How do we represent data as models and models as data? Matt Becker, CSU Long Beach

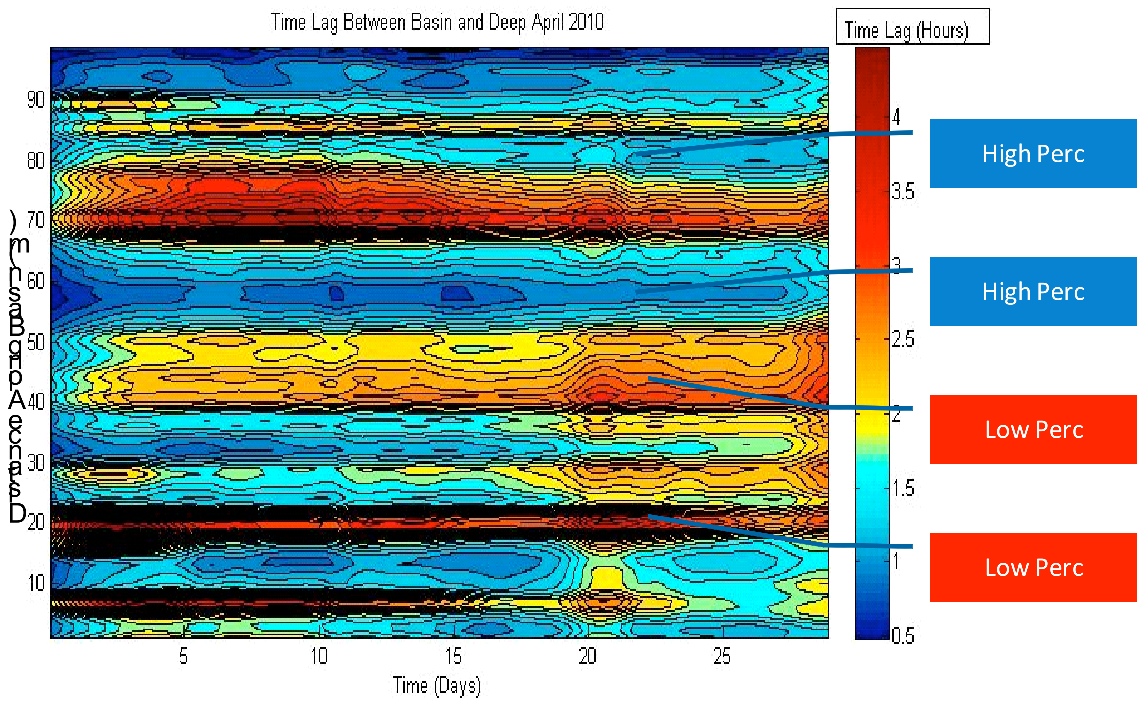
Field data are common in the natural sciences. Temperature, wind speed, hydraulic head, and ocean current velocity are all examples of phenomena that are spatially continuous and, therefore, must be represented as fields rather than objects. Geographic Information Systems (GIS) have become the standard for storing, organizing, and displaying georeferenced natural science data, but they are not adept at handling fields. GIS databases are organized around objects that are assigned attributes. Field phenomena must be gridded or "rasterized" such that continuous information is converted to a group of discrete measurements (pixels). Raster data are problematic for GIS in that they are scale dependent, storage intensive, and are not directly related to geographic entities.

One way to relate fields to an object is through a predictive or interpolative model. A kriged contour surface is a simple model, for example. However, all models carry assumptions embedded within them (e.g. a kriged field is based upon a variogram model; linear, spherical etc.). Often models are not even recognized in data (e.g. stream discharge is related to stage via a rating curve model). Databases are not typically designed to carry metadata that represents models of data, and the underlying assumptions. A model is a hypothesis and the data may be used as a hypothesis test via goodness of fit.

How should we store and communicate models of data, whether for field, vector, or scalar data? Being able to track the application of models during data workflow is critical to remove bias from the dataset that is applied during interpretation. I have been particularly interested in represent models as objects in GIS systems. We stored and manipulated numerical groundwater flow models in a GIS using Analytic Elements as vector objects in ArcGIS, for example [*Becker and Jiang*, 2007; *Fredrick et al.*, 2006; *Rabideau et al.*, 2007]. However, I think we did a very poor job dealing with metadata in a practical way.

At a different scale, we are currently using fiber optic distributed temperature sensing (FODTS) to measure temperature every meter along a fiber optic cable every half-hour along about 2 kilometers of cable. That is about 10^5 measurements per day or 10^7 measurements per year. The data storage is, therefore, approaching raster data in size. Should these data be treated as raster (e.g. temperature at a time and distance pixel?) or scalar and how can we reduce these data to something more manageable? Can we "smooth" in time and space (e.g. see figure below)?

I'd be interested in discussing (1) how models of data should be handled in networked datasets and (2) how data storage, display, and communication can be made more efficient through the use of model representations of spatially and temporally complex data.



Cited References

Becker, M. W., and Z. Jiang (2007), Flux-based contaminant transport in a GIS environment, *Journal of Hydrology*, *343*(3-4), 203-210.

Fredrick, K. C., M. W. Becker, L. S. Matott, A. Daw, K. Bandilla, and D. M. Flewelling (2006), Development of a numerical groundwater flow model using SRTM elevations, *Hydrogeology Journal*, *10.1007/s10040-006-0115-3*.

Rabideau, A. J., J. R. Craig, W. Silavisesrith, K. Fredrick, D. M. Flewelling, I. Jankovic, M. W. Becker, K. Bandilla, and L. S. Matott (2007), Analytic-element modeling of supraregional groundwater flow: Concepts and tools for automated model configuration, *Journal of Hydrologic Engineering*, *12*(1), 83-96.