spatiotemporal graph is also projected in

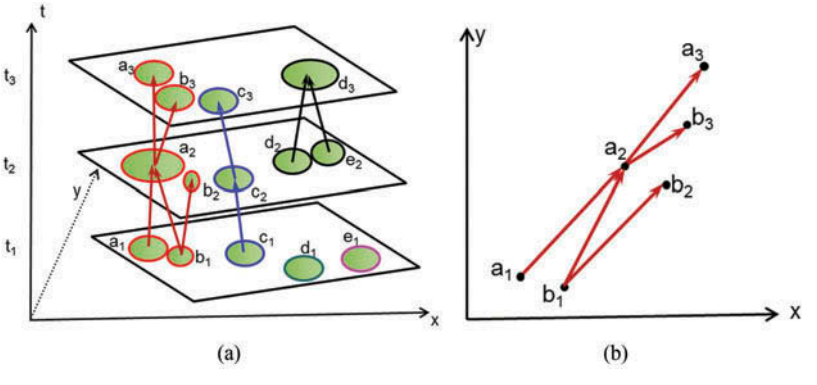

Figure 3. (a) The storm events and their directed spatiotemporal graph models identified from Figure 1. (b) The projection of the first storm graph on the x–y plane.

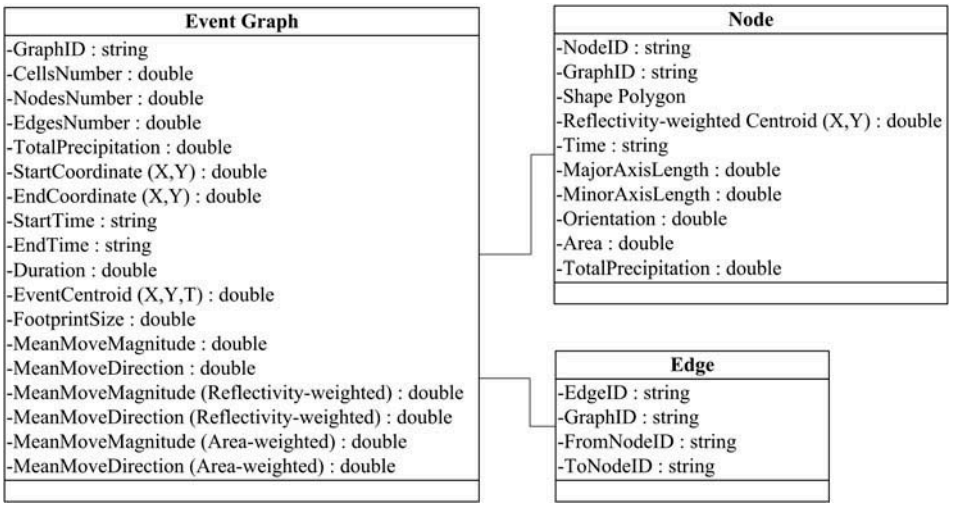

Figure 4. Components and their primary properties of the directed graph model for storm events.

two dimensions (*x, y*) in Figure 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 life span of the storm and includes the generation, continuity, split, merger, combinatorial, and dissipation filiation relationships.

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 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.

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.

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,

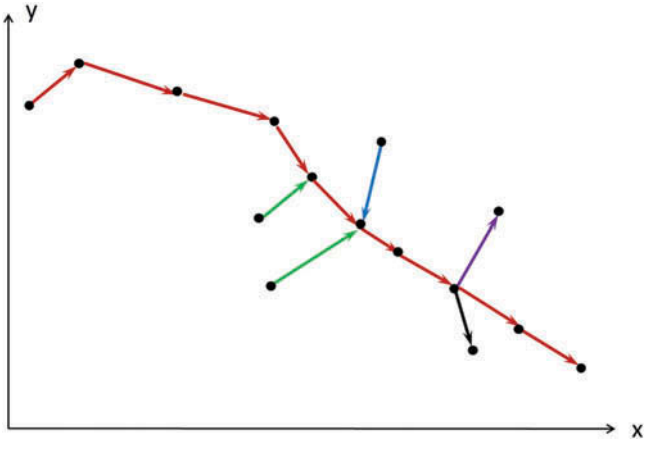

Figure 5. A storm event graph showing the maximum accumulative reflectivity path (red line), left split (purple line), right split (black line), left merger (blue line), and right merger (green line) along its movement direction.

can be achieved by using the maximum reflectivity path between the starting and ending storm objects of the event (Figure 5). The maximum reflectivity path can be found by first calculating the reverse reflectivity of the *i*th node as

$$W_i = \max(R) - R_i + \min(R) \tag{3}$$

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.

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 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 5).

Table 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

6. A case study

A convective storm outbreak that occurred on 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 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.

The data used in this case study are the final reflectivity product (N0 R) 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/nex

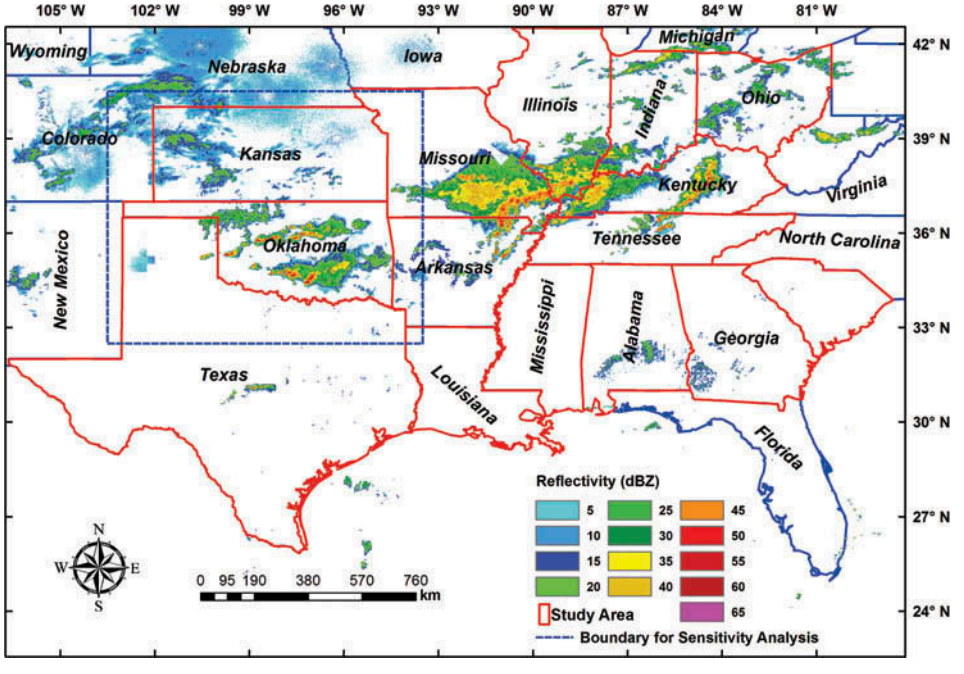

Figure 6. Case study area which covers 15 states (red) and an example radar image showing the reflectivity values at 8:00, 24 April 2011 UTC.

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

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 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.

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 7). The tracking process outlined in Section 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 life cycle) 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 7). The last component takes the event graphs as input and provides functions

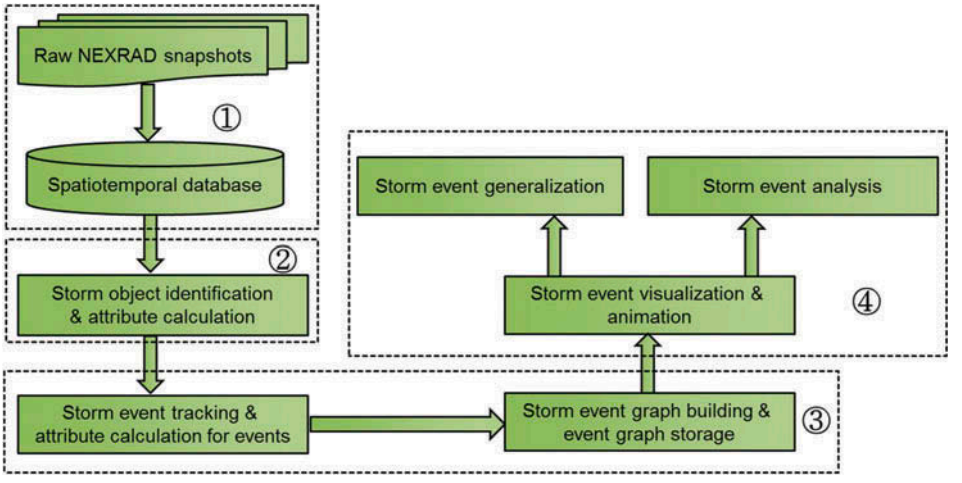

Figure 7. The general workflow of storm event identification and analysis developed using MATLAB.

for event visualization, animation, generalization, and interaction assessment. The visualization and animation interface, shown in Figure 8, could be used to browse event graph attributes, the maximum reflectivity path, and the visualization and animation of event

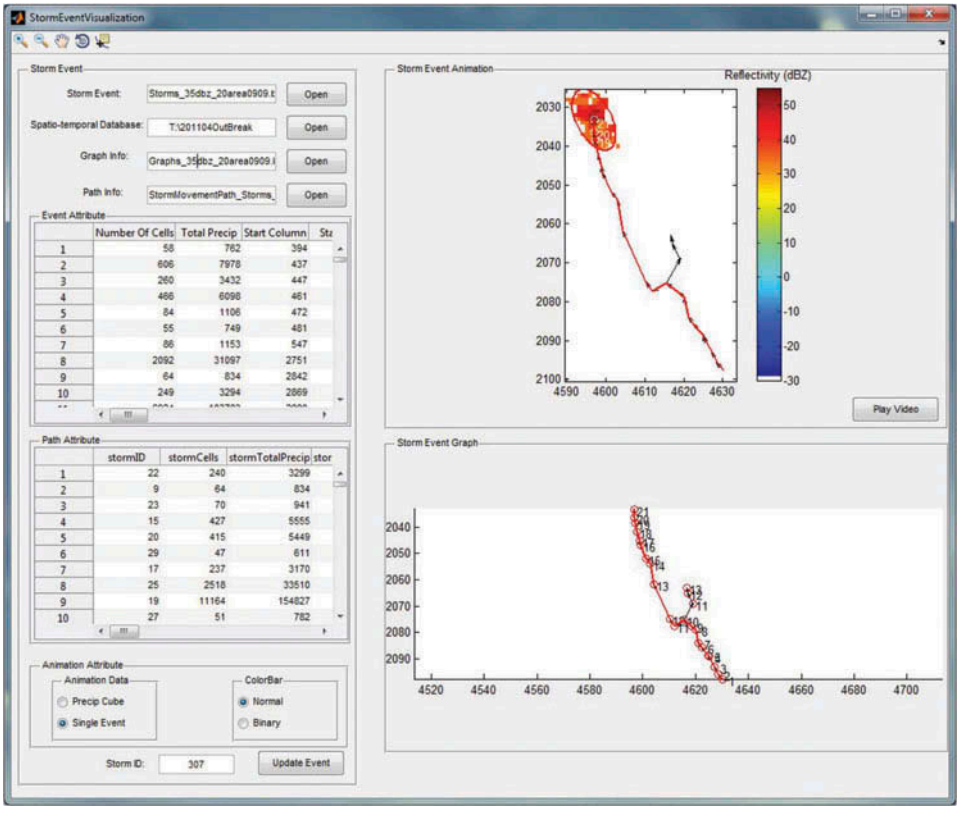

Figure 8. The user interface for event visualization and animation.

graphs. In Figure 8, a storm event is shown with the red skeleton and one right split with black line.

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 10 hours. In the study area there were a total of 7297 storm events that have a duration of at least 15 minutes, meaning they span at least three consecutive radar images. Figure 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 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 10b) depicts a positively skewed Gaussian distribution. The average and median movement speeds of the storm events are 61 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 10c) depicts an exponential decay distribution similar to Novo *et al*. (2013). The mean duration of the storm events is 36.5 minutes and 1141 out of 7297 storm events lasted more than 1 hour (15.6%). Most of storm events (44.4%) had a duration from 15 to 20 minutes. Among the 7297 storm events, just over a quarter of storm events (1928) have either splits or mergers ($\text{mean}_{\text{split}} = 0.17$, $\text{mean}_{\text{merger}} = 0.38$). There are 863 storm events that split and most of them (796) have one or two splits during their life spans. There are 1529 storm events that merged and most of them (1274) have only one or two mergers during their life spans.

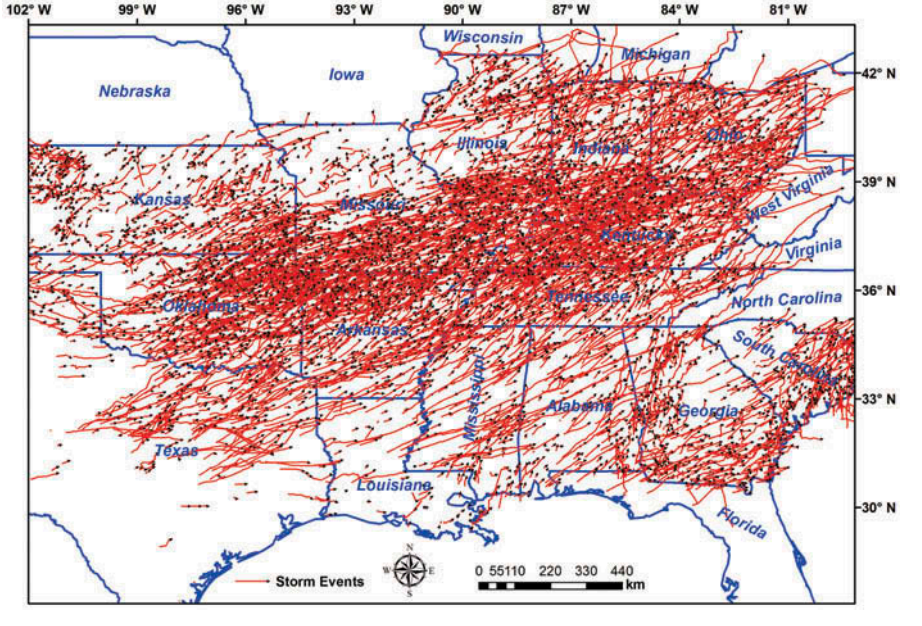

Figure 9. Map of generalized tracks of the storm events identified from the radar data between 23 April 2011 and 29 April 2011 in the study area.

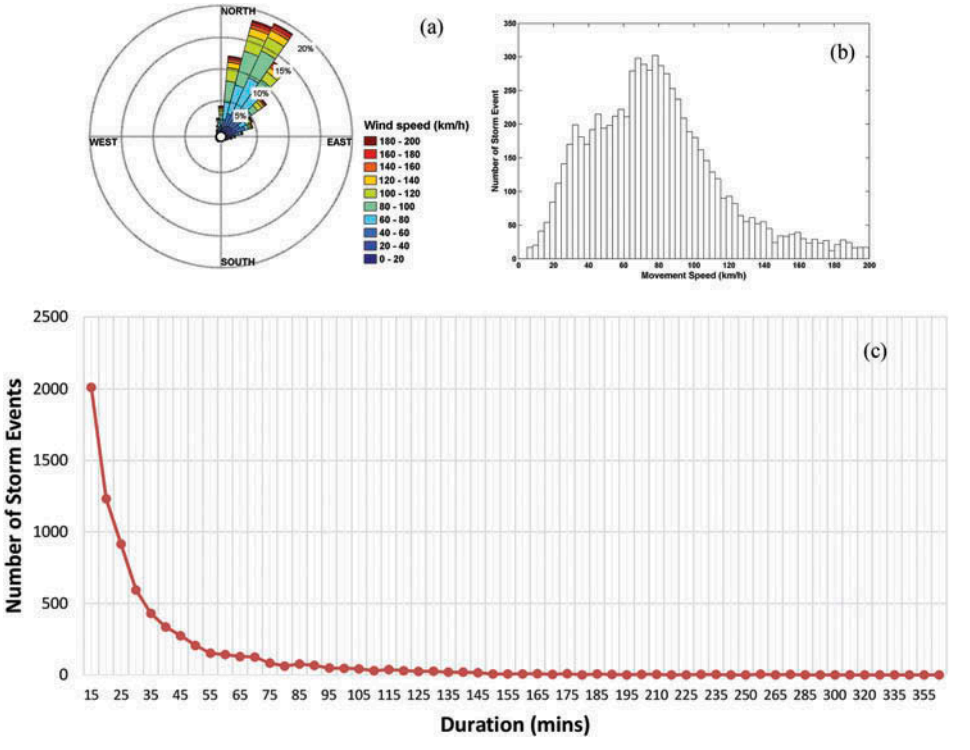

Figure 10. Characteristics of the storm events: (a) rose diagram of event velocity; (b) histogram of event speed; (c) histogram of event duration.

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 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 5. For these events, the locations of left and right split/merger are shown in Figures 11 and 12. Table 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 life spans. 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 differences between the number of left and right splits and mergers are shown in Figures 11 and 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

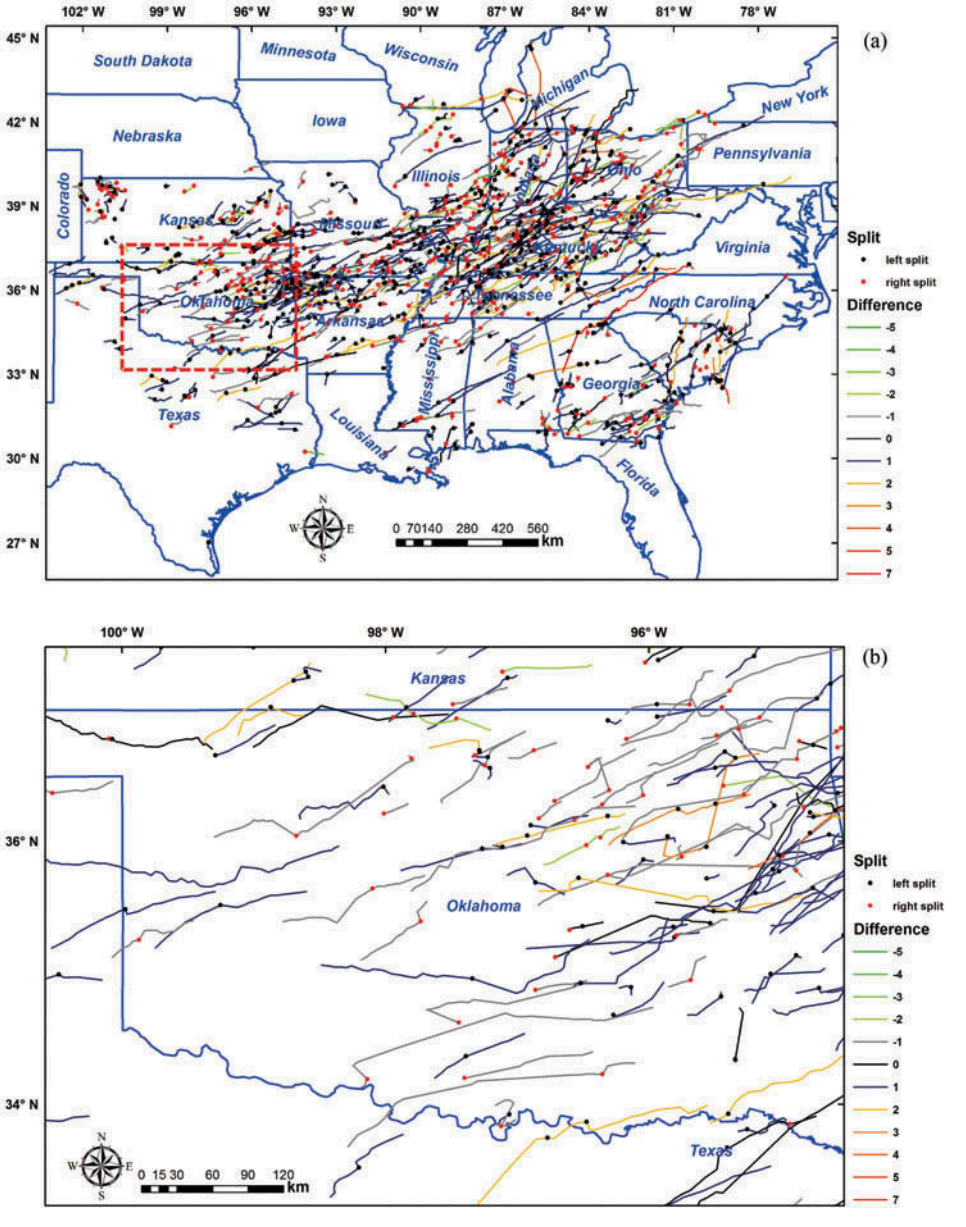

Figure 11. (a) The sides (left or right relative to storm movement direction) of the splits occurred in storm events, and the difference between the number of left splits and right splits. (b) A detailed view in the dashed box in (a).

results in Table 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.

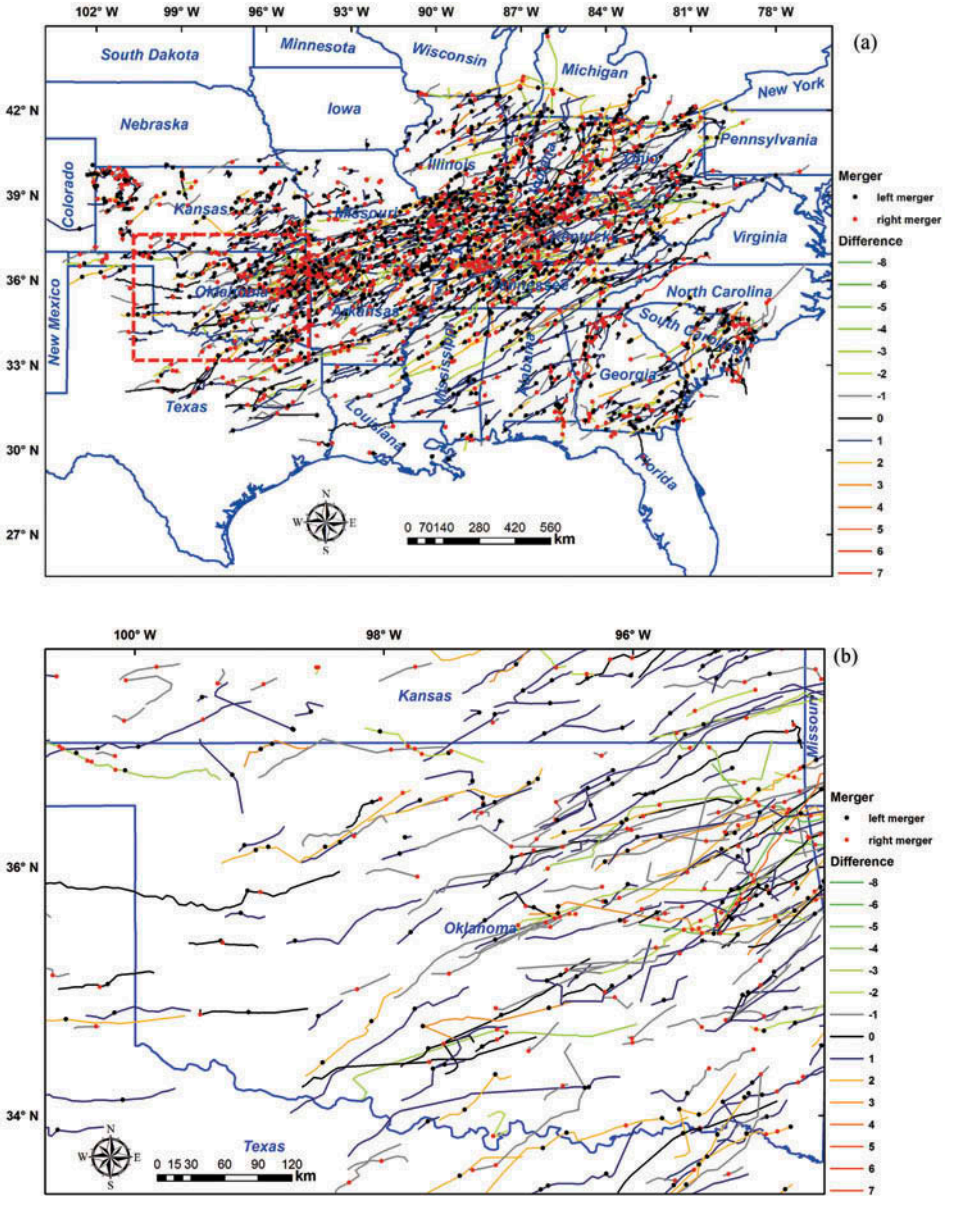

Figure 12. (a) The sides (left or right relative to storm event movement directions) of the mergers occurred in storm events, and the difference between the number of left mergers and right mergers. (b) A detailed view in the dashed box in (a).

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 life span. A directed spatiotemporal graph model was used

Table 2. The summary of split and merger appeared per storm event.

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

LS and RS represent left split and right split. LM and RM represent left merger and right merger.

Table 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}

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 by graph nodes. The interactions among storm objects were tracked using improved algorithms, and these relationships were denoted by 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 United States 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.

Acknowledgements

The authors thank the editors and the three generous anonymous reviewers for their valuable comments and suggestions which helped us greatly improve the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research is partly supported by the 'One Hundred Talents Program' of the Chinese Academy of Sciences.

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. doi:10.1111/j.1467-8306.2007.00535.x

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 (1–2), 74–91. doi:10.1016/j.jhydrol.2003.11.011

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. doi:10.1256/qj.04.100

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. doi:10.1061/(ASCE)1084-0699(2009)14:7(721)

Claramunt, C. and Thériault, M., 1995. Managing time in GIS an event-oriented approach. In: J. Clifford and A. Tuzhilin, eds. *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, eds. *Theories and methods of spatio-temporal reasoning in geographic space*. Berlin: Springer Verlag, 65–77.

Del Mondo, G., *et al*., 2010. A graph model for spatio-temporal evolution. *Journal of Universal Computer Science*, 16 (11), 1452–1477.

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. doi:10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2

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. doi:10.1175/1520-0450(2001)040<0491:MMCCSA>2.0.CO;2

Galton, A., 1995. Towards a qualitative theory of movement. In: A.U. Frank and W. Kuhn, eds. *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. doi:10.1080/13658810600965271
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. doi:10.1080/17489725.2010.537449
Han, L., *et al*., 2008. A stochastic method for convective storm identification, tracking and nowcasting. *Progress in Natural Science*, 18 (12), 1557–1563. doi:10.1016/j.pnsc.2008.06.006
Han, L., *et al*., 2009. 3D convective storm identification, tracking, and forecasting – an enhanced TITAN algorithm. *Journal of Atmospheric and Oceanic Technology*, 26 (4), 719–732. doi:10.1175/2008JTECHA1084.1
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. doi:10.1016/0004-3702(81)90024-2
Johnson, J., *et al*., 1998. The storm cell identification and tracking algorithm: an enhanced WSR-88D algorithm. *Weather and Forecasting*, 13 (2), 263–276. doi:10.1175/1520-0434(1998)013<0263:TSCIAT>2.0.CO;2
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. doi:10.1175/2008JTECHA1153.1
Lakshmanan, V. and Smith, T., 2010. An objective method of evaluating and devising storm-tracking algorithms. *Weather and Forecasting*, 25 (2), 701–709. doi:10.1175/2009WAF2222330.1
Langran, G. and Chrisman, N.R., 1988. A framework for temporal geographic information. *Cartographica*, 25, 1–14. doi:10.3138/K877-7273-2238-5Q6V
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. doi:10.1175/1520-0450(1995)034<1286:NOMAGO>2.0.CO;2
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. doi:10.1080/13658810500197652
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. doi:10.1175/2010JAMC2400.1
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. doi:10.1256/003590002320603485
Novo, S., Martínez, D., and Puentes, O., 2013. Tracking, analysis, and nowcasting of Cuban convective cells as seen by radar. *Meteorological Applications*, 21 (3), 585–595. doi:10.1002/met.2014.21.issue-3
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. doi:10.1080/02693799508902022
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. doi:10.1111/tgis.2008.12.issue-4
Smith, B.T., *et al*., 2012. Convective modes for significant severe thunderstorms in the contiguous United States. Part I: storm classification and climatology. *Weather Forecasting*, 27 (5), 1114–1135. doi:10.1175/WAF-D-11-00115.1
Stell, J., *et al*., 2011. Spatio-temporal evolution as bigraph dynamics. In: M.J. Egenhofer, et al., eds. *COSIT 2011*. Belfast, ME: Springer-Verlag, 148–167.
Thibaud, R., *et al*., 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. doi:10.1029/2008JD011093
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. doi:10.1016/j.cag.2007.01.028

Worboys, M., 2005. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, 19 (1), 1–28. doi:10.1080/13658810412331280167

Yuan, M., 2001. Representing complex geographic phenomena in GIS. *Cartography and Geographic Information Science*, 28 (2), 83–96. doi:10.1559/152304001782173718

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. doi:10.1037/0033-2909.127.1.3

Zahraei, A., *et al*., 2013. Short-term quantitative precipitation forecasting using an object-based approach. *Journal of Hydrology*, 483, 1–15. doi:10.1016/j.jhydrol.2012.09.052