The Future of Spatial Analysis in the Social Sciences

Luc Anselin

Department of Agricultural and Consumer Economics
University of Illinois at Urbana-Champaign
Urbana, IL 61801-3608

Abstract
This paper presents a personal view on some emerging research directions at the interface of social science and spatial analysis. Particular emphasis is placed on methodological challenges presented by developments in social science theory, demands for data manipulation, and the need for education and dissemination.

I. INTRODUCTION

In this paper, I formulate some general ideas pertaining to emerging research challenges and promising directions for advances in spatial analysis that are motivated by demands generated by the social sciences. These ideas are intended to paint a broad picture of a research agenda for the next decade, in particular in terms of the contribution of the "next generation" spatial analysis tools to the social sciences, in the context of the development of "spatially integrated social science" [Goodchild et al. (2000), Anselin (2000)]. While spatial analysis is often defined as encompassing a wide range of spatial data manipulations, in this paper I will take a narrower view and focus on those techniques that are particularly relevant in the process of scientific discovery, using the framework outlined in Anselin and Getis (1992), In addition to limiting my remarks to social science perspectives, I will also focus exclusively on issues pertaining to spatial data analysis and spatial statistics, and deliberately not address the broader functionality of spatial analysis, which also includes, among others, spatial decision support systems, logistics and optimization.[1]

It has now been more than ten years since Goodchild (1987) argued for the importance of the spatial analytical aspects of GIS to further the solution of generic spatial research questions [see also Goodchild (1992)]. Since then, considerable progress has been made, particularly from a technical viewpoint. Familiar examples are the incorporation of spatial analytical functionality within commercial GIS, the linkage of specialized statistical and other analytical modules with GIS, and the discussion of geocomputational issues associated with such integration. However, as we enter the 21st century, a number of new research challenges are emerging that are not satisfactorily dealt with in the current state of the art. To some extent, these research challenges are qualitatively distinct from the impediments encountered in the late 1980s in that the very incorporation of GIS, spatial data and "spatial thinking" into the standard toolbox of the scientist has led to new questions. The solution to these questions requires concepts, techniques and implementations that go beyond the traditional paradigm that originated with the quantitative revolution in geography in the 1960s and 1970s.

In addition, a relatively recent phenomenon is the renewed attention in the mainstream social sciences to geography in general, and location and spatial interaction in particular. This, in turn, has created an explosion in the demand for methods and tools that allow the explicit treatment of space in empirical applications. Consequently, spatial analysis is playing an increasingly central role in measurement, hypothesis development and validation of theoretical constructs, activities that are crucial in the development of new scientific knowledge. The distinct contribution of spatial analysis in this overall framework is that it provides the means to explicitly recognize, assess and incorporate the importance of location and interaction within the methodological toolbox of the social scientist.

Undoubtedly, spatial analysis and GIS have also revolutionized the manipulation of geographic information in broader sections of the private and public sector, such as in urban planning, marketing and logistics. However, in this essay, these aspects will not be further considered since the focus is on "science" in general and social science in particular as the driver for new developments in spatial analysis. This role for science is reflected in three important dimensions. First, social science theory itself has generated a demand for new developments in spatial analysis, in the sense that new concepts related to geography, location and interaction require an explicit treatment

[1] Other recent assessments of research directions for spatial analysis that are less focused on specific social science applications can be found in the collection of papers in Fotheringham and Rogerson (1994), Fischer et al. (1996), Fischer and Getis (1997), as well as, among others, in Openshaw (1998), Miller (1999), and Goodchild and Longley (1999).

1082-4006/99/0502-67$5.00

of space. Similarly, the empirical verification of the new models (such as the "new" economic geography) requires appropriate spatial statistics that allow for dependence and heterogeneity in the data. Second, there is a practical demand for sophisticated analysis to deal with measurement issues and the manipulation of spatial (geocoded) data that are increasingly available to empirical workers. Third, this growing demand also has a human capital component, in that there is currently a shortage of adequately trained spatial analysts.

In the remainder of the paper, I first formulate some general thoughts on the interface between spatial analysis and social science research. This is followed by three sections, each dealing with a particular dimension of the challenges to spatial analysis that emanate from the social sciences, as suggested above (theory, data, and dissemination/education), In each of these sections I outline some motivating examples and suggest a number of specific research challenges to develop a research agenda for the next decade. Finally, some concluding remarks are formulated.

II. SPATIAL ANALYSIS AND SOCIAL SCIENCE RESEARCH

The use of GIS techniques in general and mapping in particular has become increasingly common in social science applications, in fields ranging from anthropology [Aldenderfer and Maschner (1996)], to criminology [Weisburg and McEwen (1998)], epidemiology [Lawson et al. (1999a)], real estate analysis [Can (1998)] and socio-economic analysis of tropical deforestation [Liverman et al. (1998)].[2] Recently, the focus in these applications has moved from simple data manipulation and visualization to spatial data analysis, both exploratory as well as confirmatory [e.g. Anselin (1998a)].

In general terms, one could arguably distinguish three important ways in which spatial analysis contributes to the "toolbox" of the social scientist. First, it provides the basis for *data integration*, or the conversion of data collected at one spatial scale (and time dimension) to other scales and dimensions. Specifically, this is needed when geo-locational information must be manipulated or when spatial data must be obtained for locations or areal units for which they are not originally recorded. This is particularly relevant in the combination of census data, remotely sensed images, maps and survey data towards the computation of measures of access, distance and spatial linkages.

Many research questions in the social sciences pertain to the interaction between the individual and larger social entities (context), and the empirical verification of models for these interactions requires the combination of micro and macro data, as well as the combination of spatially aggregate data at various scales (neighborhood, city, state, The methodological framework for accomplishing data integration is generically known as "small area estimation" in the statistical literature [Ghosh and Rao (1994)], and as spatial interpolation or areal interpolation in the GIS literature [Goodchild et al. (1993), Mitas and Mitasova (1999)]. The application of GIS and spatial analysis tools provides a means to obtain data for any scale, but also raises important questions of accuracy and error propagation [Goodchild and Gopal (1989)], as well as the fundamental concern about the "proper" scale of analysis [the ecological fallacy or modifiable areal unit problem, see King (1997)].

A second major contribution is the application of *exploratory spatial data analysis* (ESDA) and *visualization* in an inductive approach to discovering patterns, eliciting hypotheses and suggesting associations. ESDA is a subset of exploratory data analysis (EDA) [Tukey (1977)], but with an explicit focus on the distinguishing characteristics of geographical data. It is a collection of techniques to describe and visualize spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association, clusters or hot spots, and suggest spatial regimes or other forms of spatial heterogeneity [Anselin (1999a)]. ESDA is particularly powerful when no strong prior theoretical framework exists, as is often the case in interdisciplinary social science analysis in fields such as criminology or human-environment interaction [see Messner et al. (1999)].

A third major area of application pertains to contexts where a deductive approach is more appropriate, for example when economic theory dictates the variables to be considered as well as their functional relationship. When the empirical work is based on spatial (cross-sectional) data or when the models under consideration are "spatial" in nature (spatial interaction), an application of the specialized methodology of *spatial statistics* and *spatial econometrics* is required [Cressie (1993), Anselin (1988, 1999b)]. Broadly speaking, spatial econometrics covers the specification of spatial models, their estimation, specification tests for spatial effects as well as spatial prediction. An important aspect of the methodology is the "proper" incorporation of location and spatial arrangement, which are essential elements of a GIS.

Aside from these three areas in which spatial analysis may contribute directly to the methodological toolbox of the social scientist, the "geographical perspective", or *thinking spatially* also has an important role to play

[2] For a more general perspective, see also Martin (1996).

in the refinement of the way in which space is incorporated into social science theory itself. For example, recent attention to precision farming has forced agricultural economists to start considering how to incorporate the notion of spatial cost and output *surfaces* (rather than single numbers) into production theory [Weiss (1996)]. Similarly, emerging attention to social interaction in economics [Akerlof (1997)] requires the specification of models for the flows and intensity of spatial interaction. Unfortunately, this does not always build upon the wealth of insights and conceptual frameworks developed in economic geography and regional science that deal specifically with measures of accessibility, models to incorporate friction of distance, complex interaction flows, and the like [e.g., Isard et al. (1998), Fotheringham and O'Kelly (1988), Sen and Smith (1995)]. An effective path towards spatially integrated social science would consist of recognizing and extending this existing body of work and to avoid "reinventing the wheel".

Finally, the interaction between spatial analysis and social science research is not solely one-directional, in the sense of spatial analysis being simply a "tool" to further social science research. In many respects, the methods and conceptual frameworks of spatial analysis—as a part of geographic information science, or GI Science—trace their origins to the natural and environmental sciences, and therefore may not always be well suited to the demands of the social sciences. A major contribution of social science theory to GI Science is therefore to provide superior social and behavioral foundations for concepts of space and interaction and to suggest the basis for the formal specification of spatial models. In addition, demands from the social sciences raise the threshold of sophistication required from spatial data analysis, in that they tend to focus on the manipulation of discrete and categorical variables and stress space-time dynamics.

Both methodological and theoretical considerations suggested by recent developments in the social sciences will require a rethinking of some of the central concepts in spatial analysis. This provides for a rich research agenda, which will be elaborated upon in the following three sections.

III. THEORY

An important motivation for the recent explosion in the attention paid to GIS and spatial analysis in the mainstream social sciences derives from a number of exciting theoretical developments in economics, sociology, and political science. These developments share a common interest in the *interaction* between decision makers (as interacting agents) and the feedback between the individual and the group (context). Concepts such as social norms, neighborhood effects, peer group effects, social capital, strategic interaction and copy-catting deal with interesting questions of how the individual interactions can lead to emergent collective behavior and aggregate patterns. In these conceptualizations, the roles of location, space and spatial interaction are central. Therefore, increasingly, the empirical verification of these models requires statistical and econometric techniques that acknowledge and incorporate the spatial effects.

A few examples may help to illustrate this point. Some recent work on models of social interaction and complex behavior in economics builds on principles developed in statistical mechanics, such as interacting particle systems and random field models [Brock and Durlauf (1995), Akerlof (1997), Durlauf (1997)]. The basic underlying idea in these theories is a feedback mechanism between the value of a phenomenon at a given location and the magnitude at "neighboring" locations (where the notions of location and neighborhood are not necessarily in a geographical sense). This leads to model specifications that are formally equivalent to the spatial Markov fields developed in spatial statistics [Cressie (1993)]. Similar notions underlie some of the new macroeconomics of Aoki (1996), where the interaction is in the form of a "mean field" term (some average effect of the aggregate upon the value at the micro level). Other conceptualizations that explicitly incorporate interaction as part of the theoretical construct are models for evolving trading structures [Ioannides (1997, 1999)], neighborhood spillover effects [Durlauf (1994), Borjas (1995), Glaeser et al. (1996)], yardstick competition [Besley and Case (1995), Bivand and Symanski (1997)], and strategic interaction [Case et al (1993), Brueckner (1998)].

A second important strand of theoretical literature that emphasizes the importance of location, space and spatial interaction is the "new economic geography" popularized in the work of Krugman (1991a, 1991b, 1996, 1999), Arthur (1989), Glaeser et al. (1992), and others.[3] The resulting models of increasing returns, path dependence and imperfect competition induce various forms of spatial externalities, agglomeration economies and spillovers, whose spatial imprint requires a spatial econometric approach in empirical work [see Anselin et al. (1997)].

In sociology, a recent renewed attention to the ecological perspectives pioneered by the Chicago School in the early 1920s has yielded a growing

[3] For a recent review, see the collection of papers in Pleskovic (1999), Fujita et al (1999); also Martin (1999) for a more critical assessment.

number of studies in which computerized mapping and spatial analysis techniques have become central. These new efforts follow directly from the theoretical requirements that relate individual behavior to that of the "context" and thus result in attempts to quantify notions such as social capital and neighborhood effects [see, e.g., Abbot (1997), Morenoff and Sampson (1997), Sampson et al. (2000)]. In political science as well, especially in the study of international relations, a spatial perspective is increasingly prominent in theoretical as well as empirical approaches, suggesting the formation of a new geopolitics [Starr (1991), Ward (1992), O'Loughlin et al. (1999)].

Challenges

The resurgence in attention to space and spatial interaction in social science theory provides a challenge to spatial analysis as we know it, in the sense that many (most) of the currently available data models and analysis methods are not particularly geared to deal with these evolving theoretical concepts. Three broad challenges in particular require some further discussion.

First is the concept of *"space"* itself, how it is incorporated in statistical models (regression models in particular) as well as stored in digital form in a GIS. The standard approach is to treat space as a container for spatial objects or as a field by means of which spatial distributions are described, but several of the theoretical frameworks suggest a notion of space that is endogenously determined and changes as a function of the strength of interaction (e.g., neighborhood sense, perceptual space). In addition, while the development of data models that incorporate space-time dynamics is an active area of research in GI Science, the extension of this to address individual-group interactions, perceptual space and other cognitive aspects is still in its infancy [Talen (1999)]. Similarly, several conceptualizations of spatial interaction [such as strategic interaction between states in Case et al. (1993)] suggest the importance of non-Euclidean and non-geographical notions of distance and distance decay. In terms of methodology, this has direct implications for the specification of the so-called spatial weights in spatial regression models [Anselin (1988)]. The full extent to which the standard estimators and specification tests also extend to more general metrics is not yet fully understood and an area of active research in spatial econometrics.

More generally, the incorporation of abstract spaces (such as policy space, attribute space) and distance metrics (such as economic distance, social distance, political distance) within spatial analysis will require a rethinking of some of the standard data models and algorithms. In addition, new statistical methods may be needed to ensure that proper inference is obtained when manipulating such "spatial" data. Promising directions are contained in new results obtained for the asymptotics of spatial econometric methods [Kelejian and Prucha (1999), Pinkse (2000)] on the statistical side, as well as in the work on object oriented GIS, participatory GIS and 3-D GIS. However, much remains to be done to address the sophisticated concepts of space suggested by social science theory.

A second important challenge to spatial analysis driven by theoretical concerns follows from the need to provide a meaningful theoretical interpretation for the *role of "space"* as it is incorporated in spatial statistical and spatial econometric models. The effect of neighbors (contiguous locations) as included in models for spatial dependence (e.g., through spatial weights), or the effect of location as expressed in models for spatial heterogeneity (e.g., in the form of spatially varying coefficients or spatial regimes) does not in and of itself provide an "explanation" of the phenomenon under study. Instead, it may suggest the role of a spatial multiplier effect, the extent of that effect and its strength, but it does not and cannot discover the actual socio-behavioral mechanisms that generate the effects. Similary, there is much recent excitement over methods to model local geographic heterogeneity such as LISA [local indicators of spatial association, Anselin (1995)] and GWR [geographically weighted regression, Fotheringham (1997)], although such methods do not in themselves explain the underlying heterogeneity. To many sceptics, these more sophisticated spatial formulations are nothing but models of geographic determinism in disguise. The real challenge to spatial analysis is not only to develop new techniques of "local" spatial analysis, or more sophisticated models that formally express spatial effects, but also to provide the means to discover and understand the underlying social and behavioral mechanisms that yield the revealed spatial patterns. This is further complicated by the observational equivalence between spatial dependence and spatial heterogeneity in cross-sectional settings. In many instances, this is a form of the "inverse problem" encountered in the physical sciences.[4] Here, there is an important role for cross-fertilization between theory and the tools of analysis.

A third, related, but somewhat more technical challenge to spatial analysis is to develop data models and modeling techniques to handle spatial interaction as well as *space-time interaction*. Especially in the context of theoretical models of diffusion and contagion, a proper metric for the distance in space-time (or speed of diffusion) is required. In addition, most of the theoretical models of interaction are

[4] For a recent discussion, see, e.g., Chilès and Delfiner (1999, Ch. 8).

developed at the micro-level and deal with discrete dependent variables [e.g., Brock and Durlauf (1995)], whereas the current state of the art of spatial econometric methodology pertains primarily to the standard (continuous dependent variable) regression model. Some initial advances have been made, some based on analytical econometric approaches [e.g., Pinkse and Slade (1998)], others using Bayesian and computation-intensive estimators [e.g., LeSage (1997), Waller et al. (1997)], but this still constitutes a vast area of research with important theoretical, methodological and computational ramifications.

Overall, while certainly much progress has been made since the late 1980s, the existing tools of spatial analysis, spatial econometrics and spatial statistics are still lacking and much remains to be done to develop a flexible dynamic modeling toolbox that is able to reflect the complexity of spatial components of the new theoretical frameworks that are emerging in the social sciences.

IV. DATA

Arguably, the most commonly cited reason for the increased interest in spatial analysis by social scientists is the explosion in the availability of geo-coded socio-economic data sets (i.e., data sets that also contain the location of the observational units). The existence of an extensive infrastructure of spatially referenced road networks (e.g., the Tiger files of the U.S. Census bureau) as well as digital base maps for a wide range of administrative units, and the affordability and availability of Global Positioning Systems (GPS) has made the explicit recording of location a routine matter. Where privacy laws permit, data on a wide range of socio-economic variables, from employment to crime, public health and the environment are distributed in formats that are amenable to geographical analysis. Increasingly, place-based search is becoming implemented in digital libraries and web-based initiatives to facilitate spatial data sharing and dissemination [Goodchild et al. (2000)]. Furthermore, since many federal regulations are spatially explicit (e.g., the U.S. Community Reinvestment Act to ensure equitable access to mortgage lending in "neighborhoods") they carry reporting requirements on a wide range of transactions in an explicit geographic framework [e.g., Thrall (1998)]. This has also spawned an explosion of activity in the private sector in the form of value added reselling of public data, geodemographic analyses and target marketing [Birkin and Clarke (1998), Birkin et al. (1999)].
Empirical research in a range of fields has begun to take advantage of this plethora of spatial information. For example, a spatial perspective is increasingly the standard in the analysis of human-environmental interaction, such as in investigations of changing land use and land cover and the assessment of tropical deforestation [e.g., Bockstael (1996), Wood and Skole (1998)]. Such studies are characterized by the availability of a wealth of spatial data, recorded at various scales and with different resolutions. Often they involve the combination of survey information at the individual or household level with census data at the administrative areal unit level, such as in spatial targeting, risk mapping and poverty mapping [e.g., Nelson and Gray (1997), Lawson et al. (1999b)].

Much of the spatial analysis needs driven by this explosion in the availability of spatial information can be met by today's technology. However, the very size of the data and the multitude of frames for collecting them raise a number of issues that are currently not adequately met and require further research, specifically with respect to spatial scale, the size of the data sets and spatial sampling.

Challenges

A first challenge is to address the issue of *spatial scale* in light of the increased data availability and growing analytical power of GIS. The choice of the "proper" scale of analysis has become an essential part of the design of scientific inquiry in the spatial sciences. Today, all kinds of geo-coded data sets have become easily accessible, with information collected from the individual to the global level. Moreover, powerful GIS tools allow one to move from one scale of analysis to another as well as to integrate data collected at different scales. Clearly, observations for one level of analysis (e.g., at an aggregate level) do not necessarily provide useful information about lower levels of analysis (such as individual behavior), especially when spatial heterogeneity is present. Also, as observations are re-arranged into "zones", several statistics change in value, such as correlation coefficients and measures of spatial autocorrelation. This is an old and familiar methodological problem, known to sociologists at the ecological fallacy [for a recent review, see King (1997)], to geographers as the modifiable areal unit problem, or MAUP [Openshaw and Taylor (1979)] and to earth scientists as the upscaling or change of support problem [Chilès and Delfiner (1999)].

One perspective on the ecological inference problem is that it is impossible to solve, since the properties of any predicted value however constructed remain unverifiable.[5] Alternatively, one could argue that

[5] The typical context for an ecological inference is when data are not available at some lower level of aggregation (such as a household) but inference for that level is based on observations at a higher level (such as census tracts). If the micro-data were available, they should be used, and there would not be an ecological fallacy problem. When they are not available, the accuracy of any predicted value cannot be verified.

studies should be based on scale-invariant concepts or scale-invariant variables, such as densities and surfaces. While this may have an intuitive attractiveness in the physical sciences, many processes in social science are discrete in nature, and modeling frameworks to deal with this characteristic are still in their infancy.[6] The problem remains how to construct or estimate the relevant surfaces, and while methods such as indicator kriging from geostatistics offer considerable promise [e.g. Goovaerts (1997), Chilès and Delfiner (1999)], their applicability to socio-economic phenomena remains largely untested. Similarly, the exploratory (graphical) and simulation tools for "ecological inference" (EI) proposed by King (1997) are a start, but so far they do not take into account spatial autocorrelation and other spatial aspects of the problem, and this remains an active area of research (and controversy).

A related issue is the integration of multiple scales of analysis, as in hierarchical modeling. While such analysis is now well established in social science methodology [for example, hierarchical linear model popularized by Bryk and Raudebush (1992), and its Bayesian counterparts], its extension to dealing with spatial data remains to be further developed [initial approaches can be found in Langford et al.(1999)]. This is particularly crucial in the analysis of categorical and discrete variables (such as counts of events), where there are no acceptable analytical solutions that incorporate spatial dependence. Methods based on simulation estimators, Markov Chain Monte Carlo (MCMC) and Gibbs sampling are extremely promising [Gilks et al. (1996), LeSage (2000), Beron and Vijverberg (2000)], but a number of important methodological and computational issues remain to be addressed, especially to implement these techniques in realistic large sample settings.

The sheer size of available geo-spatial databases constitutes a second challenge to spatial analysis. Most "classical" techniques of spatial data analysis were initially developed for situations where the data sets contained less than a hundred observations.[7] In contrast, the current norm is easily several orders of magnitude greater, such as in the analysis of real estate transactions [Pace and Barry (1997)]. There are several implications of this larger size. One is that exploratory spatial data analysis, under the guise of "spatial data mining" has become crucial in the process of looking for patterns, clusters, associations and other meaningful non-randomness. While many of the currently available techniques such as LISA are in principle applicable, their implementation in very large data sets can easily constitute a computational (permutation approaches for each observation) as well as conceptual (multiple comparisons) challenge. The role of geocomputation has become more important than ever. Since many spatial problems are intrinsically of order N^2, they cannot be effectively tackled in current computational environments unless special purpose algorithms are developed to handle memory management, efficient searching, sorting and data manipulation [Anselin (1998b)].

Another issue related to the large size of geospatial data sets is the choice of inferential paradigm. Classical asymptotic theory in spatial statistics and spatial econometrics has been developed to approximate the properties of estimators and tests statistics in finite samples, but they are not as meaningful when the sample at hand actually does approach infinity.[8] Alternative paradigms, based on Bayesian notions or purely computational (simulation estimators, resampling methods, permutation approaches) hold considerable promise, but their implementation in very large spatial data sets is still far from trivial.

A final issue is related to spatial sampling, a topic typically ignored in the design and application of survey research in the social sciences. Many micro data sets are currently available to social scientists and are being used to study various implications related to geographic notions such as spatial interaction, sense of community, social capital and related concepts. However, the stratification of the surveys on which they are based typically has ignored the role of spatial effects and may therefore be totally inappropriate for the purposes of spatial analysis. Three issues are important here. A first situation is encountered when the interest lies in designing or stratifying a survey in order to correct for any possible presence of spatial effects, such as spatial autocorrelation. A basic principle underlying spatial sampling in this context is to assume a form of distance decay for the autocorrelation. As a result, observations that are "far enough apart" can be considered to be spatially uncorrelated and thus can be treated in the usual fashion [e.g., the principle underlying the so-called DUST—dependent areal units sequential technique—sampling procedure in Arbia (1993)]. A second instance is when the survey is designed to capture and estimate the extent and strength of spatial interaction itself. In this instance, some form of cluster sampling design is required, where the degree of clustering should match the range of the spatial interaction process of interest. In both cases, it is essential to understand

[6] Early formulations of theoretical frameworks can be found in Isard and Liossatos (1979), but these are near impossible to implement in an empirical setting.

[7] For example, the classic Irish county data set used in Cliff and Ord (1973) to illustrate spatial autocorrelation tests and spatial autoregressive models contained only about 25 observations.

[8] This is often ignored in practice, but the essence of the problem is that in the limit most estimators and test statistics converge to a fixed constant, with zero variance and a degenerate distribution collapsed onto the fixed constant.

and measure the underlying spatial autocorrelation, since it forms the basis for respectively separating or grouping sampling units. Finally, the granularity of the sampling design is crucial for understanding processes characterized by spatial heterogeneity.

These problems are familiar in the physical sciences, where a growing literature deals with spatial sampling in the context of environmental monitoring and resource exploration [e.g., Arbia and Lafratta (1997), Müller (1998)]. However, their extension to the domain of social science survey design remains largely unexplored.

V. DISSEMINATION AND EDUCATION

A third important set of research challenges follows from the need to educate and train sufficient numbers of spatial analysts to support the current and future demand for spatial analytical expertise generated by the social sciences. There are two aspects to this question, one pertaining to the dissemination of tools and techniques, the other to the provision of sufficient human capital.

In the late 1980s, the lack of software tools in commercial GIS environments to carry out spatial analysis in general and spatial data analysis in particular was often cited as a main reason for the slow dissemination of these techniques to empirical practice [e.g., Haining (1989)]. However, to a large extent, this impediment has been removed. Not only has there been a slew of sofware tools developed in academic environments to augment existing commercial GIS with spatial data analysis capability [e.g., Anselin (1992), Anselin and Bao (1997), Zhang and Griffith (1997), Symanzik et al. (1999)], but also the commercial vendors themselves have entered this arena. Examples are the spatial analysis modules provided with ESRI's ArcView GIS and the S+ArcView link of MathSoft [for a recent review, see Bao et al. (1999)]. However, there remains considerable tension between the "lowest common denominator" approach taken in the COTS (commercial off-the-shelf) products and the advanced to esoteric methods incorporated in the academic ventures. Also, the incorporation of spatial data analytical techniques into the market-leading statistical and econometric software packages is still limited to mapping and some basic geostatistical techniques.

Great strides have been made in terms of the incorporation of geographic information sciences into the mainstream curriculum of geography departments [e.g., the model curricula developed by NCGIA and UCGIS], but outside geography, education and training in GI sciences is still the exception rather than the rule.

In light of these circumstances, three main research challenges suggest themselves: the development of a generic spatial analytical toolbox, the role of spatial analysis in social science methods curricula, and the ultimate contribution of spatial analysis to social science education and research.

Challenges

The search for a generic spatial analysis software toolbox has generated considerable debate. The research challenge associated with this is to find a compromise between the commercial requirements for sufficient market size and the sophisticated needs of an ever-changing methodological state of the art. On the one hand is the "black-box" approach often favored in commercial environments, with limited demands on the sophistication of the user through an easy to understand user interface, but typically insufficiently advanced for a research environment. On the other hand are the "programming" approaches offered in a number of advanced computing packages, which put most of the burden on the user. A compromise may lie in the development of suites of components that can be mixed and matched by sophisticated users or wrapped in a shell and interface for less computer savvy users. Such development would be greatly enhanced by the existence of an open and virtual community of scholars whose contributions would be made available. The creation of such a community is one of the cornerstones of the new Center for Spatially Integrated Social Science [Goodchild et al. (2000)].

A second challenge pertains to the integration of spatial analysis in the methodological curricula of social scientists. Geographical information science has traditionally been the "turf" of geography departments, but some recent developments may bring this into question. For example, in a number of places GI Science programs and degrees are offered in a multi-disciplinary and inter-disciplinary environment outside the traditional departmental boundaries (e.g., at the University of Texas at Dallas and the University of Utah). Some of these initiatives are promoted by industry, using state of the art technology in distance education and web-based learning (e.g., ESRI's virtual university). The question remains whether a "reinvented" Geography will claim this terrain or whether this constitutes a threat to its traditional role, in the sense that new "degrees for the 21st century" will increasingly take on this task. These issues will require considerable debate and seem far from resolved at this point in time. In addition to the matter of who will provide the education, there remains much work to be done to integrate even basic notions of spatial analysis into mainstream social science methods curricula.

Finally, an underlying theme of this essay was that indeed "space does matter," although this is by far not a widely accepted notion in the mainstream of the social sciences. An important challenge to the spatial analysis community consists of demonstrating in unequivocal terms that much is to be gained by a careful and explicit incorporation of the spatial element in social science research. Making the sales pitch credible is a formidable task and much remains to be done.

VI. CONCLUSION

A number of developments in the theory and empiricism of the social sciences form the foundation for a growing importance of spatial analysis. We are at an exciting frontier and it is likely that the next decade will bring tremendous gains in the technical and theoretical prowess associated with GI science. In order to accomplish these advances, a tight interaction between theory, data analysis and computation will be necessary. In addition, these advances are more likely to happen in an interdisciplinary and multidisciplinary environment where traditional boundaries are brought down. It is a challenge to the GI science community to seize this opportunity and provide the means to develop a "spatially integrated" social science.

ACKNOWLEDGMENTS

Earlier versions of this paper were presented at the Varenius Workshop on Status and Trends in Spatial Analysis, Santa Barbara, CA, Dec. 10–12, 1998, and as the opening address to Geoinformatics '99, Ann Arbor, MI, June 19–21, 1999. Comments from Shuming Bao, referees and participants at the workshops are gratefully acknowledged. Part of the research reported on in this paper was supported by NSF grant BCS-9978058 to the Center for Spatially Integrated Social Sciences (CSISS).

REFERENCES

[1] Abbot, A., 1997, Of time and space: the contemporary relevance of the Chicago School, *Social Forces,* 75:1149-1182.

[2] Akerlof, G., 1997, Social distance and social decisions, *Econometrica,* 65:1005-1027.

[3] Aldenderfer, M. and H. Maschner, 1996, *Anthropology, space, and Geographic Information Systems*, New York: Oxford University Press.

[4] Anselin, L., 1988, *Spatial Econometrics: Methods and Models*, Boston: Kluwer Academic.

[5] Anselin, L., 1992, *SpaceStat, a software package for the analysis of spatial data*, National Center for Geographic Information and Analysis, University of California, Santa Barbara.

[6] Anselin, L. , 1995, Local indicators of spatial association – LISA, *Geographical Analysis,* 27:93-115.

[7] Anselin, L., 1998a, GIS research infrastructure for spatial analysis of real estate markets, *Journal of Housing Research,* 9:113-133.

[8] Anselin, L., 1998b, Exploratory spatial data analysis in a geocomputational environment, In P. Longley, S. Brooks, R. McDonnell, and B. Macmillan (eds.), *Geocomputation, A Primer*, pp. 77-94, London: John Wiley.

[9] Anselin, L., 1999a, Interactive techniques and exploratory spatial data analysis. In P. Longley, M. Goodchild, D. Maguire and D. Rhind (eds.), *Geographical Information Systems*, 2nd Edition, pp. 253-266, New York: John Wiley.

[10] Anselin, L., 1999b, Spatial econometrics, In B. Baltagi (ed.), *Companion to Theoretical Econometrics*, Oxford: Basil Blackwell (forthcoming).

[11] Anselin, L., 2000, GIS, spatial econometrics and social science research, *Journal of Geographical Systems* 2 (forthcoming).

[12] Anselin, L. and S. Bao, 1997, Exploratory spatial data analysis linking SpaceStat and ArcView, In M. Fischer and A. Getis (eds.), *Recent Developments in Spatial Analysis*, pp. 35-59, Berlin: Springer-Verlag.

[13] Anselin, L. and A. Getis , 1992, Spatial statistical analysis and geographic information systems, *Annals of Regional Science* 26:19-33.

[14] Anselin, L., A. Varga, Z. Acs, 1997, Local geographic spillovers between university research and high technology innovations, *Journal of Urban Economics,* 42:422-448.

[15] Aoki, M., 1996, *New Approaches to Macroeconomic Modeling*, Cambridge: Cambridge University Press.

[16] Arbia, G., 1993, The use of GIS in spatial statistical surveys, *International Statistical Review,* 61:339-359.

[17] Arbia, G. and G. Lafratta , 1997, Evaluating and updating the sample design in repeated environmental surveys: monitoring air quality in Padua, *Journal of Agricultural, Biological and Environmental Statistics,* 2:451-466.

[18] Arthur, W. B., 1989, Increasing returns, competing technologies and lock-in by historical small events: the dynamics of allocation under increasing returns to scale, *Economic Journal* , 99:116-131.

[19] Bao, S., L. Anselin, D. Martin and D. Stralberg, 1999, Seamless integration of spatial statistics and GIS: the S-Plus for ArcView and the S+Grassland links, *Journal of Geographical Systems* (forthcoming).

[20] Beron, K.J. and W.P.M. Vijverberg, 2000, Probit in a spatial context: a Monte Carlo analysis, In L. Anselin and R. Florax (eds.), *Advances in Spatial Econometrics*, Heidelberg: Springer-Verlag (forthcoming).

[21] Besley, T. and A. Case, 1995, Incumbent behavior: vote-seeking, tax-setting and yardstick competition, *American Economic Review,* 85:25-45.

[22] Birkin, M. and G. Clarke, 1998, GIS, geodemographics and spatial modeling in the UK financial service

industry, *Journal of Housing Research,* 9:87-111.

[23]Birkin, M., G. Clarke and M. Clarke, 1999, GIS for business and service planning. In P. Longley, M. Goodchild, D. Maguire and D. Rhind (eds.), *Geographical Information Systems,* 2nd Edition, pp. 709-722. New York: John Wiley.

[24] Bivand, R. and S. Szymanski, 1997, Spatial dependence through local yardstick competition: theory and testing, *Economics Letters* ,55:257-265.

[25] Bockstael, N., 1996, Modeling economics and ecology: the importance of a spatial perspective, *American Journal of Agricultural Economics,* 78:1168-1180.

[26] Borjas, G., 1995, Ethnicity, neighborhoods, and human-capital externalities, *American Economic Review,* 85: 365-390.

[27] Brock, W. and S. Durlauf, 1995, *Discrete choice with social interactions I: Theory.* NBER Working Paper No. W5291. Cambridge, MA: National Bureau of Economic Research.

[28] Brueckner, J., 1998, Testing for strategic interaction among local governments: the case of growth controls. *Journal of Urban Economics,* 44:438-467.

[29] Bryk, A. and S. Raudenbush, 1992, *Hierarchical Linear Models,* Newbury Park: Sage Publications.

[30] Can, A, Ed., 1998, Geographic information systems in housing and mortgage Finance, *Journal of Housing Research* 9 (special issue).

[31] Case, A., H. S. Rosen and J. R. Hines, 1993, Budget spillovers and fiscal policy interdependence: evidence from the States, *Journal of Public Economics,* 52:285-307.

[32] Chilès, J-P. and P. Delfiner, 1999, *Geostatistics*: Modeling Spatial Uncertainty. New York: John Wiley.

[33]Cliff, A. and J.K. Ord, 1973, *Spatial Autocorrelation.* London: Pion.

[34] Cressie, N., 1993, *Statistics for Spatial Data.* New York: John Wiley.

[35]Durlauf, S., 1994, Spillovers, stratification and inequality. *European Economic Review,* 38:836-845.

[36] Durlauf, S., 1997, Statistical mechanics approaches to socioeconomic behavior. In W. B. Arthur, S. Durlauf, D. Lane (eds), *The Economy as an Evolving Complex System* II, pp. 81-104. Reading, MA: Addison-Wesley.

[37] Fischer, M. M. and A. Getis, 1997, *Recent Developments in Spatial Analysis.* Berlin: Springer-Verlag.

[38] Fischer, M. M., H. Scholten and D. Unwin, 1996, *Spatial Analytical Perspectives on GIS.* London: Taylor & Francis.

[39]Fotheringham, A. S., 1997, Trends in quantitative methods I: stressing the local, *Progress in Human Geography,* 21:88-96.

[40]Fotheringham, A.S. and M. O'Kelly, 1988, *Spatial Interaction Models: Foundations and Applications.* Dordrecht: Kluwer Academic.

[41]Fotheringham, A. S. and P. Rogerson, 1994, *Spatial Analysis and GIS,* London: Taylor & Francis.

[42] Fujita, M., P. Krugman, A. Venables, 1999, *The Spatial Economy: Cities, Regions and International Trade.* Cambridge, MA: MIT Press.

[43]Ghosh, M. and J. N. K. Rao, 1994, Small area estimation: an appraisal, *Statistical Sciences,* 9:55-93.

[44] Gilks, W., S. Richardson, D. Spiegelhalter, 1996, *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall.

[45] Glaeser, E., H. Kallal, J. Scheinkman, A. Schleifer , 1992, Growth in cities, *Journal of Political Economy,* 100:1126-1152.

[46] Glaeser, E., B. Sacerdote and J. Scheinkman 1996, Crime and social interactions. *Quarterly Journal of Economics,*111:507-548.

[47] Goodchild, M., 1987, A spatial analytical perspective on geographic information systems, *International Journal of Geographical Information Systems,*1:327-334.

[48] Goodchild, M., 1992, Geographical information science, *International Journal of Geographical Information Systems,*6:31-45.

[49] Goodchild, M. nd S. Gopal, 1989, *Accuracy of Spatial Databases.* London: Taylor and Francis.

[50] Goodchild, M. and P. Longley, 1999, The future of GIS and spatial analysis, In P. Longley, M. Goodchild, D. Maguire and D. Rhind (eds.), *Geographical Information Systems,* 2nd Edition, pp. 567-580. New York: John Wiley.

[51] Goodchild, M., L. Anselin, R. Appelbaum, B. Harthorn, 2000, Towards spatially integrated social science, *International Regional Science Review,*23:139-159.

[52] Goodchild, M., L. Anselin, U. Deichmann, 1993, A framework for the areal interpolation of socio-economic data, *Environment and Planning ,A,* 25:383-397.

[53] Goovaerts, P., 1997, *Geostatistics for Natural Resources Evaluation.* New York: Oxford University Press.

[54]Haining, R., 1989, Geography and spatial statistics: current positions, future developments. In B. Macmillan (ed.), *Remodeling Geography,* pp. 191-203. Oxford: Blackwell Publishes.

[55] Isard, W. and P. Liossatos, 1979, *Spatial Dynamics and Optimal Space-Time Development.* New York: North Holland.

[56] Isard, W., I. Azis, M. Drennan, R. Miller, S. Saltzman, E. Thorbecke, 1998, *Methods of Interregional and Regional Analysis.* Aldershot: Ashgate.

[57] Ioannides, Y. M., 1997, Evolution of trading structures. In W. B. Arthur, S. Durlauf, D. Lane (eds), *The Economy as an Evolving Complex System* II, pp. 129-167. Reading, MA: Addison-Wesley.

[58] Ioannides, Y.M., 1999, Economic geography and the spatial evolution of wages in the United States. In L. Anselin and R. Florax (eds), *Advances in Spatial Econometrics.*Heidelberg:Springer-Verlag (forthcoming).

[59] Kelejian, H. and I. Prucha, 1999, *On the asymptotic distribution of the Moran I test statistic, with applications.* Working Paper, Department of Economics, University of Maryland, College Park, MD.

[60] King, G., 1997, *A Solution to the Ecological Inference Problem.* Princeton: Princeton University Press.

[61] Krugman, P., 1991a, Increasing returns and economic geography, *Journal of Political Economy* 99:438-499.

[62] Krugman, P., 1991b, *Geography and Trade.* Cambridge, MA: MIT Press.

[63] Krugman, P., 1996, Urban concentration: the role of increasing returns and transport costs, *International Regional Science Review* 19, 5-30.

[64]Krugman, P., 1999, The role of geography in development, *International Regional Science Review* 22: 142-161.

[65] Langford, I., A. Leyland, J. Rasbash and H. Goldstein, 1999, Multilevel modeling of geographical distributions of diseases. *Applied Statistics,* 48:253-268.
[66] Lawson, A., A. Biggeri, D. Böhning, E. Lesaffre, J-F Viel, R. Bertollini, Eds. 1999a, *Disease Mapping and Risk Assessment for Public Health.* Chichester: John Wiley.
[67]Lawson, A., D. Böhning, A. Biggeri, E. Lesaffre, J.-F. Viel 1999b, Disease mapping and its uses, in Lawson, A., A. Biggeri, D. Böhning, E. Lesaffre, J-F Viel, R. Bertollini (eds.), *Disease Mapping and Risk Assessment for Public Health,* pp. 3-13. Chichester: John Wiley.
[68]LeSage, J., 1997, Bayesian estimation of spatial eautoregressive models. *International Regional Science Review* 20:113-129.
[69]LeSage, J., 2000, Bayesian estimation of limited dependent variable spatial autoregressive models, *Geographical Analysis* 32 (forthcoming).
[70] Liverman, D., E. Moran, R. Rindfuss, P. Stern, Eds., 1998, *People and Pixels, Linking Remote Sensing and Social Science.* Washington, D.C.: National Academy Press.
[71] Martin, D., 1996, *Geographic Information Systems: Socioeconomic Applications.* London: Routledge.
[72] Martin, R., 1999, The new geographical turn in economics: some critical reflections, *Cambridge Journal of Economics,* 23:65-91.
[73] Messner, S., L. Anselin, R. Baller, D. Hawkins, G. Deane, S. Tolnay, 1999, The spatial patterning of county homicide rates: an application of exploratory spatial data analysis, *Journal of Quantitative Criminology,* 15: 423-450.
[74] Miller, H., 1999, Potential contributions of spatial analysis to geographic information systems for transportation(GIS-T), *Geographical Analysis* 31:373-399.
[75] Mitas, L. and H. Mitasova , 1999, Spatial interpolation. In P. Longley, M. Goodchild, D. Maguire and D. Rhind (eds.), *Geographical Information Systems,* 2nd Edition, pp. 481-492. New York: John Wiley.
[76] Morenoff, J. and R. J. Sampson, 1997, Violent crime and the spatial dynamics of neighborhood transition: Chicago 1970-1990, *Social Forces,* 76:31-64.
[77] Müller, W., 1998, *Collecting Spatial Data: Optimum Design of Experiments for Random Fields.* Heidelberg: Physica-Verlag.
[78] Nelson, G. and D. Gray, 1997, *The use of spatial analysis at the World Bank.* Environmental Economics Series Paper No. 56. Washington, DC: The World Bank.
[79] O'Loughlin, J., M.D. Ward, C.L. Lofdahl, J.S. Cohen, D.S. Brown, D. Reilly, K.S. Gleditsch and M. Shin 1999, The diffusion of democracy, 1946-1994. *Annals of the Association of American Geographers,* 88:545-574.
[80]Openshaw, S., 1998, Some trends and future perspectives for spatial analysis in GIS,*Geographic Information Sciences,* 4:5-13.
[81]Openshaw, S. and P. Taylor, 1979, A million or so correlation coefficients: three experiments on the modifiable areal unit problem, In N. Wrigley (ed.), *Statistical Methods in the Spatial Sciences,* pp. 127-144. London: Pion.
[82]Pace, R.K. and R. Barry, 1997, Quick computation of spatial autoregressive estimators, *Geographical Analysis,* 29:232-246.
[83]Pinkse, J., 2000, Moran-flavored tests with nuisance parameters, In L. Anselin and R. Florax (eds.), *Advances in Spatial Econometrics.* Heidelberg: Springer-Verlag (forthcoming).
[84] Pinkse, J. and M. Slade, 1998, Contracting in space: an application of spatial statistics to discrete-choice models, *Journal of Econometrics,* 85:125-154.
[85] Pleskovic, B., 1999, Guest editorial: challenges for the new economic geography in the twenty-first century. *International Regional Science Review,* 22:139-141.
[86] Sampson, R. J., J. Morenoff and F. Earls, 2000, Beyond social capital: neighborhood mechanism and structural sources of collective efficacy for children, *American Sociological Review* (forthcoming).
[87] Sen, A. and T. E. Smith, 1995, *Gravity Models of Spatial Interaction Behavior.* New York: Springer-Verlag.
[88]Starr, H., 1991, Democratic dominoes: diffusion approaches to the spread of democracy, *Journal of Conflict Resolution,* 35: 6-31.
[89] Symanzik, J., T. Kötter, S. Schmelzer, S. Klinke, D. Cook, D. Swayne, 1999, Spatial data analysis in the dynamically linked ArcView/Xgobi/XploRe environment, *Computing Science and Statistics* (forthcoming).
[90] Talen, E., 1999, Constructing neighborhoods from the bottom up: the case for resident-generated GIS. *Environment and Planning, B,* 26:533-554.
[91]Thrall, G., 1998, GIS applications in real estate and related industries, *Journal of Housing Research,* 9:33-59.
[92] Tukey, J., 1977, *Exploratory Data Analysis.* Reading: Addison-Wesley.
[93] Waller, L., B. Carlin and A. Gelfand, 1997, Hierarchical spatio-temporal mapping of disease rates, *Journal of the American Statistical Association,* 92:607-617.
[94] Ward, M. D., 1992, *The New Geopolitics.* London: Gordon and Breach.
[95] Weisburd, D. and T. McEwen, Eds. ,1998, *Crime Mapping and Crime Prevention.* New York: Criminal Justice Press.
[96] Weiss, M., 1996, Precision farming and spatial economic analysis: research challenges and opportunities, *American Journal of Agricultural Economics,* 78:1275-1280.
[97] Wood, C. and D. Skole, 1998, Linking satellite, census, and survey data to study deforestation in the Brazilian Amazon, In D. Liverman, E. Moran, R. Rindfuss, and P. Stern (eds.), *People and Pixels: Linking Remote Sensing and Social Science,* pp. 70-93. Washington, DC: National Academies Press.
[98] Zhang, Z. and D. Griffith, 1997, Developing user-friendly spatial statistical analysis modules for GIS, an example using ArcView, *Computers, Environment and Urban Systems* , 21:5-29.