Emerging Trends and Challenges for Space-Time Analytics and Software in the Social Sciences

SERGIO REY
School of Geographical Sciences and Urban Planning
Arizona State University
Email: srey@asu.edu

The availability of software to support spatial data analysis has played a critical role in the adoption of a spatial perspective across the social sciences. Beginning in the 1980s the appearance of specialized code to carry out basic spatial econometric and exploratory spatial data analysis ushered in an era where these analytical tools have moved from the fringes of the social science toolkit to occupy central positions. By and large, however, the existing tools have been developed primarily for the cross-sectional context. Yet, all of the processes of interest to social scientists operate over both space and time. Demographers are not exempt.

The increasing availability of space-time data made possible through expanded data collection efforts by the volunteer geographic community and government agencies, the advent of sensor networks, GPS, and other geospatial data technologies have opened up a rich empirical context for the social sciences. However, this poses significant analytical challenges given the largely cross-sectional focus of our existing toolkit. I believe there is a critical need to develop new methods for space-time analytics that will enable demographers and other social scientists to leverage these new data sources. While there is a rich taxonomy of space-time data types (Goodchild 2008), I have focused my work on longitudinal spatial data (LSD), that is data observed for the same spatial unit over multiple periods. In what follows I describe the evolution of software designed to explore this form of space-time data together with some challenges I see going forward.

Space-Time Analytics: Evolution
My research developing space-time analytics had its origins in a long-standing interest in spatial inequality and regional growth dynamics. During the 1990s this literature was composed of largely separated groups of economists and sociologists studying the dynamics of income distributions, and regional scientists and geographers exploring spatial patterns and clustering in poverty and inequality. Both groups relied on different methodological frameworks with virtually no cross-fertilization between the camps, yet it was clear to me that the processes understudy had temporal, spatial, and distributional dimensions that needed to be considered in a comprehensive and simultaneous fashion. The toolsets of the day simply did not support this.

Space-Time Analysis of Regional Systems (STARS) was borne out of this need. It is an open source package that brings together a suite of new space-time analytical statistical measures together with an array of dynamic and interactive visualization tools (Rey and Janikas 2006). The space-time analytics consist of two broad sets of methods, the first departs from exploratory spatial data analysis methods (ESDA) and extends them to incorporate a dynamic component
giving rise to methods of ESTDA. The second begins with methods that have previously been developed for exploratory temporal data analysis (ETDA) and proceeds to introduce space into the analytical framework (ETSDA) (Rey 2001).

ETSDA extends the workhorse Markov chain methods for income distribution dynamics to incorporate a geographical dimension. The spatial Markov chain explores the role of regional context in the evolution of income distributions and gives empirical expression to the construct of spatial poverty traps. Spatial rank concordance and mobility measures (Rey 2011) take a similar approach to enhancing widely used a-spatial measures of distributional dynamics to consider spatial dependence and heterogeneity.

ESTDA methods extend both global and local measures of spatial association to the dynamic context. These allow for an assessment of the temporal stability of the local spatial clustering as well as new indicators of space-time association. The latter include Directional LISA statistics (Rey, Murray, and Anselin 2011) and a LISA Markov (Rey 2001).

STARS provides access to these, and other, analytical measures through a graphical user interface following the success of the desktop application model of GeoDa (Anselin et al. 2006). Although STARS supports limited scripting for more advanced users, there is a clear need for a library of spatial and space-time analytics to support the development of the next generation of spatial analysis software. PySAL (Rey and Anselin 2010) represents collaborative work to help address this need.

PySAL includes modules for spatial dynamics; classic ESDA, regionalization and districting; computational geometry; spatial inequality; and spatial econometrics, among others. As a library, PySAL supports a number of delivery mechanisms such as in cyberGIS gateways, web services, web clients, ArcGIS and QGIS plugins as well as desktop applications. Examples of the latter include work at ASU on the new version of STARS (planned release Spring 2012) and GeoDaSpace (spatial econometrics—planned release January 2012).

Challenges

While the methods above have gained increasing adoption (beyond scratching my own research itches), I can see a number of challenges standing before us as we work towards the next generation of space-time analytics. Some of these are methodological challenges, while others are more cultural in nature.

Methodological Challenges

A particularly thorny, yet fascinating, problem in a space-time context pertains to the nature of boundaries when defining either regions, as is done in the economic inequality literature, or neighborhoods in the vast literatures in sociology, urban epidemiology, criminology and geodemographics. The dominant approach in the investigation of social dynamics has been to define neighborhood boundaries at one point in time and then to measure compositional or attribute change in-situ. Any inferences drawn regarding these dynamics are contingent upon these boundaries remaining fixed over the period, yet neighborhoods, housing markets, and labor markets are dynamic constructs and this calls the assumption of static boundaries into
question. How to incorporate endogenous boundaries into the analytical framework is a major challenge (Rey et al. 2011).

A second methodological challenge is associated with the tremendous opportunities that are now becoming available in the form of high performance computing and cyberinfrastructure resources. Parallel and distributed computing will be increasingly common and can enable a massive expansion in the scope and nature of computational problems that can be addressed. Much of the existing spatial analytical software we rely on in the social sciences has not, however, been designed with these environments in mind, and a significant refactoring effort is required before this dream can be realized (Anselin and Rey 2011).

The third methodological challenge arises from an embarrassment of riches in the spatial analytical tools that social scientists have at their disposal. This diversity can be daunting to the untrained scientist, and we are far from consensus on many of the possible scientific workflows that would consist of a sequence of tasks necessary to carry out a particular type of spatial analysis. This situation will become even more complex when new space-time analytics are added to the toolkit. Ontologies for spatial data manipulation and provenance are receiving increasing attention, yet we currently lack anything remotely similar for spatial analytical workflows and models.

Cultural Challenges

To these methodological challenges I would add two cultural challenges. The first pertains to an often-heard criticism that social scientists applying spatial methods have failed to articulate the linkages between the uncovered spatial effects and the relevant substantive theory. Implicit here is that the problem lies with the tools at hand, yet one can also point to the overly abstract and reductive nature of most substantive theory as it pertains to spatial dimensions. Clearly there is a need for tighter linkage between spatial methods and substantive theory, but both ends of that linkage require development.

A second culture shift that, in my view, has to occur is from the current mode of scientist as consumer of spatial analysis software to scientist as collaborative producer of new analytical tools. The current situation is more akin to the cathedral model of software development (Raymond 1999) where small individual teams implement new analytical methods and release those packages to the wider research community. This is a recursive form of interaction, where scientific practice is very much constrained by the package functionality. This is problematic on several accounts. Gone are the days when the substantive questions of interest were easily addressed with off-the-shelf software. Increasingly novel twists extending theoretical frameworks will require new analytical tools. Scientists unwilling to get their hands dirty with the demands of scientific computation are likely to find themselves increasingly distant from the cutting edge.

In terms of new advances in tools, the cathedral model is also a constraint. I hope things begin to change to more of a bazaar model where larger numbers of geographically separated researchers become collaborators in the development of the next generation of tools. While this increases the start-up costs borne by the researcher, and requires new forms of training, it does rebalance the nature of inquiry to have scientific questions driving tool development rather than
the reverse. Tapping into the power of open source development and communities offers a new way to advance spatial analytical software and the practice of spatial social sciences (Rey 2009).

References