WEATHER INTELLIGENCE: A GIS APPROACH TO ENRICH WEATHER DATABASES

6.5

May Yuan* and John McIntosh
The University of Oklahoma, Norman, Oklahoma

1 INTRODUCTION

On Earth, weather arguably has the longest history of observations with extensive coverage at intensive rates. Recent advances in weather observation technologies further accelerate weather data production from in-situ automatic sensors, weather radars, and meteorological satellites. Traditionally, in-situ observations are stored as text files or spread-sheets indexed by station identifiers or dates of observations. Recently, relational database technology has been employed to archive observations from ground weather networks, such as the Oklahoma Mesonet. The use of relational database technology (e.g. Oracle, Informix, and Microsoft SQL server) enables support of extracting weather data records by a set of criteria. For example, the user may select stations and days with temperature observations greater than 90°F. Satellite and radar data, on the other hand, are usually archived in binary files using various data formats, such as GRIB, HDF, NetCDF, and NIDS/NEXRAD. Data retrieval is often restricted to area and time of interest.

What has been missing in the above forms of weather data archives is information about the structures and development of weather systems, which we refer to as *weather intelligence*. Meteorologists often loop weather data or images to examine the structure and development of storms or cyclones. Looping weather data enables meteorologists to perceive the development and movements of weather systems, and therefore to gain insights of how the weather has evolved and furthermore to forecast how the evolution will continue. If we can incorporate weather intelligence into weather databases, then weather data retrieval can be based on spatiotemporal characteristics and structures of a given weather system. For example, it will be able to support the selection of weather data representing lines of convective storms that initiated from the Texas panhandle, moved eastward, and produced supercells in central Oklahoma from 2000-2002. In doing so, meteorologists can retrieve weather data based on the defined weather characteristics to assess weather forecasting, compare weather cases, and, with other environmental data, such as land cover, terrain, and soil moisture, examine regional or local influences on weather development.

We hereby propose a Geographic Information Systems (GIS) framework to enrich weather databases by incorporating weather intelligence. The use of GIS technology is necessary because information about the structure and development of weather systems cannot be derived without careful examination of spatial and temporal relations in weather data. Distinguished from the other information technologies, GIS offer a suite of functions to analyze and relate spatial data. While GIS lack functions to handle temporal data, we have developed a data structure to enable temporal indexing and tracking. The proposed GIS framework integrates both object and field (grid) representations to capture spatiotemporal characteristics of weather systems in a hierarchical structure. Using precipitation data as an example, we have developed a prototype to demonstrate the proposed GIS framework and its support for spatiotemporal query and analysis of weather systems.

In the next section, we outline some recent approaches in weather information archives to reveal the needs for alternative solutions. We then present the proposed GIS framework and a case study to

* *Corresponding author address*: May Yuan, Univ. of Oklahoma, Dept. of Geography, Norman, OK 73019-1007; e-mail: myuan@ou.edu

demonstrate weather information retrieval through the decipherment of weather intelligence embedded in massive data records. While the proposed framework is still a work in progress, we note strengths and limitations of the proposed framework and directions for future research in the final section.

2 RECENT APPROACHES IN THE ARCHIVAL AND DISSEMINATION OF WEATHER AND CLIMATE DATA

Recent information technology research has explored various ways to archive and disseminate massive data. Weather and climate is among the major areas that have received much attention because weather and climate research, such as numerical modeling, model validation, and climate analysis, often involves large quantity of data from various sources. In addition to data collected from the world-wide networks of meteorological observations and satellite measurements, there are model output data contributing to the ever increasing volumes of weather data archival.

Data mining and distributed information systems are two popular approaches to manage and explore massive data sets. The data mining approach investigates various ways to filter information from massive data sets. Most data mining methods rely on pattern detection, clustering, and machine learning techniques, to decipher information from massive data sets (Han and Kamber 2001). Of importance to effective data mining is the development of an appropriate science catalog from a data set by finding all events of interest in the data and recording relevant measurement of properties (Fayyad and Smyth 1999).

The distributed system approach offers a single entry to a wealth of data nodes online with tools for data analysis and retrieval. One example of the distributed system approach is NASA's Working Prototype Earth Science Information Partnership (WS-ESIP) program. Seasonal to interannual ESIP (SIESIP) attempts to provide data to support research in monsoons; El Nino/Southern Oscillation (ENSO); large-scale precipitation and wind patterns; the Intertropical Convergence Zone; and the Tropical Biennail Oscillation (TBO) and associated influences in the tropics and other associated geophysical climate variability. An innovative feature of SIESIP attributes to the recognition of the need that users may or may not know what exactly what data to retrieve. It allows the user to identify significant correlations and trends worthy of further analysis prior to download the data of interest (Kafatos 1998). In addition, SIESIP provides content-based browsing that enables the user to explore phenomena such as teleconnections between El Nino and vegetation cover in Africa by plotting time series and correlations, and statistically derived parameters (Li 1998).

Both approaches signify the importance of information support for content-based data retrieval, which aims to retrieve relevant data according to what phenomena to be studied. While different techniques are adopted in implementation, both approaches appear limited to a static view of phenomena because they do not go beyond data retrieval by specifying the region or location, time, and phenomena of interest. What can further advance the retrieval and access of weather/climate data is the ability to search information according to characteristics, relationships, and interactions of weather or climate phenomena in space and time. One example is retrieving observations associated with droughts that expanded throughout the southern plain over 6 months. Another example can be retrieving observations that represent storms initiating in Texas Panhandle and producing tornadoes in western Oklahoma from 2000-2002. The use of spatial characteristics, relationships, and interactions as criteria in such queries warrants the need for a GIS solution.

3 THE PROPOSED GIS FRAMEWORK

Our GIS solution takes advantage of its abilities to elicit spatial objects and uncover spatial relationships. While conventional GIS is considerably weak on temporal data handling, we have developed methods to incorporate time into GIS data models (Yuan 1999; Yuan 2000)

The proposed GIS framework builds upon the idea of representing meteorological phenomena by a set of events, processes, sequences, and zones (Fig. 1). We define these terms as follows. Phenomena are subjects of interest, such as storms, droughts and El Nino. Meteorologists define what phenomena are and develop hypotheses to explain or model their dynamics. Events are realization or occurrences of phenomena in space and time. An event takes place through multiple stages, and a transition from one stage to another constitutes a process. An event may consist of multiple processes when it occurs at multiple locations at the same time. For example, a rainfall event may be sporadic with several isolated storms. The development of a storm in space and time corresponds to a process. Within each process, mergers or splits may occur, such as two storm cells may become one or vice versa. Each merger or split marks formation of a sequence which indicates an identifiable

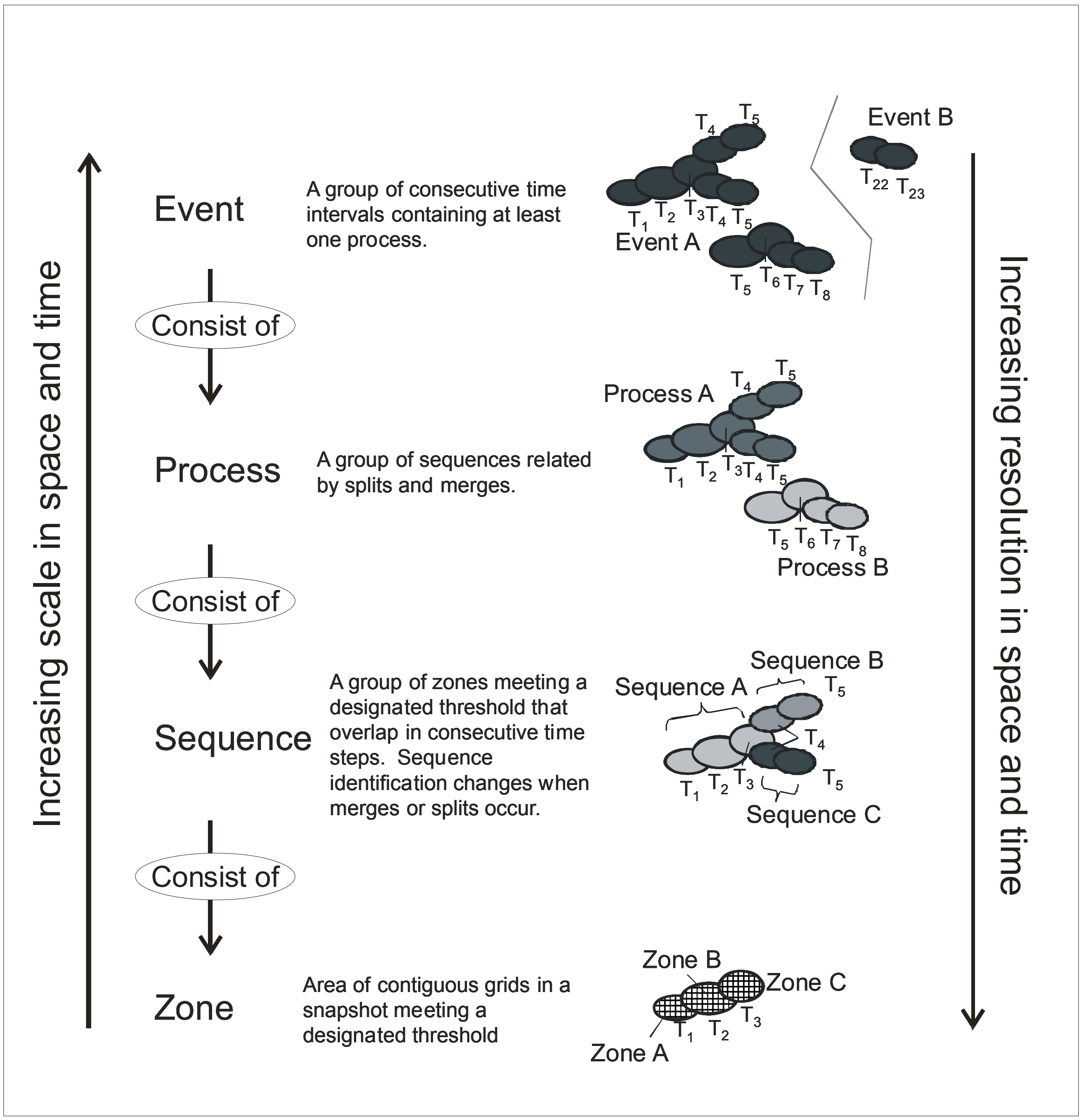


Figure 1: The proposed GIS framework of events, processes, sequences, and zones.

unit of process development. Finally, within each sequence, there are zones where the properties of interested phenomenon exhibit, such as zones of rainfall where rain is observed. From events to zones, we zoom in to the phenomenon of interest to examine its spatiotemporal characteristics and behaviors at finer resolutions. On the other hand, from zones to events, we zoom out to the phenomenon of interest to investigate its generalized and aggregated spatiotemporal properties at greater scales. Hence, the proposed framework supports spatiotemporal queries at multiple resolutions and scales. We implemented the framework in a relational data structure (Fig.2).

4 A CASE STUDY

Using digital precipitation arrays (DPA) data from the National Weather Service's Arkansas-Red River Forecast Center, we tested the information support capability of the proposed GIS framework. Of particular interest is the support for retrieval of information based on the dynamics of the focal phenomenon, such as the direction of movement and rotation of rain storms in the case study (Fig.3). The case DPA data, around 31 MB, covers the entire state of Oklahoma and portions of surrounding states from March 15, 2000 to June 15, 2000 (Fig 4). In the case study, an event object indicates observations of rainfall in

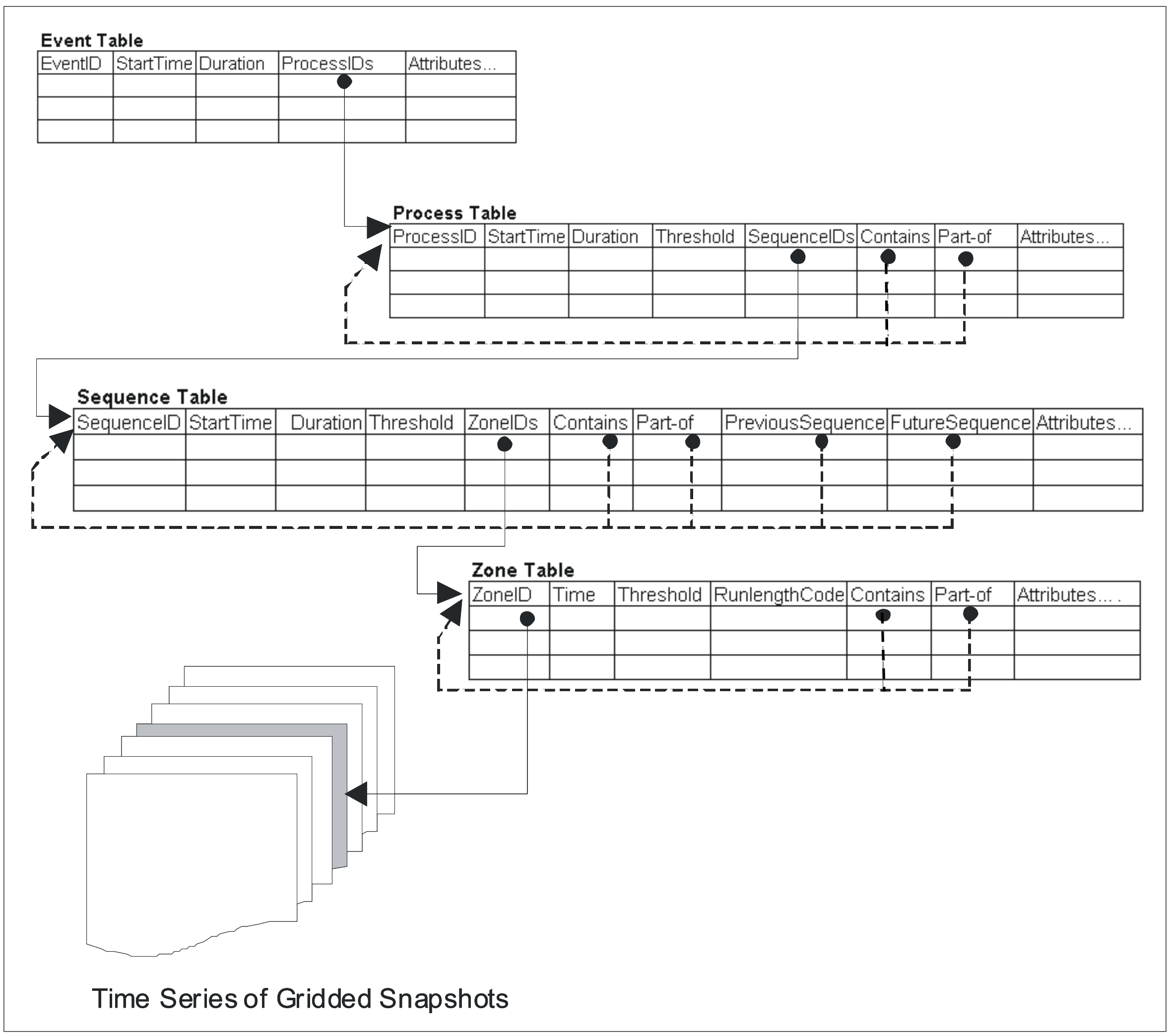


Figure 2: Data structure used to implement the proposed frameworkof events, processes, squences, and zones. The solid arrows indicate primary keys to relate tables. The dashed arrows indicate fields that encode relationships among objects within a table.

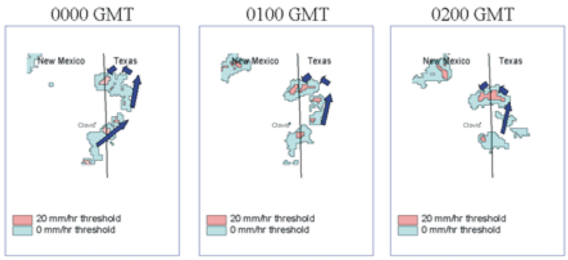


Figure 3: Movement and rotation of a rainstorm can be computed by tracing zones, sequences, and processes over space and time.

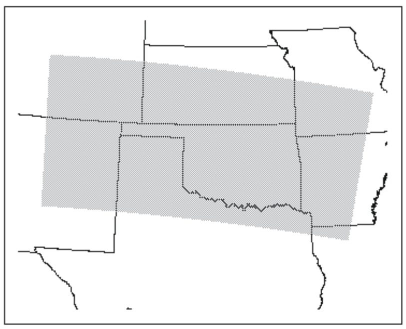


Figure 4. The coverage of DPAs from the Arkansas-Red River Forecast Center.

the study area. A process denotes the development and movement of a rainstorm in space and time. A sequence represents a segment of a rainstorm if the rainstorm has experienced mergers or splits. Finally, a zone represents a continuous rain area at a point in time; hence it corresponds to a snapshot of a rainstorm.

We implemented the proposed framework and data structure through the development of Avenue™ scripts to incorporate data extraction, query analysis, and appropriate user interfaces in the ArcView® GIS environment (Environmental System Research Institute, Redlands, California). Figure 5 exhibits the case data that was fit into the proposed data structure.

Within the proposed framework, it is straightforward to trace rainstorms at the levels of zones, sequences, and processes because each of these objects has unique identifiers (Fig. 2 & Fig. 5). Therefore, the movement of a storm, for example, can be directly retrieved by specifying its process identifier (Fig. 6). Should we be only interested in rainstorms that initiated in southwestern Oklahoma, we can select only those rainstorms that satisfy the criterion through a common GIS query by location. Retrieval of precipitation data can also be based on spatial relationships of rainstorms and a given geographic feature, such as a watershed because spatial and temporal properties of storm processes are explicitly stored in the proposed framework. Figure 7 shows an example of retrieving rainstorms that passed a given watershed. Since the current GIS technology lacks support for temporal queries, we developed a prototype system to complement its spatial query support.

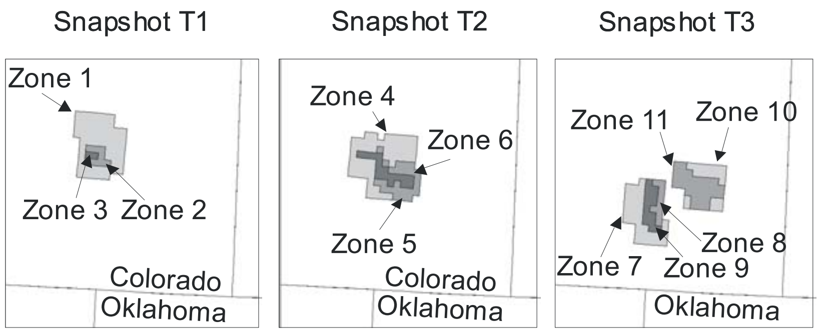
In addition, data retrieval based on rainstorm structures is supported because the proposed data structure explicitly stores how objects spatially and temporally relate to each other (Fig. 2 & Fig. 5), such as “contain”, “part-of”, “previous”, and “future”. For example, the following query will retrieve all rain areas from the zone table where rainfall sequences with precipitation greater than 20mm per hour and contain at least one sequence of higher intensity rainfall:

(min([contains].[Threshold]) > 20)

Figure 8 shows an example of such a query with precipitation started before April 1, 2000 in the case data set. By including attribute information such as elongation and orientation, the framework supports queries based on characteristic patterns similar to those used in the springtime rainstorm typologies proposed by Houze et al (1990) and Scheisser et al (1996). Including additional attributes can further refine the searches. For example, querying for low intensity processes that contain higher intensity processes that have a significant range in speeds may suggest rainstorms with rotation or direction. A sample query is illustrated in Fig. 3.

5 CONCLUDING REMARKS

The proposed framework enriches weather databases by incorporating meteorological phenomena, structures, as well as spatial and temporal relationships into data structures. Hence, it enables information retrieval by specifying spatial or temporal characteristics of the phenomena



Event Table

EventID	StartTime	Duration	ProcessIDs	Attributes...
1	1	3	1,2,3	

Process Table

ProcessID	StartTime	Duration	Threshold	SequenceIDs	Contains	Part-of	Attributes...
1	1	3	low	1,4,6	2,3		
2	1	3	middle	2,5,7	3	1	
3	1	2	high	3		1,2	

Sequence Table

SequenceID	StartTime	Duration	Threshold	ZoneIDs	Contains	Part-of	PreviousSequence	FutureSequence	Attributes...
1	1	2	low	1,4	2,3			4,6	
2	1	2	middle	2,5	3	1		5,7	
3	1	3	high	3,6,9		1,2,4,5			
4	3	1	low	7		3,5	1		
5	3	1	middle	8		3	2		
6	3	1	low	10	7		1		
7	3	1	middle	11		6	2		

Zone Table

ZoneID	Time	Threshold	RunlengthCode	Contains	Part-of	Attributes....
1	1	low	………	2,3		
2	1	middle	………	3	1	
3	1	high	………		1,2	
4	2	low	………	5,6		
5	2	middle	………	6	4	
6	2	high	………		4,5	
7	3	low	………	8,9		
8	3	middle	………	9	7	
9	3	high	………		7,8	
10	3	low	………	11		
11	3	middle	………		10	

Figure 5. The implementation of the proposed data structure using the case data set.

of interest. Using events, processes, sequences, and zones, it captures the phenomena's characteristics at various spatial and temporal scales with multiple levels of resolution. Hence, it enhances the flexibility of information access and offers opportunities to cull weather intelligence based on meteorological significance.

However, the proposed framework assumes that spatial and temporal characteristics of the phenomena of interest can be adequately captured by observations, and the phenomena are continuous distributed at the granularity of observations. The assumption serves as the basis for identification, queries, and analyses of events, processes, sequences, and zones. Should the assumption not hold, the framework cannot adequately represent the phenomena of interest.

So far, we have only experimented with data of a single theme (rainfall) from a single source (the Arkansas-Red River Forecast Center). Plans are under way to work on multiple meteorological themes (such as precipitation, temperature, and pressure) with diverse data sets (such as observatory data, radar estimates, and satellite data).

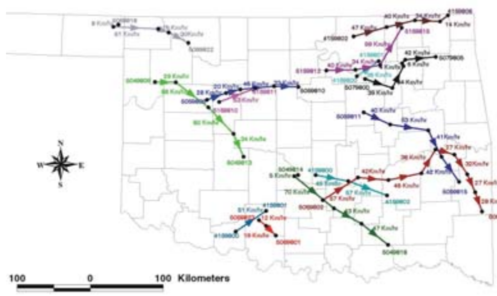


Figure 6: A query on movement of rainstorms that initiated in Oklahoma (Yuan 2001).

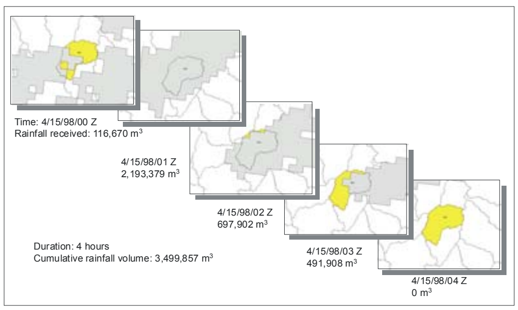


Figure 7: An example of the prototype system's response to a query on interactions between a rainstorm and a watershed. It showed how the watershed received different rainfall as the storm passed the watershed (Yuan 2001).

References

Fayyad, U. M. and P. Smyth (1999). "Cataloging and Mining Massive Datasets for Scientific Data Analysis." *Journal of Computational and Graphical Statistics* **8**(3): 589-610.

Han, J. and M. Kamber (2001). *Data Mining:Concepts and Techniques.* San Francisco, CA, Morgan Kaufmann Publishers.

Houze, R., Smull, B., and Dodge, P., 1990, Mesoscale Organization of Springtime Rainstorms in Oklahoma, *Monthly Weather Review*, 118, Pp. 613-654.

Kafatos, M. W., X.S.; Li, Z.; Yang, R.; Ziskin, D. (1998). *Information technology implementation for a distributed data system serving Earth scientists: seasonal to interannual ESIP.* Tenth International Conference on Scientific and Statistical Database Management, Pp. 210-215.

Li, Z. W., X.S.; Kafatos, M.; Yang, R. (1998). *A pyramid data model for supporting content-based browsing and knowledge discovery.* Tenth International Conference on Scientific and Statistical Database Management, Pp. 170-179.

Schiesser, H., Houze, R., and Huntrieser, H., 1995, The Mesoscale Structure of Severe Precipitation Systems in Switzerland, *Monthly Weather Review*, 123, Pp. 2070-2097.

Yuan, M. 1999, Representing geographic informatin to enhance GIS support for complex spatiotemporal queries. *Transactions in GIS*, 3(2); 137-160.

Yuan, M. 2001, Representing complex geographic phenomena with both object- and field-like properties. *Cartography and Geographic Information Science,* 28(2): 83-96.

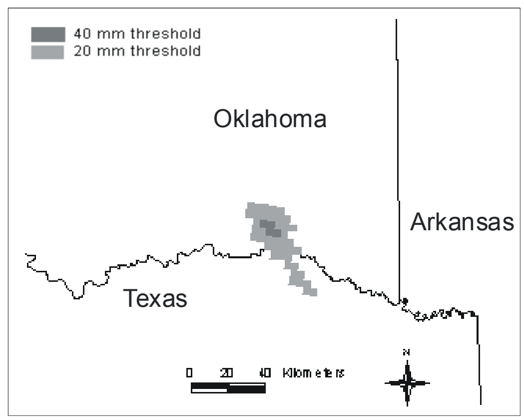


Figure 8: One of the 7 sequences returned by query described in the text. This sequence is part of a rainfall event that occurred over the border of Oklahoma and Texas on 3/26/00 and 3/27/00. The 40 mm/hour threshold sequence shown in the figure had a duration of one hour and occurred on 3/27/00 at 03 GMT.

Acknowledgement

This research was supported by the National Imagery and Mapping Agency (NIMA) through the University Research Initiative Grant (NMA202-97-1-1024) and the National Science Foundation (SBE #0074620). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the NIMA or NSF.