J Geograph Syst (2000) 2:11–15

Journal of Geographical Systems

Part 2 The Link between GIS and spatial analysis

GIS, spatial econometrics and social science research

Luc Anselin

Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, 326 Mumford Hall – MC 710, 1301 W Gregory Drive, Urbana IL, 61801-3608, USA (e-mail: anselin@uiuc.edu)

Abstract: Some ideas are formulated on the challenges presented to GIS, spatial analysis and spatial econometrics that result from recent trends in social science research. These new developments are characterized by a focus on the geography of phenomena. Particular emphasis is placed on the need to extend concepts of space, to broaden the analytical toolbox and to develop software and advance education.

Key words: Spatial econometrics, spatial analysis, social sciences

The subset of spatial analysis that pertains to the statistical analysis of spatially referenced data has recently gained a growing acceptance as a methodology in the mainstream social sciences. I will focus my remarks on this specific issue, leaving the discussion of aspects of spatial analysis such as optimization and decision support systems to others.

In contrast to what happened in the physical sciences in general, and specifically in natural and environmental analysis (e.g., soil science, hydrology), the acceptance and application of a spatial analytic perspective in the social sciences (*outside* of the discipline of human geography) is very recent. It is evidenced by a growing number of journal special issues, book volumes, etc. devoted to the application of "mapping", "GIS", or "spatial analysis" to subfields as varied as criminology, epidemiology, real estate analysis, and socio-economic analysis of tropical deforestation [some representative examples are Weisburd and McEwen (1998), Lawson et al. (1999), Can (1998), Liverman et al. (1998)]. This recent explosion of activity is often attributed to the rapid spread of affordable GIS technology to the desktop and the availability of a vast array of geographically referenced socio-economic data. This has led to the common use of GIS for data organization, data integration and visualization. In addition, simple mapping is increasingly augmented with new methods to detect meaningful local and global patterns and associations as part of an inductive approach to exploring data (exploratory spatial data analysis or ESDA). In other words, the widespread growth in the availability of *spatial data* can be considered as one of the main drivers in the acceptance and increased demand for spatial analysis in conjunction with GIS.

While this has undeniably been an important factor, an equally important driver has been the need to operationalize "new" *theoretical constructs* that explicitly incorporate space in the analysis of human (economic) behavior. Many of these concepts are similar (though not always acknowledged) to the models proposed by economic geographers and regional scientists in the 1960s, and stress the importance of location, neighborhood, region and spatial (social) interaction. Current examples in economics are the emphasis on spatial externalities and regional clusters (e.g., Krugman, Arthur, Porter), theories of interacting agents and interdependent decision making (e.g., Pollak, Ioannides, Durlauf, Brock, Brueckner), the importance of social interaction and group effects (e.g., Akerlof, Aoki) and neighborhood effects (Borjas). Similar examples can be cited in recent work in other social sciences, such as sociology, political science and criminology, where the central role of "space" is introduced in models of the diffusion of crime and innovative behavior, the structure of neighborhood and sense of community and notions of social capital (a more in-depth review is given in Anselin 1999a).

Empirical validation of the new "spatial" concepts and models requires a statistical and econometric methodology that takes into account location and spatial interaction. In other words, *spatial econometric* methods become crucial, in the sense that they tackle issues of spatial dependence and spatial heterogeneity, as well as their extensions to the space-time domain. Spatial econometrics is a subset of spatial statistics in that rather than being "statistics for (any) spatial data", it concerns itself with statistics for spatial (socio-economic) *models*, where the model specification is dictated by theory (Anselin 1988). These subtle differences aside, it is important to acknowledge that in the past few years a growing number of mainstream econometricians have started to contribute to the spatial econometric methodology and that spatial econometrics has gained recognition as a useful subset of the econometric toolbox. This has led to the development of new estimation methods (such as method of moments), specification tests and formal theoretical frameworks (asymptotics for dependent and heterogeneous spatial processes) to derive the properties of the various techniques (for overviews, see Anselin and Bera 1998, Anselin 1999b, Anselin and Florax 1999).

These recent developments in the mainstream social sciences in general and in economics in particular raise a number of challenges for the next generation of "spatial analysis." Central to this is the need to move "beyond mapping" (generally recognized in the GIS community, but not necessarily in the mainstream disciplines) and to tackle the methodological and theoretical issues that address the complexities of the current models. I see the potential for new developments in three important domains:

- *Extending concepts of "space"*
  Spatial analysis needs to go beyond dealing with Euclidean space and physical geographical locations to include location in "social" and "perceptual" space (social distance, economic distance). These general spaces do not necessarily fit in the current GIS data models and may require new sets of abstract frameworks to register, manipulate and visualize the relevant objects. This will also require further consideration and development

of distance metrics for these general spaces, such as distance metrics for "social" space, ways to conceptualize space-time dynamics and notions of "topology" in space-time (the counterpart of the "weights" matrix in spatial autocorrelation analysis). Similarly, it will become increasingly important to consider the intrinsic meaning of "space" and how the pervasiveness of spatial heterogeneity can be melded with general theory.

Space is more that geographic determinism, but a general theory of location and spatial interaction is still lacking in many respects. In the past, most of the attention has been spent on cross-sectional settings. However, the more interesting problems (both from a theoretical perspective as well as in the policy arena) deal with complex space-time dynamics and situations where space (e.g., neighborhood, community) is defined endogenously. Current spatial analysis is ill-equipped to deal with these complexities, but promising avenues are constituted by current work on GIS data models, object-oriented GIS, three-dimensional and space-time GIS, multimedia integration, and the like.

– *Broadening the analytical toolbox*

The toolbox of spatial econometrics and spatial analysis needs to be extended to deal with the challenges posed by the new theoretical frameworks in the social sciences. While much progress has been made, particularly with respect to the analysis and modeling of continuous cross-sectional data (the classical multivariate setting and linear regression model), several unresolved issues remain when it comes to space-time dynamics in general and the treatment of discrete and categorical data in particular. Central concerns pertain to modeling the intensity and extent of spatial interaction (the presence of spatial autocorrelation, the range of interaction, the construction for spatial weights and spatial lag operators) and its extension to hierarchical interactions between individuals and groups, local and global entities. Related methodological issues that have received only limited attention to date concern spatial effects in spatial interaction and transportation models (dependence in flow-data), spatial duration models, modeling changing choice sets, distinguishing spatial dependence from spatial heterogeneity, and effective visualization of model fit for spatially dependent (and heterogeneous) data.

In addition to new demands from the theoretical end, different conceptualizations are needed to handle the analysis of the very large spatial data sets that increasingly form the empirical setting for social science research. Notions of standard error and traditional significance tests lose much of their meaning when applied in near-asymptotic settings. Exploration of very large data sets, or, spatial data mining is increasingly seen as a replacement for traditional hypothesis testing and statistical modeling. Moreover, computations that are straightforward for data sets with hundreds of observations often become totally intractable when applied to millions of data points, especially since spatial interaction involves "matrices" of interaction.

The large size of today's realistic data sets as well as the complexity of the theoretical specifications often preclude analytical solutions, or, even when they exist, make them impossible to implement in reasonable computational time. Pure number crunching solutions are often the only alternative. These procedures depend heavily on efficient simulation routines, such as required by simulation estimators, Markov Chain Monte Carlo and

Gibbs sampling. For many of these techniques, the incorporation of spatial dependence is still in its infancy. Considerable advances in computational geometry and computational geography, both in terms of data models and algorithms, are required to satisfactorily address the complexity of spatial dependence in the models and large data sets used by social scientists.

– *Technology transfer: Software and education*
Most of today's commercial GIS software comes in the form of (partially) open environments that allow the user to include customizations and extend the functionality. In this modern component oriented computing environment, there has been a change in the focus of the discussion from determining the range of functionality that should be contained in a spatial analytical toolbox associated with the GIS to developing mechanisms that allow the analyst to mix and match components in order to accomplish specific tasks. Commercially provided ready-made shrink-wrapped toolboxes tend to deliver the lowest common denominator in terms of methodology and always lag the state of the art. For example, while many GIS software package nowadays include some form of rudimentary statistical analysis such as regression models, ironically, only very few contain facilities to estimate some of the more sophisticated *spatial* regression model. Moreover, there are two important drawbacks of the "black-box" approach. One is that uninitiated users identify "spatial analysis" with the (limited) set of techniques offered by a software vendor. The other is that the analysis is presented as being "easy" and underlying assumptions, algorithms and limitations are hidden from the user. However, the alternative of user-defined software solutions puts a much heavier burden on the analyst in terms of both methodological and software know-how. This may not be practical for social scientists whose main objective is to carry out "sound" spatial analysis. A promising avenue for future development in this respect is to build onto a common (open) platform and provide analytical tools (combined with methodological background) through a clearing-house, involving a virtual community of both scholars as well as software developers.

A related issue pertains to the training and education of future spatial analysts and how this will be situated in the methodological curricula of the mainstream social sciences. Clearly, there is some tension here between the traditional "turf" of the geography discipline (and geography departments) and the increased spatial awareness in the mainstream social science disciplines. This provides for exciting opportunities, either in terms of geography re-integrating itself into the mainstream, or for the development of new truly multi- and interdisciplinary GI "science" education as part of a new set of "degrees for the 21st century". In this context, there is an additional tension between education and training, and between traditional classroom formats and the fast-developing digital web-based distance education approach. In order to address the needs of social science research, some basic "spatial" content will have to penetrate into the traditional methodological toolbox. However, the dissemination of advanced spatial analysis expertise will probably have to rely on more cost-effective, web-based and virtual approaches. This constitutes a fascinating area of research involving both content development as well as the design and implementation of effective delivery mechanisms.

The demands of applied data analysis and the theoretical questions posed in the mainstream social sciences offer an important challenge to the methodology of spatial analysis. In return, spatial analysis has a lot to offer to mainstream social sciences, not only in terms of providing an productive toolbox, but also in terms of a long tradition of research on location and spatial interaction, which could be more effectively integrated into the social sciences.

This tension constitutes a major opportunity for the spatial analytical perspective (as part of a geographic information *science*) to contribute to both the theoretical as well as the methodological debates in the core social science disciplines and thereby move towards the development of more effective "spatially enabled" social sciences (Goodchild et al. 2000).

References

Anselin L (1988) *Spatial econometrics, methods and models*. Kluwer Academic, Boston

Anselin L (1999a) The future of spatial analysis in the social sciences. *Geographic Information Sciences* (forthcoming)

Anselin L (1999b) Spatial econometrics. In: Baltagi B (ed) *Companion to theoretical econometrics*. Basil Blackwell, Oxford (forthcoming)

Anselin L, Bera A (1998) Spatial dependence in linear regression models, with and introduction to spatial econometrics. In: Ullah A, Giles D (eds) *Handbook of applied economic statistics*. Marcel Dekker, New York

Anselin L, Florax R (eds) (1999) Advances in spatial econometrics. Springer, Berlin Heidelberg New York (forthcoming)

Can A (ed) (1998) Geographic information systems in housing and mortgage finance. *Journal of Housing Research* 9 (special issue)

Goodchild M, Anselin L, Appelbaum R, Harthorn B (2000) Toward spatially integrated social science. *International Regional Science Review* 23:139–159

Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertollini R (eds) (1999) *Disease mapping and risk assessment for public health*. John Wiley, Chichester

Liverman D, Moran E, Rindfuss R, Stern P (eds) (1998) *People and pixels, linking remote sensing and social science*. National Academy Press, Washington, DC

Weisburd D, McEwen T (eds) (1998) *Crime mapping and crime prevention*. Criminal Justice Press, New York