Thematic Background Note for UGEC-NASA Workshop

Geoff Henebry [geoffrey.henebry@sdstate.edu], **South Dakota State University** **27MAR2011**

> *Data*: The support of applications at different scales is supported by the growth of Spatial Data Infrastructures, geo-portals, and private sector initiatives (*e.g.*, Google Earth, Microsoft Virtual Earth, etc.) has resulted in a massive increase in geographical data availability globally at multiple scales. This growth has not been fully coupled by an increase of knowledge to support spatial decisions. Spatial analytical techniques and geographical analysis and modeling methods are, therefore, required to analyze data and to facilitate the decision process. With cities, conceptualized as a concentration of people, it is most striking to find coherence between land use and socio-demographic as well as socio-economic parameters. The statistical analysis of census data infers information on the human usage of the land, the human exposure to potential hazards in the city, and the configuration of each neighborhood indicating the urban quality of life. For example, combining maps of socio-demographic features with land use maps provides information on gender and age distribution connected with proximity to urban green/open spaces, income and building density, or water consumption and level of provision of infrastructure. In this context URS helps by providing spatial information where linked social and physical indicators explain the interrelations between ecological conditions and socio-spatial development.
>
> In terms of forecasting urban growth, a majority of urban growth models focus on cities in high income countries where data are typically widely available. However, forecasts suggest that in the next two decades most of the urban growth will occur in low-income countries, such as in Asia and Africa. In these cases, data may be nonexistent, incomplete, inaccurate, unreliable, or all the above. One goal is to identify the challenges with developing models in these data-poor contexts? How do we improve on issues regarding data availability and accuracy in these areas? What linkage may exist with remote sensing technology to fill this gap?

This description of the data theme by the workshop organizers raises some critical questions and provides an excellent starting point for this essay. At the risk of being provocative, let me ask another: ***What are the appropriate units of analysis for monitoring, modeling, and forecasting urban growth?*** The phrase "units of analysis" refers here not to the units of specific measurement systems, but rather to the conceptual entities that are subject to measurement and analysis. Tools and techniques commonly used to represent and manipulate geospatial data carry strong assumptions as to what constitute the units of analysis. Geospatial entities include axiomatic geometric objects (*e.g.*, points, lines, polygons, polyhedra) that are located within a spatial reference system. But geospatial entities can also be synthetic geometric objects derived from sensor systems.

An expected response might be, in the context of URS data, the pixel. Let me assert that individual pixels are neither appropriate nor sufficient! This is not a novel position (*cf.* Fisher 1997; Cracknell 1998). Fisher's title conveys the problem: the pixel as a snare and a delusion. Why? Pixels are—almost always—heterogeneous objects. They do indeed convey measurements associated—often loosely—with their nominal geospatial and temporal coordinates. However, it is important not to confuse the packaging of the observations with the things of interest that motivate the observations (Strahler *et al.* 1986), particularly when those things are observed at different scales. A related problem arises with socio-economic data packaged in irregular tessellations: the Modifiable Areal Unit Problem, *cf.* Openshaw (1984), Arbia (1989).

There are no natural *a priori* spatial units. We impose units by our observational processes. Thus, delineations between patches are arbitrary and may be imprecise in location, transitory in duration, and irrelevant to underlying processes of interest. Further, there is no *a priori* ordering of the directionality of causation in space comparable to the "arrow of time." While topological relationships indicate who is neighbor of whom, more information is required to know who the <u>effective</u> neighbors are. This requires the user to inform the geospatial database about the flows of influence among spatially ordered data. Different processes can have different effective neighborhoods at different scales.

Further, in urban remote sensing the images or pictures are themselves *not* the endpoint for scientific analysis; rather, what is of scientific interest is the dynamic of pattern and process that the pictures portray. Consider the analogy of sparse sampling of individual frames or even frame sequences from a movie. One level of analysis could aim at reconstructing motion from these data, but a more sophisticated analysis could aim at *reconstructing the plot*. If we are to delve into the image archives with the aim of advancing our understanding of cities and their regional penumbrae, then we need to advance a program of reconstructing plots, comparing plots, characterizing typical plots, and identifying unusual plots, as well as interesting deviations from typical plots.

Some plots relevant to urbanization and global environmental change include (i) growth and decay patterns of human settlements in various kinds of resource environments and (ii) responses of urbanized areas to disturbances—both anthropogenic and natural—on a variety of timescales, among many others. We can observe aspects of these phenomena from orbital platforms by sensing *reflected solar* radiation (visible to middle infrared), *emitted terrestrial* radiation (middle infrared through thermal infrared and microwaves), and *backscattered anthropogenic* radiation (RaDAR, LiDAR).

The process of observation in remote sensing is a more subtle issue than first it may appear. There is the general problem of observability in a strictly technical sense: Is it possible to sample adequately the phenomenon of interest? Given the loosely coupled and contingent nature of ecological (and even socio-economic) relationships, this question must be addressed at multiple scales (Allen and Hoekstra 1992). But multi-scale sensing has rarely been practiced, though it has often been demonstrated.

In considering the future of spatial analysis and GIS, Openshaw (1994) argued for a "concepts-rich approach to spatial analysis, theory generation, and scientific discovery in GIS using massively parallel computing." He diagnosed a source of malaise that continues to affect the spatial analysis community and then pointed to a possible remedy:

> Pattern searching is not the same as hypothesis testing because there is no relevant null hypothesis. This point was lost on the original quantitative geographers [during the 1970's]. ... [They] failed to develop a statistical theory of spatial analysis as distinct from providing examples of statistical methods being applied to spatial data in search for largely aspatial patterns. **The danger now is that the same mistake will be repeated 20 years later in the GIS era by a failure to appreciate that spatial patterns are themselves geographic objects that can be recognized and extracted from spatial databases.** [Emphases added.]

A key notion here is that spatial and, by extension, spatio-temporal patterns are *observable entities* and *appropriate units of analysis*. Here is a lever by which to build a theoretical framework for spatio-temporal analysis of image time series. To date, theory development for spatial-temporal analysis has been hampered by lack of a suitable framework for identification and quantification of spatio-temporal patterns. Numerous metrics have been proposed for quantifying spatial properties of image data; however, scant attention has been paid to the effective use of these metrics for capturing or summarizing spatio-temporal dynamics, whether in urbanized areas, croplands, grazinglands, or wildlands.

Openshaw's critique also points to the problem of baseline models: "...because there is no relevant null hypothesis." The testing of null hypotheses is one particular form of using neutral models to compare and contrast phenomena. Neutral models are touchstones. They serve a crucial role in scientific investigation by providing archetypes of expectation that guide the development of theory, the design of experiments, and the collection, analysis, and interpretation of data. The most powerful inferential tools in traditional probability theory rely upon the concept of zero-dimensional randomness and its formal model, the Gaussian probability distribution function. Similarly, one-dimensional randomness and its formal model, white noise, provide the touchstone for time series analysis. Various spatially random patterns and processes, such as doubly stochastic Poisson processes, self-avoiding random walks, percolation theory, and conditional and simultaneous spatial autoregressive models provide neutral models for two-dimensional data. With the discovery of fractal geometry and the emergence of complexity theory, new neutral models have become useful to characterize distributed-disordered systems: fractional Brownian motion, Ising and Potts models, Levy flights, self-organized criticality, *etc*.

Notice, however, in this litany of neutral models that abiotic randomness motivates each. This pattern points to a fundamental problem in the use of such neutral models for investigation of biospheric dynamics: the biotic world is not random but—as our ecological understanding demonstrates—it is knowable, albeit *truly* complex. Many sciences must indirectly observe the responses of "their" dynamical systems to various stimuli, either intentional or coincidental. The problem of inferring process from pattern arises from many-to-one mappings in the absence of domain-specific models to *inform* that inference.

So, where can we find domain-specific models to guide informed inference from the rising flood of digital data? We need to build them. Perhaps we need to reinvigorate the concept of cities as nodal regions and work to make this concept interoperate with "traditional" land cover land use change studies. Consider of the manifold dimensions of nodal regions in terms of mass flow of water/carbon/nitrogen, in addition to concentrations of people/information/wealth. We need to explore the resource flow linkages between various subregions of cities and conurbations as well as between smaller cities and their hosting ecoregions to enable forecasting growth velocity, direction, intensification. We need, in particular, to rescale storylines of possible global futures (*e.g.*, SRES, MEA) into *regionally relevant and plausible* alternative scenarios that articulate drivers and constraints on the local level.

That we lack much data in many places is clear. Remote sensing cannot fill those voids, but it can help to constrain the possible. And it may provide us with as-yet-discovered spatio-temporal signatures that are characteristic of dynamics in urbanized environments. These signatures may then provide the units of analysis for investigating the dynamics of the built environment.

References

Allen TFH, T Hoekstra, 1992. *Toward a Unified Ecology*. New York: Columbia University Press.

Arbia G, 1989. *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Boston: Kluwer.

Cracknell AP, 1998. Review article. Synergy in remote sensing—what's in a pixel? *International Journal of Remote Sensing* 19:2025-2047.

Fisher P, 1997. The pixel: a snare and a delusion. *International Journal of Remote Sensing* 18:679-685.

Openshaw S, 1984. *The Modifiable Areal Unit Problem*. Concepts and Techniques in Modern Geography. Norwich, UK: Geo-Books.

Openshaw S, 1994. A concepts-rich approach to spatial analysis, theory generation, and scientific discovery in GIS using massively parallel computing. In: (M. Worboys, ed.) *Innovations in GIS*. London: Taylor & Francis. pp. 123-137.

Strahler AH, CE Woodcock, JA Smith, 1986. On the nature of models in remote sensing. *Remote Sensing of Environment* 20:121-139.