DATA RETRIEVAL BASED ON SPATIOTEMPORAL CHARACTERISTICS OF WEATHER EVENTS

17.14

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

1 INTRODUCTION

Effective and efficient access to weather data is critical to the development of scientific data stewardship (Trenberth et al. 2002). In today's data-rich environment, many weather and climate data servers offer advanced and attractive graphic user interfaces to allow the user query and access data by specifying the data of interest (such as mean surface temperature), sensors (such as radars or GEOS), a geographic extent (mostly by longitudes and latitudes), and time of interest (such as May 1994). Search for data based on themes, sensor of interest, geographic location, and time limits our ability to locate data of known meteorological significance. The user must have a prior knowledge about what data they are looking for and have a clear idea about a specific weather or climate event under study, for example, Oklahoma City Tornado Outbreak on May 3rd 1999.

Meteorological research based on pre-identified cases has indeed greatly enhanced scientific understanding of the atmospheric system. Complementarily, the emerging *informational science* offers an alternative scientific practice that emphasizes scientific discovery by digging into massive data sets without prior knowledge of where and when interesting events may have had occurred. In meteorology, vast amounts of data records and observations create ample opportunities for us to discover weather events with similar characteristics, their development, and how they correlate in space and time. However, without efficient information access tools, the majority of weather observations will remain out of reach.

Much effort has been devoted to improve meteorological data accessibility. One significant project is Thematic Real-time Environmental Distributed Data Services (THREDDS) to provide efficient real-time access and effective tools for data analysis from distributed servers (Domenico et al. 2002). The core of THREDDS is the use of Publishable Inventories and Catalogs (PICat) from distributed data servers to enable data visualization and analysis client sites. Another emphasis of THREDDS is meta-data that provides actual field names, coordinate systems, units, and other data descriptors to facilitate data access.

Another distinct project on the venture to ease data access and facility scientific discovery is NASA Earth Science Information Partnership (ESIP) program, in which Seasonal to Interannual ESIP (SIESIP) centers on data that support research in monsoons, El Nino/Southern Oscillation (ENSO), large-scale precipitation, and wind patterns, the Intertropic Converge Zone (ITC), and the Tropical Biennial Oscillation (TBO) and associated influences in the tropics and other associated geophysical climate variability. Recognizing that the user may or may not know exactly what data to retrieve, ESIP has functions to aid the user to identify significant correlations and trends worthy of further analysis before downloading the data (Kafatos 1998). Most notably, 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). This is indeed a marked advance in the informational science in that it provides the user information about how data from multiple regions and periods correlate so

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

that the user can judge if the data will be worthy of downloading for further analysis.

With the marching strikes made in the previous informational studies, further progress can be made in the dimension that is even more directly linked to human conceptualization on scientific inquiry and discovery. For example, a scientist may be interested in how the jet stream relates to the development of severe storms in the southern plain. A convenient way to get the data necessary to answer the research question is to pose a query that will retrieve data corresponding to the two weather features. An example will be to ask an information system to retrieve temperature, winds, precipitation, and sounding data that are associated with the jet stream and severe storms developed in the southern plain during a specified period. We refer this kind of data search as process-based because the user specifies the processes of interest and/or the correlation of interest to retrieve relevant data. The rest of our paper describes the approach that we propose to enable such data retrievals.

2 THE PROPOSED APPROACH

Central to our approach is the idea that geospatial information systems, including the ones for meteorological information, should have events and processes explicitly represented in databases. In other words, there should be built-in database objects (or data objects) that correspond to a weather event, such as a storm, and its development. In doing so, spatiotemporal characteristics of individual weather events can be queried and analyzed. For example, we will be able to pose a query to select all storms that initiated in southwest Oklahoma, moved northeast, and developed hails and tornadoes in central Oklahoma. Another example is to retrieve sea surface temperature data associated with El Nino during droughts in the southern plain. In this case, we need an information system that has “El Nino” and “drought” represented as data objects. The system will first select drought objects that occurred in the southern plain, identify the time periods of the selected droughts, and then retrieve El Nino objects that occurred within proximity of the selected time periods. We believe that such a system will release meteorologists from tedious and laborious search for data of interest and enable them to focus on research questions of interest.

Development of digital representation of weather events and their development is not a trivial task because it requires the integration of semantic properties (identity and attributes), spatial properties (geometry, location, and topology), and temporal properties (time of occurrences and time of observations). Figure 1 shows an example of a storm moving from northern Oklahoma and moving to the southeast Oklahoma. In order to represent the storm as a database object, we need to have the database object account for the size, shape, and location of the storm through time (from 7Z to 15Z).

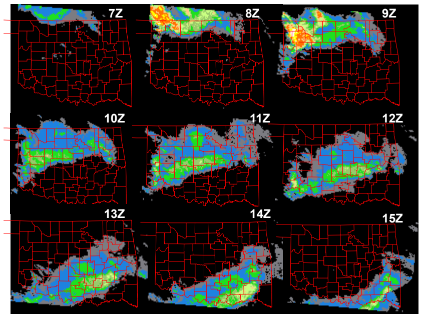


Figure 1: An example of a storm moving through Oklahoma.

To meet the needs, we develop a framework of composite database objects to link geometry and location through a hierarchy of spatial and temporal aggregation (Figure 2).

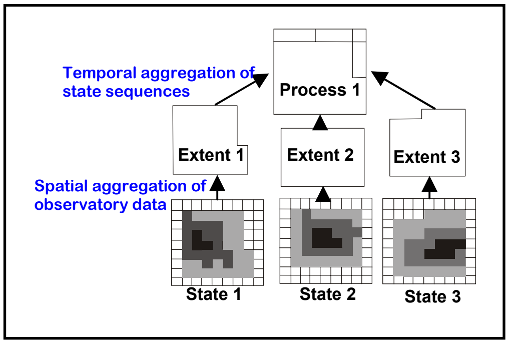
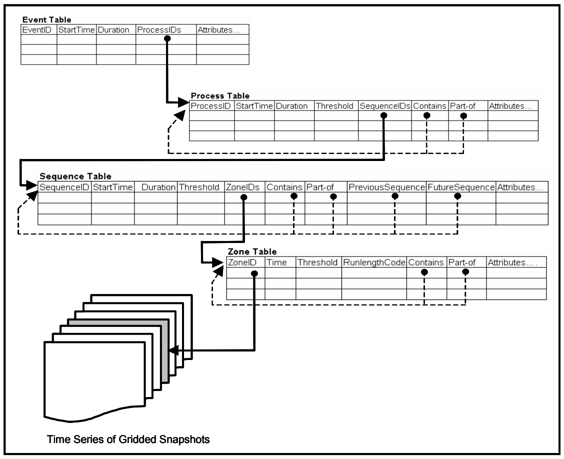

Temporal aggregation of state sequences
Process 1
Extent 1
Extent 2
Extent 3
Spatial aggregation of observatory data
State 1
State 2
State 3

Figure 2: Spatial and temporal aggregation of observations to form composite data object that can count for the change of geometry and location over time.

In addition, proper attribute tables are built to record semantic descriptors of the composite object. For example, if a composite object represents a storm, then its attribute table will have information about the total damage from the storm, magnitude of the storm, and so on. However, in our implementation, the hierarchy is expanded to incorporate the need for intermediate objects that form through different levels of spatial and temporal aggregation (Figure 3). The hierarchy starts with observations of snapshots, such as a radar image. Storms or weather events of interest occurred in the observations are then identified and form a data object "zone" that represents the geographic extent, geometry, and location of the identified event at that time. A zone table is built to stored attribute information about individual zones and linkages to their upper level data objects "sequences." A sequence is a temporal lineage of zones. It represents how a storm cell moves from one location to another over time. Its attribute table stores information about the movement and its development during the move accordingly. When two storm cells merge to one or a new storm cell spins out of an existing one, we keep track of the process and record the information in the process table. Finally, all processes are considered as a part of a weather event, such as a tornado outbreak or a drought. Descriptors of individual events are recorded in the event table.



Event Table

EventID	StartTime	Duration	ProcessIDs	Attributes...

Process Table

ProcessID	StartTime	Duration	Threshold	SequenceIDs	Contains	Part-of	Attributes...

Sequence Table

SequenceID	StartTime	Duration	Threshold	ZoneIDs	Contains	Part-of	PreviousSequence	FutureSequence	Attributes...

Zone Table

ZoneID	Time	Threshold	RunlengthCode	Contains	Part-of	Attributes....

Figure 3: The data structure

3 DEMONSTRATIVE EXAMPLES

We use hourly digital precipitation arrays for March 15 to June 15, 2000 from the Arkansas Red River Basin Forecast Center as a case study. Based on the data and our proposed framework, we are able to implement the following queries.

- Find storms occurring at certain time and duration. We develop a query builder dialog to support queries based on the modeled relationships and object attribute values (Figure 4)

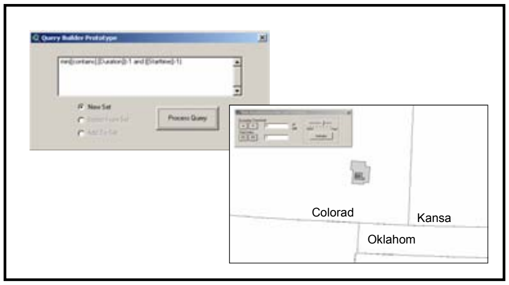


Figure 4: An example of queries on a storm.

- Find storms with similar change from T1 to T2. We develop measures to determine the similarity of storm development based on geometric characteristics and individual distributions. Based on the measures, our system is able to find storms that show similar change (Figure 5)

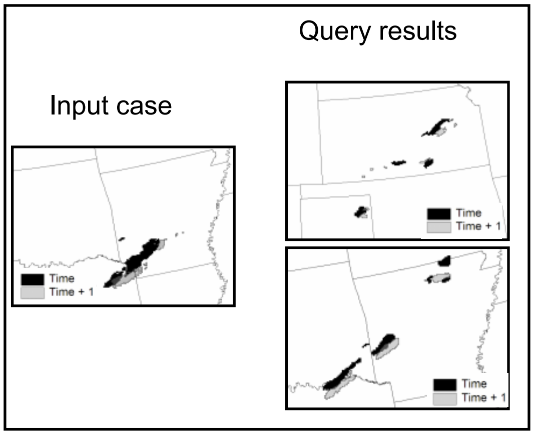
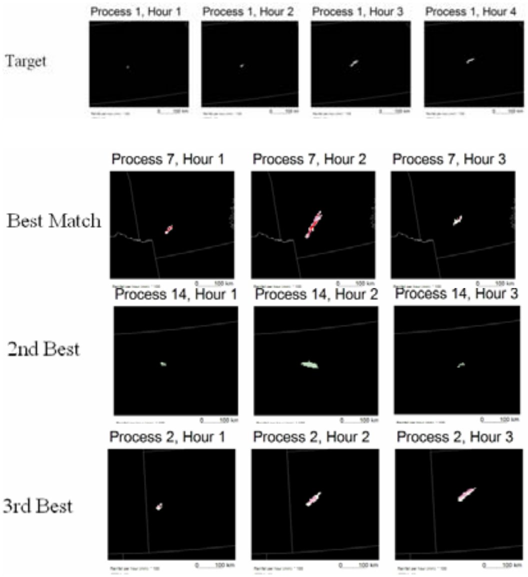


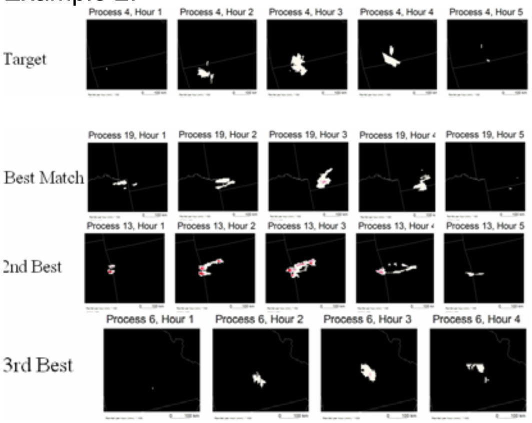
Figure 5: An example of query on storms with similar change from two observations.

- Find storms with similar development (from initiation to dissipation). We develop a time warping technique to examine changes along the development of individual storms and measure their difference (similar to gene-sequencing analysis). Below are examples of target storms which are submitted to the system and matched storms which are returned by the system as storms exhibit similar characteristics and behaviors in space and time. In this test, only 20 storms were considered for comparison. Notably, our technique does not require storms with the same duration to be considered as similar. Similarity is mainly judged by a combination of geometry, elongation, movement, distribution (if more than one exists), and evolution (such as split or merge).

Example 1:



Example 2:



4 CONCLUDING REMARKS

Our research emphasizes the need for event-based query and data retrieval to facilitate scientific discovery. We believe the importance of information support that can go beyond how data are collected to how data are related to real-world events and processes. Conventionally, meteorological information retrieval is based on sensors, geographic extents, properties (such as temperature), and time. The user needs to have prior knowledge about a region and time of significance before the search. Alternatively, we argue for the need of information support that forces the user to focus on (1) what events may be of interest or (2) how events may be related in space and time. Data retrieval will then be based on identified events or types of events.

We proposed a data framework and developed a prototype for proof of concepts. We will soon post our prototype in the world wide web for free downloads. While the data set we used to test the proposed data framework is small, the prototype demonstrates the usefulness of the proposed data framework on data retrieval according to meteorological processes of interest. We continue refining the data framework and investigating scaling issues related to very large data sets.

References

Domenico, B., J. Caron, E. Davis, R. Kambic and S. Nativi (2002). "Thematic Real-time Environmental Distributed Data Services (THREDDS): Incorporating Interactive Analysis Tools into NSDL." Journal of Digital Information **2**(4).

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.

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.

Trenberth, K. E., K. R. Thomas and T. W. Spence (2002). "The Need for a Systems Approach to Climate Observations." Bulletin of the American Meteorological Society **83**(11): 1593-1602.

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.