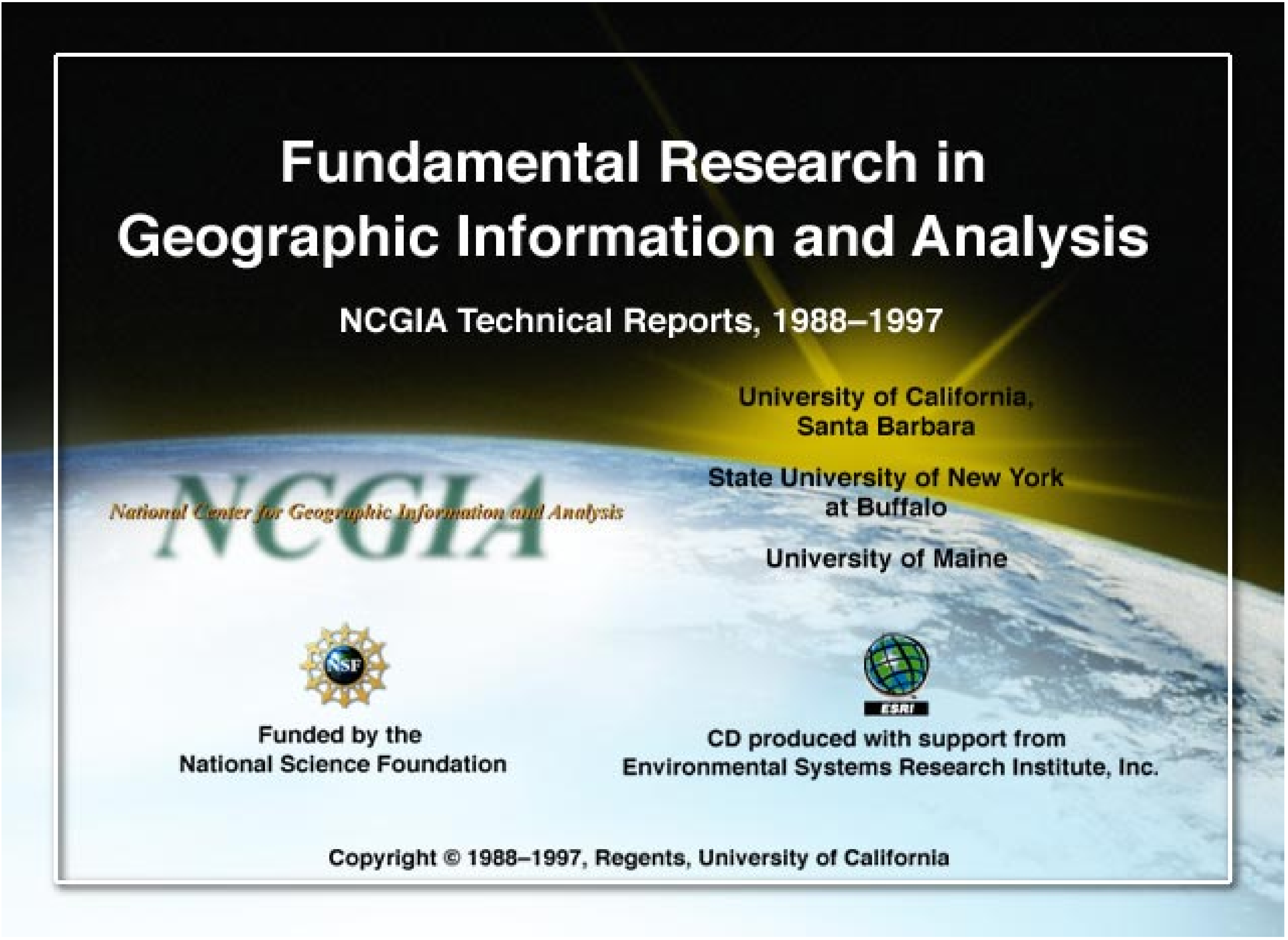

Fundamental Research in Geographic Information and Analysis

NCGIA Technical Reports, 1988–1997

University of California,
Santa Barbara

State University of New York
at Buffalo

University of Maine

NCGIA
National Center for Geographic Information and Analysis

Funded by the
National Science Foundation

CD produced with support from
Environmental Systems Research Institute, Inc.

WHAT IS SPECIAL ABOUT SPATIAL DATA?

ALTERNATIVE PERSPECTIVES ON SPATIAL DATA ANALYSIS

by

Luc Anselin
Department of Geography and Department of Economics
and
National Center for Geographic Information and Analysis
University of California
Santa Barbara, CA 93106

Acknowledgment: Paper prepared for presentation at the Spring 1989 Symposium on *Spatial Statistics, Past, Present and Future*, Department of Geography, Syracuse University. The research reported on in this paper was supported in part by Grants SES 86-00465 and SES 87-21875 from the National Science Foundation, and by the National Center for Geographic Information and Analysis (NCGIA). An earlier version was presented at the NCGIA Specialist Meeting on Accuracy of Spatial Databases, Montecito, CA, December 13-16,1988.

WHAT IS SPECIAL ABOUT SPATIAL DATA?
ALTERNATIVE PERSPECTIVES ON SPATIAL DATA ANALYSIS

INTRODUCTION

The analysis of spatial data has always played a central role in the quantitative scientific tradition in geography. Recently, there have appeared a considerable number of publications devoted to presenting research results and to assessing the state of the art. For example, at an elementary level, Goodchild (1987a), Griffith (1987a), and Odland (1988) introduce the concept of spatial autocorrelation, and Boots and Getis (1988) review the analysis of point patterns. At more advanced levels, Anselin (1988a) and Griffith (1988) deal with a wider range of methodological issues in spatial econometrics and spatial statistics. Extensive reviews of the current state of the art for different aspects of spatial data analysis are presented in Anselin (1988b), Anselin and Griffith (1988), Getis (1988), Griffith (1987b), and Odland, Golledge and Rogerson (1989). In addition, spatial data analysis has received considerable attention as an essential element in the development of Geographic Information Systems (GIS), e.g., in Goodchild (1987b) and Openshaw (1987), and as an important factor in regional modeling, in Anselin (1989a).

In this paper, I will take some distance from specific methods and techniques, and instead outline a few general ideas on fundamental issues related to the distinctive characteristics of *spatial* data analysis, as opposed to data analysis in general. I will focus on two issues that are often overlooked in technical treatments of the methods of spatial statistics and spatial econometrics. One is the relevance for spatial data analysis of the ongoing debate about methodology in the disciplines of statistics and econometrics. I will review and evaluate a number of different approaches towards modeling and analyzing spatial data, and put them in the context of the debate. Recent examples of the opposing viewpoints that are taken in this debate can be found in Leamer (1978), Hendry (1980), Sims (1980, 1982), Lovell (1983), Swamy et al. (1985), Zellner (1985, 1988), Efron (1986), Pagan (1987), Kloek and Haitovsky (1988), and Durbin (1988). The second issue is much narrower and pertains to the role of *spatial errors* in modeling and analysis. This topic has recently received considerable attention in the context of GIS (e.g., as evidenced in the NCGIA Research Initiative on "errors in spatial databases”), but many aspects of its relation to spatial data analysis remain to be explored.

The discussion in this paper is not intended to be comprehensive, but it is selective in the sense that I will focus on issues that seem to be most relevant to current modeling practice and most promising to lead to future research advances. Clearly, this selective treatment reflects my own bin and interests, and is focused on applications in regional science and analytical human geography.

The remainder of the paper consists of six sections. First, I formulate some general remarks on two opposing viewpoints regarding spatial analysis and spatial data: a data-driven approach vs. a model-driven approach. This is followed by a review of a number of competing inferential frameworks that can be used as the basis for spatial data analysis. Next, I focus on spatial errors and on the implications of various forms of spatial errors for spatial data analysis. I close with some concluding remarks on future research directions in spatial statistics and spatial econometrics.

SPATIAL ANALYSIS AND SPATIAL DATA

In general terms, spatial analysis can be considered to be the formal quantitative study of phenomena that manifest themselves in spare. This implies a focus on location, area, distance and interaction, e.g., as expressed in Tobler's (1979) First Law of Geography, where "everything is related to everything else, but near things are more related than distant things." In order to interpret what "near" and "distant" mean in a particular context, observations on the phenomenon of interest need to be referenced in space, e.g., in terms of points, lines or areal units.

There are two opposite approaches towards dealing with spatially referenced data (Anselin 1986b; Haining 1986). In one, which I will call the *data-driven* approach, information is derived from the data without a prior notion of what the theoretical framework should be. In other words, one lets the "data speak for themselves" (Gould, 1981) and attempts to derive information on spatial pattern, spatial structure and spatial interaction without the constraints of a pre-conceived theoretical notion.

In most respects, this approach falls under the category of "exploratory data analysis" (EDA) popularized by Tukey (1977) and Mosteller and Tukey (1977). It is also similar to the philosophy underlying time series analysis and forecasting of the Box-Jenkins (1976) type, and its extensions to vector autoregressive processes and the. like (e.g., Doan et al., 1984; and the critique of Cooley and LeRoy, 1985).

The data-driven approach in spatial analysis is reflected in a wide range of different techniques, such as point pattern analysis (Getis and Boots, 1978; Diggle, 1983), indices of spatial association (HubeM 1985; Wartenberg, 1985), kriging (Clark, 1979), spatial adaptive filtering (Foster and Gorr, 1986), and spatial time series analysis (Bennett, 1979). All these techniques have two aspects in common. First, they compare the observed pattern in the data (e.g., locations in point pattern analysis, values at locations in spatial autocorrelation) to one in which space is irrelevant. In point pattern analysis this is the familiar Poisson pattern, or "randomness," while in many of the indices of spatial association it is the assumption that an observed data value could occur equally likely at each location (i.e., the basis for the randomization approach). The second common aspect is that the spatial pattern, spatial structure, or form for the spatial dependence are derived from the data only. For example, in spatial time series analysis, the specification of the autoregressive and moving average lag lengths is derived from autocorrelation indices or spatial spectra.

The data-driven approach is attractive in many respects, but its application is not always straightforward. Indeed, the characteristics of spatial data (dependence and heterogeneity) often void the attractive properties of standard statistical techniques. Since most EDA techniques are based on an assumption of independence, they cannot be implemented uncritically for spatial data. In this respect, it is also important to note that dependence in spare is qualitatively more complex than dependence in the time dimension, due to its two-directional nature. As a consequence, many results from the analysis of time series data will not apply to spatial data. As discussed in detail in Hooper and Hewings (1981), the extension of time series analysis into the spatial domain is limited, and only applies to highly regular processes. It goes without saying that most data in empirical spatial analysis for irregular areal units do not fit within this restrictive framework.

The second approach, which I will call *model-driven*, starts from a theoretical specification, which is subsequently confronted with the data. The theory in question may be spatial (e.g., a spatial process or a spatial interaction model, as in Haining 1978, 1984) or largely a-spatial (e.g., a multi-regional economic model, as in Folmer 1986), but the important characteristic is that its estimation or calibration is carried out by means of spatial data. The properties of this *data*, i.e., spatial dependence and spatial heterogeneity, necessitate the application of specialized statistical (or econometric) techniques, irrespective of the nature of the theory in the model.

Most of the methods that I would classify under this category deal with estimation and specification diagnostics in linear models in general, and regression models in particular (e.g., Cliff and Ord 1981; Anselin 1980, 1988a). The main conceptual problem associated with this approach is how to formalize the role of "space." This is reflected in three major methodological problems, which are still largely unresolved to date: the choice of the spatial weights matrix (Anselin 1984, 1986a); the modifiable areal unit problem (Openshaw and Taylor 1979, 1981); and the boundary value problem (Griffith 1983, 1985).

In order for the data-driven or the model-driven approaches to be operational, the various tests, diagnostics and estimators need to be incorporated in an inferential framework. More precisely, the uncertainty associated with a random variable, sampling error, or any other stochastic aspect of the data analysis needs to be assessed within a consistent framework that forms a logical basis for decisions. A number of competing frameworks have been suggested. They are discussed next.

INFERENTIAL FRAMEWORKS IN SPATIAL DATA ANALYSIS

Spatial data analysis is not immune from the implications of the philosophical debates that go on in the broader disciplines of statistics and econometrics. Although the results of applied and empirical work are often presented as if only one particular view of statistics existed, there are in fact many competing perspectives (or even paradigms). Rather than repeating the various philosophical arguments, I will outline five dimensions of conflict or competition, and discuss some implications of the alternative viewpoints for spatial data analysis. Some of these dimensions are more fundamental than others, but all have direct applications to the practice of spatial statistics and spatial econometrics.

Classical vs. Bayesian Inference

The debate between the classical (Neyman-Pearson) and Baysian approaches to statistical inference (or decision making) is undoubtedly the most fundamental one ongoing in the discipline. The arguments of both sides are well known and a compromise does not seem likely in the near future (e.g., Efron 1986; Durbin 1988; Zellner 1988). In a nutshell, the classical approach is "objective," and practical, but fraught with philosophical problems when applied in a strict sense: problems with multiple comparisons, the need to assume a "true" model, etc… On the other hand, the Bayesian approach is generally considered to be superior in terms of overall consistency and as a perspective on “learning," but is "subjective" and difficult to apply to many practical problems, due to the need to construct complex prior distributions and to carry out numerical integration in multiple dimensions.

In spatial data analysis, the classical approach is by far the dominant one. The Bayesian perspective is the exception, and has found only limited application, although some Bayesian concepts are fairly familiar in image processing of remotely sensed data (e.g., Richards 1986; for applications to spatial data analysis in human geography, see, e.g., March and Batty 1975; OdIand 1978; Hepple 1979; and Anselin 1982, 1988b). However, an uncritical application of the Neyman-Pearson classical inferential framework to spatial data analysis is inappropriate in a lot of respects. The many assumptions, judgements and multiple comparisons carried out in the practice of estimation and data analysis (both data-driven as well as model-driven) make a mockery out of the rigorous and elegant probabilistic calculus that underlies the classical approach (for more details, see Anselin 1988b). It would therefore seem, at least from a conceptual viewpoint, that a number of spatial "problems" could be most fruitfully attacked from a Bayesian perspective. Examples are pattern recognition, or 'learning" from data in general, the prior assumptions about a spatial weights matrix, spatial interpolation and dealing with boundary effects. However, the practical implementation of a Baysian analysis of these issues is not straightforward.

Specifically, it has so far not been possible to develop useful prior distributions for the full range of patterns of spatial dependence (spatial weight matrices) that would be operational in spatial data analysis. Overall, dealing with the two-directional nature of spatial dependence in a Bayesian framework is still very much an unresolved research topic.

Parametrics vs. Non-Parametrics and/or Robustness

In applied spatial data analysis, the standard assumptions of normality and of perfect knowledge of the model specification are often rather crude abstractions of reality. Consequently, the relevance of a strict parametric approach has been increasingly questioned, and a non-parametric, qualitative or robust perspective has sometimes been suggested as an alternative, e.g., in Gould (1981), Costanzo (1983), Nijkamp, Leitner and Wrigley (1985) and Knudsen (1987). However, it is not as if non-parametric and robust procedures have not been introduced to spatial analysis. On the contrary, a number of well known indices for spatial association have been based on randomization, permutation and other non-parametric techniques. Examples range from a robust Moran index in Cliff and Ord (1973), and Sen and Soot (1977), to the general measures of spatial association in Hubert et al (1981, 1985). Most of these methods would fall under the data-driven category of spatial data analysis. However, there have been some recent applications in the model-driven category as well, primarily based on the use of the Jackknife and bootstrap estimation techniques, e.g., in Stetzer (1982), Folmer and Fischer (1984) and Anselin (1989b).

In spite of the concerns about its appropriateness, the parametric approach remains the most common one in spatial data analysis. Most tests are based on an underlying distribution which is normal (for values) or Poisson (for point patterns) and the estimation method of choice is the maximum likelihood technique. As is well known, the parametric approach is optimal in a number of ways if the underlying assumptions are indeed satisfied. It is when this is not the case that problems occur. Since the robust and non-parametric techniques are not grounded in such a restrictive set of assumptions, they remain valid in a wider range of situations. However, this robustness comes at the cost of a loss in generality and precision. For example, the spatial association indices that are based on a permutation approach only pertain to the data at hand, and cannot be generalized to hold for a "population." Similarly, the variance estimates for parameters obtained by means of the bootstrap or Jackknife will tend to be larger than for the maximum likelihood approach, and thus will lead to a more conservative inference (Le, it will be "harder’ to find significant coefficients).

A further obstacle for the acceptance of robust or non-parametric techniques in spatial data analysis is that a great many of the techniques developed in mathematical statistics and econometrics (e.g., as reviewed in Huber 1981; Koenker 1982; Efron 1982; and Robinson 1988) are not directly transferable, since they are based on an assumption of independence. An appropriate "spatial” theoretical framework for robust analysis remains to be developed.

Random Sample vs. Stochastic Process

The dependence that is inherent in many (if not most) spatial data runs directly counter to the postulate of a random sample of independent observations on which most common statistical procedures are based. Nevertheless, much applied spatial data analysis still proceeds as if the standard assumptions held, and notions of sampling error, sampling variance, etc… abound in the empirical literature. Clearly, this is incorrect, and the loss of information that results from the dependence in the observations should be accounted for.

In most instances, the proper perspective is not to consider spatial data as a random sample with many observations, but instead as a single realization of a stochastic process. In contrast to the sampling approach, where each observation is taken to provide an independent piece of information, the dependence (and heterogeneity) embodied in a stochastic process implies that only one observation is available, i.e., the full spatial pattern (or space-time pattern) of values. Provided that the underlying stochastic process is sufficiently stable (stationary, isotropic, etc ...) the observed pattern will yield information on the characteristics of that process. In contrast to the random sampling approach, where the notion of independence is exploited to derive exact statistical properties for estimates and hypothesis tests, an asymptotic reasoning is needed in the stochastic process approach. Specifically, the theory of mixing processes, which allow a degree of dependence as well as heterogeneity, forms a solid basis for the inference for spatial stochastic processes (for details, see Anselin 1988a, Chapter 5).

The consequence of spatial dependence is that the observations contain less information than if there had been independence. In other words, in order to obtain the same degree of information as in an independent set of observations, a larger data set of dependent observations will be needed. Sometimes, the latter can be transformed into the former, by deleting observations that are contiguous or within a given distance of each other. For example, if only those observations are selected that are far enough apart so that no dependence can reasonably be expected (i.e., dependence related to distance only) this new "sample" can be considered to be independent for most practical purposes. This "re-coding" lies at the basis of the so-called "conditional” approach to spatial modeling (Haining, 1986). Its advantage is that most standard statistical techniques can be applied unchanged to the re-coded data. 'However, the re-coding itself is not unique and somewhat arbitrary. Also, this is only a practical approach if the loss of information from discarding the "dependent" observations is not critical. Unfortunately, in many practical situations such a luxury does not exist and the "simultaneous" (joint probability) stochastic process approach is the only feasible parametric framework.

A related issue is the extent to which spatial data constitute a sample, a realization of a stochastic process, or instead form the complete population of interest It is sometimes argued that the latter is the only correct perspective and thus that no inferential statistics are possible, but only a descriptive approach is valid (e.g., Summerfield 1983)i Although this may be an acceptable viewpoint in the case of extreme heterogeneity (i.e., each place is "unique" and no generalization is possible) it is more the exception than the rule. There are two crucial issues that need to be considered. The first pertains to the imperfect nature of measurement, and the inherent error (or noise). Since we observe a mixture of signal and noise in empirical practice, the stochastic nature of the data can easily be generated from the randomness in errors of measurement As a consequence, the population in question pertains to the family of stochastic processes that may have generated a particular error pattern. Thus, a "statistical" approach is the only way in which conclusions can be formed about the underlying "signal" and an understanding of *spatial errors* is crucial.

The second issue pertains to the nature of space as a framework within which observations are ordered. In essence, the spatial unit of observation needs to be a representative unit for the phenomenon" that is under study. Only then will it be possible to formulate and test general statements about "space." The real issue is whether the observations at hand are compatible with the complexity of the phenomenon of interest. If they are not, this does not mean that a statistical approach should be rejected, but rather that other types of data are needed. For example, this may necessitate the collection of micro-behavioral data to avoid problems of ecological fallacy, or may require the extension of a cross-section into the time dimension in order to formulate general conclusions about a specific region.

Finite-Sample vs. Asymptotics

The stochastic process approach to spatial data analysis is based on asymptotic properties for an "abstract" and infinitely large data set. This conceptual framework contrasts sharply with the reality of small data sets with a finite number of observations. Two issues merit some consideration. The first is practical and pertains to the extent to which the asymptotic properties are valid in finite samples. As is well known, this is not necessarily the case, and many properties of equivalence and optimality of asymptotic tests and estimators are not reflected in realistic data sets. Moreover, few analytic results are available and the properties of a number of approximations are questionable (see, e.g., Taylor 1983; and Anselin 1988b for spatial data). In other words, considerable caution (a conservative inference) is needed when interpreting the findings of spatial data analysis that are based on asymptotic properties.

The second issue related to asymptotics is more conceptual and pertains to the relevance of the notion of an infinitely large data set for spatial analysis. In essence, an asymptotic reasoning is only meaningful if an infinitely large number of replications of the observed spatial units can be conceived of. While this is fairly straightforward in the case of regularly spaced points or grids, it is not at all obvious' for irregular areal units (e.g., a given set of counties in a state). There are two approaches to this conceptual problem. In one, the data for irregular spatial units are transformed (interpolated) to regular spatial units. Although this forms an elegant solution to the problem, it is only valid if the underlying process is sufficiently smooth and homogeneous. In the other approach, the dependence and heterogeneity in the data are recognized as a limiting factor and the only way to obtain meaningful information from the observations is by adding an additional dimension, i.e., the time dimension. In other words, by pooling time series data for a fixed set of cross-sectional units, the asymptotics in the time dimension provide the framework to carry out statistical inference about the spatial dimension. In either case, it is necessary to evaluate whether the complexity of the proposed hypotheses or models is compatible with the information available in the data. Unfortunately, in many situations encountered in applied empirical work this will not be the case. In those instances the stochastic framework for inference will be suspect, and give rise to legitimate concerns about the relevance of a "statistical" approach.

Analytics vs. Computing Power

A final issue which has come to the fore as a result of the recent advances in computer technology is the choice between procedures based on rigorous analytics and those that replace the analytics by computation. The latter have led to the development of combinatorial methods and re-sampling schemes, in which the stochastic properties are derived from a large number of replications of pseudo-data (e.g., Efron 1979; Hubert 1985; Knudsen 1987). With the advent of large spatial data bases and geographic information systems, the distinction between description, analysis, modeling and simulation has become blurred. The technological possibilities are virtually unbounded, and have- opened up new horizons for spatial data analysis. An example of a recent development in this respect is the creation of a so-called "geographical analysis machine" (GAM), as a combination of a GIS, a spatial statistical analysis and expert system which is designed to carry out an automated spatial data analysis (Openshaw et al. 1987). This concept has many attractive features, but in its current form, the GAM is still rudimentary and limited to a specific application. Important further developments are needed before this a-theoretical approach will be able to replace (or complement) the more traditional analytic approach for a wide range of spatial data analysis problems.

SPATIAL ERRORS

Basic to both the data-driven and the model-driven analysis of spatial data is an understanding of the stochastic properties of the data. The use of "space" as the organizing framework leads to a number of features that merit special attention, since they are different from what holds for a-spatial or time series data. The most important concept in this respect is that of *error*, or, more precisely for data observed in space, *spatial error*. The distinguishing characteristics of spatial error have important implications for description, explanation and prediction in spatial analysis. I discuss some of these in the next section. In this section, I present a simple taxonomy of the nature of spatial errors and outline some alternative perspectives on how error can be taken into account.

The Nature of Spatial Errors

Spatial errors can be due to *measurement error* or *specification error*, or to a combination of both. Measurement error occurs when the location or the value of a variable are observed with imperfect accuracy. The former is an old cartographic problem and is still very relevant in modem geographic information systems (e.g., errors due to a lack of precision in digitizing). The main problem is that the geometric and graphical representation of the location of points, lines or areal boundaries (i.e., a map) gives an imperfect impression of the uncertainty associated with errors in their measurement. Since these locational features are important elements in the evaluation of distance and relative position, and in the operations of areal aggregation and interpolation, the associated measurement error will affect many of the "values" generated in a spatial information system as well. Although similar errors occur in the time dimension, they are much simpler to take into account since they only propagate in one direction.

Other spatial errors of measurement have to do with the imperfect way in which data on socio-economic phenomena are recorded and grouped in spatial units of observation (e.g., various types of administrative units). This interdependence of location and value in spatial data leads to distinctively *spatial* characteristics of the errors. These are the familiar spatial dependence and spatial heterogeneity. Dependence is mostly due to the existence of spatial spillovers, as a result of a miss-match between the scale of the spatial unit of observation and the phenomenon of interest (e.g., continuous processes represented as points, or processes extending beyond the boundaries of administrative regions). Heterogeneity is due to structural differences between locations and leads to different error distributions (e.g., differences in accuracy of census counts between low-income and high-income neighborhoods).

Specification error is particular to the model-driven approach in spatial data analysis. It pertains to the use of a wrong model (e.g., recursive vs. simultaneous), an inappropriate functional form (e.g., linear vs. nonlinear), or a wrong set of variables. In essence it is no different from miss-specification in general, but it generates spatial patterns of error due to the use of spatial data. These spatial aspects can occur as a result of ignoring location-specific phenomena, spatial drift, regional effects or spatial interaction. When a false assumption of homogeneity is forced onto a model in those instances, spatial heterogeneous errors will result. Similarly, when the spatial scale or extent of a process does not correspond to the scale of observation, or when the nature of a process changes with different scales of observation, spatial dependent errors will be generated.

Perspectives on Spatial Errors

The treatment of spatial errors in data analysis is fundamentally different between the data-driven and the model-driven approaches. In the data-driven approach, errors are considered to provide *information*. The focus of attention is on how the spatial pattern of the errors relates to data generation processes. For example, in attempts to provide measures of uncertainty for spatial information in a GIS, the spatial pattern of errors is related to data collection and manipulation procedures. The spatial pattern of errors can often provide insight into the form of the underlying substantive spatial process, e.g., as exploited in the model identification stage of a spatial time series analysis. Important and still unresolved research questions deal with the formulation of useful spatial error distributions, in which error is related to location, distance to reference locations, area, etc...

In the model-driven approach to spatial data analysis, the errors are considered to be a *nuisance*. The main focus is on how to identify the spatial distribution of the error process and to eliminate e the effect of errors on the statistical inference. In other words, once errors are identified, they are eliminated by means of transformations, corrections or filters. Alternatively, robust estimation and test procedures can be applied that are no longer sensitive to the effect of errors. A major research question in this respect is how diagnostics can be developed that are powerful in detecting various types of errors, and are able to distinguish between them (e.g., to distinguish between spatial dependence and spatial heterogeneity, or “real" vs. "apparent" contagion).

IMPLICATIONS OF SPATIAL ERRORS FOR SPATIAL DATA ANALYSIS

The presence of errors with a distinctive spatial pattern has obvious implications for the analysis of spatial data. These implications vary between the analysis of spatial pattern, the estimation and prediction of spatial models, and their validation.

Analysis of Spatial Pattern

Given the importance of distance and contiguity in the analysis of spatial pattern, errors of measurement in the location of points, lines and areal units will greatly affect the distributional properties of tests and other indices. This aspect of spatial error is largely ignored in current statistical practice, but merits closer attention, particularly in light of the increased availability of large computerized spatial data bases with the explosion of the GIS field. Some indices of spatial pattern and spatial association that are routinely derived in a GIS (e.g., based on nearest-neighbors) provide a misleading sense of precision, since they ignore the uncertainty associated with the location of spatial units themselves. Conceptually, the solution to this problem is straightforward, in that a spatial distribution needs to be specified for each location (and the associated values). However, the choice of the most appropriate distribution and its effect 'on the properties of the various spatial statistics are still largely unresolved topics of research.

Estimation and Prediction

The effect of spatial errors on the estimation and prediction of standard linear models is probably the best understood aspect of spatial data analysis. In particular, for the linear regression model with normally distributed disturbance terms, many tests and estimators have been developed (see Anselin and Griffith 1988, for a review). In those models, the error is taken to pertain to the dependent variable only and its effect is incorporated in the regression disturbance term. The more realistic situation where error is present in both dependent and independent variables has received much less attention, and is considerably more complex. The specification of the interaction between the various spatial errors is largely unresolved, and so far only a robust estimation approach seems to hold promise.

Most methodological results obtained so far are also limited to the normal distribution case. Spatial effects in models with limited dependent variables, censored and truncated distributions, or in models for count data have been largely ignored. A major problem in this respect is that multivariate dependent distributions other than the normal are highly complex. Moreover, their application in an operational context is often hampered by limitations on carrying out numerical integration in multiple dimensions. Since the non-normal case is probably the rule rather than the exception in actual spatial data, a considerable agenda of research questions remains to be addressed.

Model Validation

In model validation, the focus is on assessing the uncertainty associated with the output (interpretation) of alternative specifications. Clearly, this will be a function of the probabilistic model that has been adopted for the underlying (unobserved) spatial errors. A particular problem in spatial data analysis is how to provide a meaningful summary measure of spatial accuracy. If spatial heterogeneity is present, the accuracy is likely to vary systematically by location. On the other hand, if spatial dependence is present, the accuracy at one location Will be affected by the accuracy associated with "neighboring" locations. A summary or holistic measure of accuracy will be an imperfect reflection of this partitive (observation by observation) accuracy. What is needed is a meaningful objective function (loss or risk function) that incorporates the relative importance of accuracy for particular locations or regions in space. It is unlikely that such an objective function can be developed with universal applicability, but instead, a flexible approach can be taken which is consistent with the use of spatial information systems as decision support systems.

CONCLUSION

The wide array of philosophical and methodological dilemmas that confront the analysis of spatial data necessitates an eclectic perspective. Many different ways of looking at a data set or at a model specification should be compared, and sensitivity analysis should play a central role. If different approaches yield the same conclusions, one can be fairly confident that meaningful insights have been gained. On the other hand, if the statistical findings turn out to be very sensitive to the approach taken, there is likely to be something wrong with the data and/or with the model and not much faith should be put in the precise quantitative results.

The characteristics of errors that affect observations of spatial data clearly motivate the need for a specialized methodology of spatial statistics and spatial econometrics. However, much of the current state of the art in these fields pertains to highly artificial and rather simplistic data structures. A major emphasis of future research should be to focus on *realistic* perspectives an spatial data. With the vast power of a user-friendly GIS increasingly in the hands of the non-specialist, the danger is great that the "wrong" kind of spatial statistics will become the accepted practice. Since the "easy" problems have more or less been solved, a formidable challenge lies ahead.

REFERENCES

Anselin, Luc, 1980. *Estimation Methods for Spatial Autoregressive Structures*. Ithaca, NY: Cornell University, Regional Science Dissertation and Monograph Series #8.

Anselin, Luc, 1982. "A Note on Small Sample Properties of Estimators in a First-Order Spatial Autoregressive Model." *Environment and Planning* A, 14, 1023-1030.

Anselin, Luc, 1984. "Specification Tests on the Structure of Interaction in Spatial Econometric Models." *Papers, Regional Science Association*, 54, 165-182.

Anselin, Luc, 1986a. "Non-Nested Tests on the Weight Structure in Spatial Autoregressive Models: Some Monte Carlo Results." *Journal of Regional Science*, 26, 267-284.

Anselin, Luc, 1986b. "Some Further Notes on Spatial Models and Regional Science." *Journal of Regional Science*, 26, 799-802.

Anselin, Luc, 1988a. *Spatial Econometrics. Methods and Models*. Dordrecht: Kluwer Academic.

Anselin, Luc, 1988b. "Niodel Validation in Spatial Econometrics: A Review and Evaluation of Alternative Approaches." *International Regional Science Review*, 11,279-316.

Anselin, Luc, 1989a. "Quantitative Methods in Regional Science: Perspectives on Research Directions." Paper Presented at-a Plenary Session of the Third World Congress of the Regional Science Association, April 2-7, Jerusalem, Israel.

Anselin, Luc, 1989b. "Some Robust Approaches to Testing and Estimation in Spatial Econometrics." *Regional Science and Urban Economics*, 19 (forthcoming).

Anselin, Luc and Daniel A- Griffith, 1988. "Do Spatial Effects Really Matter in Regression Analysis?" *Papers, Regional Science Association*, 65, 11-34.

Bennett, Robert, 1979. *Spatial Time Series*. London: Pion.

Boots, Barry N. and Arthur Getis, 1988. *Point Pattern Analysis*. Newbury Park, CA: Sage Publications.

Box, G. and G. Jenkin , 1976. *Time Series Analysis, Forecasting and Control*. San Francisco: Holden Day.

Clark, I., 1979. *Practical Geostatistics*. London: Applied Science Publishers.

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

Cliff, A and J.K Ord, 1981. *Spatial Processes; Models and Applications.* London: Pion.

Cooley, T. and S. LeRoy, 1985. "A theoretical Macro-Econometrics: A Critique." *Journal of Monetary Economics*, 16, 283-308.

Costanzo, C. Michael, 1983. "Statistical Inference in Geography: Modem Approaches Spell Better Times Ahead." *The Professional Geographer*, 35, 158-165.

Diggle, P., 1983. *Statistical Analysis of Spatial Point Patterns*. New York: Academic Press.

Doan, T., R. Litterman, and C. Sims, 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews*, 3, 1-100 (with discussion).

Durbin, James, 1988. "Is a Philosophical Consensus for Statistics Attainable?" *Journal of Econometrics*, 37,51-61.

Efron, B., 1979. "Computers and the Theory of Statistics: Thinking the Unthinkable." *SIAM Review*, 21, 460-480.

Efron, B., 1982. *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia, PA.: Society for Industrial and Applied Mathematics (SIAM).

Efron, B., 1986. "Why isn't Everyone a Bayesian?" *The American Statistician*, 40, 1-11 (with discussion).

Folmer, Hendrik, 1986. *Regional Economic Policy.- Measurement of Its Effect*. Dordrecht: Martinus Nijhoff.

Folmer, H. and M. Fischer, 1984. 'Bootstrapping in Spatial Analysis." Paper Presented at the Symposium of the IGU Working Group on Systems Analysis and Mathematical Models, Besancon, France.

Foster, S. and W. Gorr, 1986. "An Adaptive Filter for Estimating Spatially-Varying Parameters: Application to Modeling Police Hours Spent in Response to Calls for Service." *Management Science*, 32, 878-889.

Getis, Arthur, 1988. "Second-Order Theory in Spatial Analysis." Paper Presented at the Symposium of the IGU Working Group on Mathematical Models, August 16-19, Canberra, Australia.

Getis, Arthur and Barry Boots, 1978. *Models of Spatial Processes*. London: Cambridge University Press.

Goodchild, Michael, 1987a. *Spatial Autocorrelation*. CATMOG.

Goodchild, Michael, 1987b. "A Spatial Analytical Perspective on Geographical Information Systems." *International Journal of Geographical Information Systems*, 1, 327-334.

Gould, Peter, 1981. "Letting the Data Speak for Themselves." *Annals of the Association of American Geographers*, 71, 166-176.

Griffith, Daniel A, 1983. The Boundary Value Problem in Spatial Statistical Analysis." *Journal of Regional Science*, 23, 377-387.

Griffith, Daniel A, 1985. "An Evaluation of Correction Techniques for Boundary Effects in Spatial Statistical Analysis: Contemporary Methods." *Geographical Analysis*, 17,81-88.

Griffith, Daniel A, 1987a. *Spatial Autocorrelation: A Primer*. Washington, D.C.: Association of American Geographers.

Griffith, Daniel A., 1987b. 'Toward a Theory of Spatial Statistics: Another Step Forward." *Geographical Analysis*, 19,69-82.

Griffith, Daniel A, 1988. *Advanced Spatial Statistics*. Dordrecht: Kluwer Academic.

Haining, Robert, 1978. "Estimating Spatial Interaction Models." *Environment and Planning A*, 10, 305-320.

Haining, Robert, 1984. 'Testing a Spatial Interacting Market Hypothesis." *The Review of Economics and Statistics*, 66,576-583.

Haining, Robert, 1986. "Spatial Models and Regional Science: A Comment on Anselin's Paper and Research Directions." *Journal of Regional Science*, 26, 793-798.

Hendry, David F, 1980. 'Econometrics - Alchemy or Science?" *Economica*, 47, 387-406.

Hepple, 1, 1979. 'Bayesian Analysis of the Linear Model with Spatial Dependence." *In* C. Bartels and R. Ketellapper, *Exploratory and Explanatory Statistical Analysis of Spatial Data*, pp. 179-199. Boston: Martinus; Nijhoff.

Hooper, P. and GJ-D. Hewings, 1981. "Some Properties of Space-Time Processes." *Geographical Analysis*, 13,203-223.

Huber, P., 1981. *Robust Statistics*. New York. Wiley.

Hubert, L., 1985. "Combinatorial Data Analysis: Association and Partial Association." *Psychometrika*, 50, 449-467.

Hubert, L, R. Golledge and C_ Costanzo, 1981. "Generalized Procedures for Evaluating Spatial Autocorrelation." *Geographical Analysis*, 13, 224-233.

Hubert, L, R. Golledge, C_ Costanzo and N. Gale, 1985. "Measuring Association Between Spatially Defined Variables: An Alternative Procedure, *Geographical Analysis*, 17,36-46.

Kloek, Teun and Yoel Haitovsky, 198& *Competing Statistical Paradigm in Econometrics*. Special Issue, Journal of Econometrics, 37 (1).

Knudsen, Daniel, 1987. "Computer-Intensive Significance-Testing Procedures." *The Professional Geographer*, 39, 208-214.

Koenker, R., 1982. "Robust Methods in Econometrics." *Econometric Reviews*, 1, 213-255.

Leamer, Edward, 1978. *Specification Searches. Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

Lovell, M.C., 1983. 'Data Mining." *The Review of Economics and Statistics*, 65, 1-12.

March, L and M. Batty, 1975. "Generalized Measures of Information, Bayes' Likelihood Ratio and Jaynes' Formalism." *Environment and Planning B*, .2, 99-105.

Mosteller, F. and J.W. Tukey, 1977. *Data Analysis and Regression*. Reading, Mass: Addison-Wesley.

Nijkamp, P., H. Leitner and N. Wrigley, 1985. *Measuring the Unmeasurable*. Dordrecht: Martinus, Nijhoff.

Odland, John, 1978. "Prior Information in Spatial Analysis." *Environment and PlanningA*, 10, 51-70.

Odland, John, 1988. *Spatial Autocorrelation*. Newbury Park, CA. Sage Publications.

Odland, John, Reginald G. Golledge, and Peter Rogerson, 1989. "Recent Developments in Mathematical and Statistical Analysis in Human Geography." *In* G. Gaile and C. Wilmott, *Geography in America*, pp. 719-745. Columbus, OH: Merrill.

Openshaw.- Stan, 1987. "An Automated Geographical Analysis System." *Environment and Planning A*, 19, 431-436.

Openshaw, Stan and Peter Taylor, 1979. "A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem." *In* N. Wrigley, *Statistical Applications in the Spatial Sciences*, pp. 127-144. London: Pion.

Openshaw, Stan and Peter Taylor, 1981. "The Modifiable Unit Problem." *In* N. Wrigley and R. Bennett; *Quantitative Geography, a British View*, pp. 60-69. London: Routledge and Kegan.

Openshaw, S., M. Charlton, C. Wymer, and A. Craft, 1987. "A Mark 1 Geographical Analysis Machine for the Automated Analysis of Point Data Sets." *International Journal of Geographical Information Systems*, 1, 335-358.

Pagan, Adrian, 1987. 'Three Econometric Methodologies: A Critical Appraisal." *Journal of Economic Surveys*, 1, 2-24.

Richards, John A., 1986. *Remote Sensing Digital Image Analysis.- An Introduction*. New York. Springer Verlag.

Robinson, P, 1988. "Semiparametric Econometrics: A Survey." *Journal of Applied Econometrics*, 3, 35-5 1.

Sen, A. and S. Soot, 1977. "Rank Test for Spatial Correlation." *Environment and PlanningA*, 9,897-903.

Sims, Christopher, 1980. "Macroeconomics and Reality." *Econometrica*, 48, 1-48.

Sims, Christopher, 1982. "Scientific Standards in Econometric Modeling." *In* H. Hazewinkel and A.N.G. Rinnooy Kan, *Current Developments in the Interface. Economics, Econometrics, Mathematics*, pp. 317-337. Dordrecht: D. Reidel Publishing.

Stetzer, F., 1982. 'The Analysis of Spatial Parameter Variation with Jackknifed Parameters." *Journal of Regional Science*, 22, 177-188.

Summerfield, M., 1983. "Populations, Samples and Statistical Inference in Geography." *The Professional Geographer*, 35, 143-149.

Swamy, P, P. R. Conway, and P. von zur Muehlen, 1985. "Me Foundations of Econometrics - Are There Any?" *Econometric Reviews*, 4, 1-6 1.

Taylor, WE, 1983. "On the Relevance of Finite Sample Distribution Theory." *Econometric Reviews*, 1, 213-255.

Tobler, Waldo, 1979. "Cellular Geography." *In* S. Gale and G. Olsson, *Philosophy in Geography*, pp. 379-386. Dordrecht: Reidel.

Tukey, J.W., 1977. *Exploratory Data Analysis. Reading*: Addison-Wesley.

Wartenberg, Daniel, 1985. "Multivariate Spatial Correlation: A Method for Exploratory Geographical Analysis." *Geographical Analysis*, 17, 263-283.

Zellner, Arnold, 1985. “Bayesian Econometrics." *Econometrica*, 53, 253-269.

Zellner, Arnold, 1988. "Bayesian Analysis in Econometrics." *Joural of Econometrics*, 37,27-50.