Title: **Toward Spatio-temporal Models of Biogeophysical Fields for Ecological Forecasting**

Authors: Geoffrey M. Henebry[1]*, Tony Fountain[2], Jan Chomicki[3], and K. Jon Ranson[4]

Address: [1]Center for Advanced Land Management Information Technologies (CALMIT)
University of Nebraska-Lincoln
113 Nebraska Hall
Lincoln, NE 68588-0517
Phone: 402-472-6158
Fax: 402-472-4608
Email: ghenebry@calmit.unl.edu

[2]San Diego Supercomputer Center
University of California-San Diego
9500 Gilman Drive
La Jolla, CA 92093-0505
Phone: 858-534-8374
Fax: 858-534-5152
Email: fountain@sdsc.edu

[3]Department of Computer Science and Engineering
University at Buffalo, The State University of New York
201 Bell Hall
Buffalo, NY 14260-2000
Phone: 716-645-3180 x 103
Fax: 716-645-3464
Email: chomicki@cse.buffalo.edu

[4]Biospheric Sciences Branch
Code 923
NASA GSFC
Greenbelt, MD 20771
Phone: 301-286-4041
Fax: 301-286-1757
Email: jon.ranson@gsfc.nasa.gov

TYPE: SHORT PAPER

DEMO?: NO

* Corresponding author

CONTEXT We are now in an era of intensive earth observation. Orbital platforms generate myriad remote sensing datastreams across a range of spatial, temporal, spectral, and radiometric resolutions. The number and variety of "eyes in the skies" are scheduled to increase significantly over the next few years. This veritable data deluge necessitates new ways of thinking about transforming remote sensing data into information about ecological patterns and processes. These datastreams hold the promise for environmental decision support. Effective use of remote sensing datastreams to characterize and monitor landscape dynamics requires analysis of the temporal variations in spatial patterns. We can distinguish four main phases in the analysis of image time series (Henebry and Goodin 2002): (1) change detection—perceiving the differences; (2) change quantification—measuring the magnitudes of the differences; (3) change assessment—determining whether the differences are significant; and (4) change attribution—identifying or inferring the proximate cause of the change. There is a critical need for theories and tools that will enable efficient and reliable characterization of spatio-temporal patterns contained in image time series. Such tools ought to be based on ***ecological expectations*** of land surface dynamics, analogous to climatological expectations. Ecological expectations would summarize across specific regions the typical temporal development of spatial pattern in biogeophysical fields. To make ***ecological forecasting*** an operational possibility (Clark et al. 2001; Maier et al. 2001), there is need for computational strategies to establish and to update complex spatio-temporal baselines that will enable prediction of the usual and the identification, quantification, and assessment of the unusual. Beyond computational considerations, there is an urgent need for environmental scientists to dialogue with computer scientists to develop effective and robust spatio-temporal models of biogeophysical fields for database and datamining applications aimed at the investigation of these baselines and their associated anomalies in datastreams. One objective in our Biodiversity and Ecosystem Informatics (BDEI) project is to sponsor an international cross-disciplinary workshop to examine the challenges facing the development and implementation of the next generation of spatio-temporal data models. This workshop is to held 8-10 April 2002 at the San Diego Supercomputer Center. We shall report the results at the dg.o conference. Here we offer some background on principal issues and a list of questions to be examined at the workshop.

THE CHALLENGES OF GEOSPATIAL DATA IN TIME Burrough and Frank (1995), inquiring about the generality of GIS implementations, observed an unresolved and possibly irresolvable tension between the universal data models that computer scientists seek and the *ad hoc* data models that GIS practitioners use to address specific problems. They further identified three major groups of GIS users: managers of defined objects (e.g., cadastres, utilities, facilities management); planners and resource managers (e.g., multi-attribute evaluation & decision-making); and space-time modelers (e.g., environmental scientists broadly construed). What Burrough and Frank (1995) discovered was a profound conceptual disconnect in the GIS community between the units of analysis and the baseline models employed by different disciplinary subgroups. Current GIS implementations are *not* generic and they do not adequately support space-time modeling (Burrough and Frank 1995; Couclelis 1999; Peuquet 2001).

Inclusion of time in GIS is not as straightforward an exercise as might be expected (Ott and Swiaczny 2000). A major source of difficulty stems from how the increased dimensionality of the data affects what can be assumed about the data. Consider an unordered list, the simplest database structure. It is a collection of zero-dimensional data, database records that lack spatial or temporal relationships with other records. While this structure is easy to implement and enables efficient querying about the relationships between records, it can permit inferences about relationships that are nonsensical when viewed within the broader context of the data. Temporal databases introduce an explicit, unidirectional, one-dimensional structure to the data. The "arrow of time" makes temporally oriented queries and logical inferences possible (Snodgrass 1992; Chomicki and Toman 1998). Spatial databases represent spatial relationships as locations (raster/fields) and/or as entities (vector/objects). Although coordinate systems supply topology, there is no *a priori* ordering of the directionality of causation in space. This has the important consequence of requiring the user to inform the database about the flows of influence among spatially ordered data. The user must specify a model of spatial relationships in order to make meaningful queries. For example, many GIS implemenations have a module that introduces the influence of gravity into the database topology in order to analyze drainage patterns. While topological relationships indicate who is the neighbor of whom, additional information is required to know who are the *effective* neighbors. Different processes can have different effective neighborhoods or corridors at different scales. The addition of time into a spatial model further complicates the issue of influence and places more responsibility on the user to identify relevant neighborhoods and to supply meaningful orderings (Henebry 1995; Peuquet 2001; Henebry and Merchant 2002).

Databases that contain data with both spatial and temporal dimensions need to support sophisticated spatio-temporal queries. For example, let's consider some spatio-temporal queries that could be asked of a time series of vegetation index composites (e.g., AVHRR or MODIS NDVI) concerning the "green wave" that accompanies the onset of spring in the northern temperate and boreal zones (Myneni et al. 1997; Schwartz and Reed 1999; Cayan et al. 2001; Shabanov et al. 2002):

[Q1] ***Where did spring arrive earlier this year than last?*** Assume SP is a predicate with the following meaning: SP(Y, P, x,y) is true if the database contains information that indicates Spring is present in the year Y, instant P (measured in days), and location x,y. Thus, SP assumes that data have been processed to bring a series of remotely-sensed images into a series of images of a biogeophysical variable (here, fractional vegetation cover). The query [Q1] can then be expressed by the logical formula:

$$\exists \mathbf{Pa}.\ \exists \mathbf{Pb}.\ \neg\mathbf{SP(2000, Pa, x,y)} \wedge \mathbf{SP(2000, Pa+1, x,y)} \wedge \neg\mathbf{SP(2001, Pb, x,y)} \wedge \mathbf{SP(2001, Pb+1, x,y)} \wedge \mathbf{Pb<Pa}.$$

[Q2] ***Was the area in which spring arrived earlier than the previous year greater than the area in which spring arrived later than the previous year?***

[Q3] ***Where is spring likely to arrive earlier next year than this year?***

The first query [Q1] requires a point-wise comparison of relationships between different time instants. The second query [Q2] additionally requires the ability to do spatial aggregation and its translation to a logical formula would require non-standard constructs. The third query [Q3] moves beyond querying about what has been observed to ask for a forecast, a prediction about the future based on prior and current knowledge. This is the kind of predictive query could be addressed using machine learning/data mining techniques *informed by domain expertise* and applied to the image time series and its derived products. These sorts of quantitative results are what biogeophysical scientists want to elicit from time series of images and/or GIS data layers.

AN UNORDERED LIST OF QUESTIONS TO BE EXAMINED AT THE WORKSHOP

- Which kinds of spatio-temporal questions are of interest? What do these data look like? What do investigators want to get out of these data?
- Which types of data mining techniques are useful for analyzing biogeophysical data for ecological forecasting?
- Which types of data mining algorithms are appropriate for biogeophysical data?
- Which kind of spatio-temporal data model is appropriate to enable/facilitate the queries of interest? How generic is this model?
- How should querying to spatio-temporal vs. spatio-spectral vs. spectral-temporal slices from image time series be handled?
- How should domain expertise be integrated into advanced query processing and mining?
- Which interface tools are needed to enable the scientists to accomplish their goals with the database and mining tools.
- Which types of computational infrastructure would be useful for efficient querying/mining/analysis? How should tasks be mapped onto hardware and software resources; in particular, what are the key connections between the database and the other analysis/mining routines.
- Which operators of the relational algebra should be available to formulate spatio-temporal queries? Should operator nesting be restricted? Should new operators be introduced? What about relational calculus and SQL?
- Which query evaluation and optimization techniques may be suitable in this context?
- What is impact of missing and/or uncertain data on efficiency and accuracy of pattern characterization/query processing?
- How achieve spatio-temporal queries that imply change detection and/or change quantification?
- How to handle missing or poor quality data in spatio-temporal queries?
- How to quantify/evaluate the success of a complex query?
- How to model propagation of error or uncertainty in complex queries and/or mined relationships?
- How much statistical analysis can be productively moved into a RDBMS?

ACKNOWLEDGEMENTS We are grateful for support from BDEI NSF #0131937.

REFERENCES

Behnke, J., E. Dobinson, S. Graves, T. Hinke, D. Nichols, P. Stolorz, and P. Newsome. 2000. Final Report: NASA Workshop on Issues in the Application of Data Mining to Scientific Data. Available online at www.cs.uah.edu/NASA_Mining.

Burrough, P.A., and Frank, A.U. 1995. Concepts and paradigms in spatial information: are current geographical information systems truly generic? *International Journal of Geographical Information Systems* 9:101-116.

Cayan, D.R., Kammerdiener, S.A., Dettinger, M.D., Caprio, J.M., and Peterson, D.H., 2001, Changes in the onset of spring in the western United States. *Bulletin of the American Meteorological Society* 82:399-415.

Chomicki, J. and D. Toman. 1998. "Temporal logic in information systems." In Logics for Databases and Information Systems, edited by J. Chomicki and G. Saake, 31-70. Boston: Kluwer.

Couclelis, H. 1999. Space, time, geography. In Geographical Information Systems. Volume 1 Principles and Technical Issues 2/e, edited by P.A. Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind, 29-38. New York: Wiley.

Clark, J.S., and 16 others. 2001. Ecological forecasts: An emerging imperative. *Science* 293:657-660.

Henebry, G.M. 1995. Spatial model error analysis using autocorrelation indices. *Ecological Modelling* 82:75-91.

Henebry, G.M., and D.G. Goodin. 2002. Spatio-temporal analysis of landscape dynamics from image time series. *Photogrammetric Engineering and Remote Sensing.* In review.

Henebry, G.M., and J.W. Merchant. 2002. Geospatial data in time: limits and prospects for predicting species occurrences. In Predicting Species Occurrences: Issues of Scale and Accuracy, (J.M. Scott, P.J. Heglund, M. Morrison, editors). Island Press, Covello, CA. Chapter 23. pp 291-309.

Maier, D., E. Landis, J. Cushing, A. Frondorf, A. Silberschatz, M. Frame, and J.L. Schnase. 2001. Research Directions in Biodiversity and Ecosystem Informatics. Report of an NSF, USGS, NASA Workshop on Biodiversity and Ecosystem Informatics (NASA Goddard Space Flight Center, June 22-23, 2000, Greenbelt, Maryland). 30 pp.

Myneni, R. B., C. D. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani. 1997. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* 386:698-702.

Ott, T., and F. Swiaczny. 2000. Time-Integrative Geographic Information Systems: Management and Analysis of Spatio-Temporal Data, Springer-Verlag, New York. 234 pp.

Peuquet, D.J. 2001. Making space for time: Issues in space-time data representation. *Geoinformatica* 5(1):1-32.

Schwartz, M.D., and B.C. Reed. 1999. Surface phenology and satellite sensor-derived onset of greenness: an initial comparison. *International Journal of Remote Sensing* 20(17):3451-3457.

Shabanov, N.V., L. Zhou, Y. Knyazikhin, , R.B. Myneni, and C.J. Tucker. 2002. Analysis of interannual changes in northern vegetation activity observed in AVHRR data from 1981-1994. *IEEE Transactions on Geoscience and Remote Sensing* 40:115-130.

Snodgrass, T.R. 1992. Temporal databases. In Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, edited by A.U. Frank, I. Campari, and U. Formentini, 22-64. New York: Springer-Verlag.