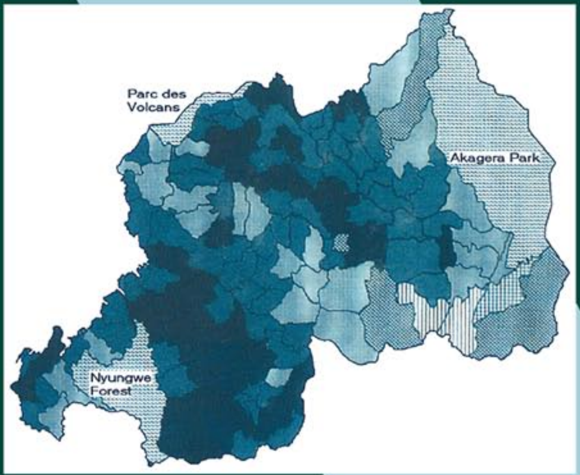
PRACTICAL HANDBOOK

of

Spatial Statistics



CRC PRESS

Editor

Sandra L. Arlinghaus

PRACTICAL HANDBOOK

of

Spatial Statistics

EDITOR-IN-CHIEF
Sandra Lach Arlinghaus, Ph.D.
Institute of Mathematical Geography
and The University of Michigan

SPECIALIST ASSOCIATE EDITOR
Daniel A. Griffith, Ph.D.
Syracuse University

ASSOCIATE EDITORS
William C. Arlinghaus, Ph.D.
Lawrence Technological University

William D. Drake, Ph.D.
The University of Michigan

John D. Nystuen, Ph.D.
The University of Michigan

CRC PRESS

Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Practical handbook of spatial statistics / editor-in-chief, Sandra Lach Arlinghaus ; specialist associate editor, Daniel A. Griffith ; chapter contributors other than editors, Daniel G. Brown ... [et a;].
p. cm.
Includes bibliographical references (p. -) and index.
ISBN 0-8493-0132-7 (alk. paper)
1. Geography—Statistical methods. 2. Spatial analysis (Statistics) I. Arlinghaus, Sandra L. (Sandra Lach). II. Griffith, Daniel A.
G70.3..P73 1995
910'.021 95-24710
CIP

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

Visit the CRC Press Web site at www.crcpress.com

International Standard Book Number 0-8493-0132-7
Library of Congress Card Number 95-24710
Printed in the United States of America 6 7 8 9 10 11 12 13 14 15

Foreword

THE IMPORTANCE OF POSITION IN SPACE

Sandra L. Arlinghaus and John D. Nystuen

"The rain is raining all around,
It rains on land and sea..."

unknown children's song.

Most of us have watched the rain fill potholes in our streets or the snow pile up in the bird baths in our yards as weather commentators report that these are "scattered showers" or "snow flurries." The reports do not tally with our own observations. The problem is that the official forecasts of widespread phenomena, such as rain or snow, are based on a small set of observations at separated locations; the measurements derived from these locations are then attributed to an entire region. But precipitation is spatially and temporally heterogeneous, perhaps even fractal in nature. Finer and finer meshes of observations yield higher and higher variances, yet the pattern is not random in space or time. Nearby locations experience similar intensities of precipitation as rainfall sweeps over a region leaving moist rain tracks. After many repeated rain events, amounts of rain recorded at each location begin to converge. This type of phenomenon calls for a measurement strategy that focuses on the spatial character of widespread phenomena. [Nystuen, McGlothin, and Cook, 1993] Spatial statistics does just that.

On April 29, 1986, U. S. scientists detected the nuclear incident at Chernobyl prior to official Soviet acknowledgment that any event of note had occurred (see Sadowski and Covington, 1987 for a technical discussion). Unusually high energy emissions, well above normal reflected energy levels, were detected by multispectral scanners operating with 30 to 80 meter resolutions from a commercial satellite orbiting the Earth. Immediate reference to data from another satellite with ten meter resolution confirmed that the hotspot was confined to a very small region implying an even higher, very localized, energy release. Calculations revealed that the temperature of the energy release was that of burning graphite. A meltdown was occurring. Spatial, spectral, and temporal precision made this detection possible. A strategy to detect rare events in space and time was employed. Positional accuracy in data recording is a prerequisite for most spatial statistics applications whether the phenomena are widespread or located at unique points.

I. DEPENDENCE OF OBSERVATION ON SPATIAL POSITION

As Griffith notes in the next chapter,

> *Spatial statistics differs from classical statistics in that the observations analyzed are not independent...*

In the case of the rainfall problem, the weather stations scattered across a region are generally viewed, for the purposes of forecasting, as independent. However, we all know that for the most part, they are not independent--the distance between them, and the distance between local residents and rain gauges, matters. The fundamental assumption of spatial statistics acknowledges this difficulty, involving dependence on spatial position, at the outset.

This handbook displays various techniques from spatial statistics and illustrates them using real world data. It also illustrates how techniques for spatial statistics differ from, or are similar to, corresponding techniques for classical statistics.

II. A CLASSICAL VIEWPOINT

The roots of useful ideas often have multiple branches that penetrate different disciplinary horizons. The notion of distance decay is one that can be traced to Isaac Newton; in geographical interpretations, it later rested in the hands of Tobler [1961, 1992], Warntz [1965], and others. Currently, note that Griffith's discussion of spatial weights matrices, and Griffith's and Can's consideration of land use patterns around the central city, refer to distance decay in their chapters in this handbook.

Another powerful notion---that of a transformation---has served as the backbone of twentieth century mathematics. It has altered disciplinary focus from the study of individual mathematical systems to the study of relations between mathematical systems. D'Arcy Thompson explored this transformational approach to biology [1917]; Tobler has employed transformations and some of Thompson's work in various aspects of cartographic analysis [1961, 1992] as has Bookstein in the measurement of biological shape [1978]. Clarke (1995) presents a transformational view of cartography in his text on analytical and computer cartography. Thirty years ago Michael Dacey and others contributed to the development of spatial statistics in highly original ways. Dacey used the idea of a dimensional transformation to permit evaluations of the spatial association of point and area phenomena. Indeed, one of the figures in Vasiliev's chapter calls to mind some of Dacey's earlier efforts [1964]. Today, note that Wong employs the transformation of spatial aggregation to move across a hierarchy of scale changes; Feng sees the fertility transition as one transformation in the broader transition theory framework (for a complementary current perspective on transition theory see Drake in Ness, Drake, and Brechin, 1993); Long uses the Jacobian as a normalizing factor to transform a correlated mathematical space into an uncorrelated one; Brown considers the role of the transformation in a Geographic Information Systems (GIS) context; and, Li adopts a transformational viewpoint in both content and written format in his essay on parallel processing. The concept of transformation is critical in these, and in other, chapters of this handbook.

Ideas capable of widespread application have deep and far-flung roots. It is the aim of this handbook to show the clearly visible, tangible branching structure of the spatial statistics "tree;" we hint here at its roots in the hope that the diligent reader will be encouraged to dig into the rich and fertile minds from which this tree was raised.

REFERENCES

Bookstein, Fred L. *The Measurement of Biological Shape and Shape Change.* Berlin, Springer-Verlag, 1978.

Clarke, Keith C. *Analytical and Computer Cartography,* 2nd edition. Prentice-Hall, Englewood Cliffs NJ, 1995.

Dacey, Michael F. Imperfections in the uniform plane. *Michigan Inter-University Community of Mathematical Geographers,* John D. Nystuen, Editor, May, 1964. Reprinted, *Solstice: An Electronic Journal of Geography and Mathematics,* Summer, 1994.

Ness, Gayl D.; Drake, William D.; Brechin, Steven R., eds. *Population-Environment Dynamics: Ideas and Observations.* Ann Arbor, University of Michigan Press, 1993.

Nystuen, Jeffrey A.; McGlothin, Charles C.; and, Cook, Michael S. The underwater sound generated by heavy rainfall. *Journal of the Acoustical Society of America,* 93 (6), 3169-3177, 1993.

Sadowski, Frank G. and Covington, Stephen J. *Processing and Analysis of Commercial Satellite Image Data of the Nuclear Accident near Chernobyl, U.S.S.R.* U. S. Geological Survey, Bulletin 1785. U.S. Government Printing Office, Washington, D. C., 1987.

Thompson, D'Arcy W. *On Growth and Form.* Cambridge University Press, 1917.

Tobler, Waldo R. *Map Transformations of Geographic Space.* Ph.D. Thesis, University of Washington, 1961.

Tobler, Waldo R. Preliminary representation of world population by spherical harmonics. *Proceedings of the National Academy of Sciences of the United States of America,* 89:6262, 4 Jul. 15 1992

Warntz, William. *Macrogeography and Income Fronts.* Philadelphia, Regional Science Research Institute, Monograph #3, 1965.

Preface

Anyone who already uses statistics and maps will benefit from using spatial statistics.

This handbook is a reference work that illustrates the differences between classical statistics, and spatial statistics--those techniques which account, in some manner, for geographical position. It does so at both the abstract level and the real world level.

Useful features in this handbook include:

- 1. Comparisons of classical and spatial statistical techniques;
- 2. Rules-of-thumb capturing the essence of selected techniques;
- 3. Real-world data used to illustrate abstract concepts;
- 4. Real-world locales of timely current nature;
- 5. Cutting-edge topics in spatial statistics;
- 6. Spatial index that maps relative locations of terms by chapter;
- 7. Reference lists grouped by chapter and for the book as a whole.

Editors and authors alike have worked hard to create a useful and uniform document; however, as is always the case, there is no perfect document. If you find something that you wish to share with us for the next edition--new material, corrections, or whatever--please communicate with:

William C. Arlinghaus
ATTN: CRC Handbooks
Department of Mathematics and Computer Science
Lawrence Technological University
21000 West Ten Mile Road
Southfield, MI 48075
arlinghaus@LTU.edu

Thank you in advance; we have tried to be careful and to contribute something that is useful and different; in the end, despite all the care of the many who have generously offered time, effort, and advice, the blame for errors, omissions, or poor judgment must rest with the Editor-In-Chief, alone.

Sandra Lach Arlinghaus, Ann Arbor, MI June 7, 1995.

Affiliations of Editors and Authors

Editor in Chief:
Sandra Lach Arlinghaus, Ph.D. Geography.
Founding Director, Institute of Mathematical Geography (independent),
2790 Briarcliff, Ann Arbor, MI 48105; and,
Adjunct Professor of Mathematical Geography and Population-
Environment Dynamics
School of Natural Resources and Environment
The University of Michigan
Ann Arbor, MI 48109

Associate Editors:
William C. Arlinghaus, Ph.D. Mathematics.
Professor and Chairman,
Department of Mathematics and Computer Science,
Lawrence Technological University
21000 W. Ten Mile Rd.
Southfield, MI 48075

William D. Drake, Ph.D. Operations Research.
Professor,
School of Natural Resources and Environment,
School of Public Health, and,
College of Architecture and Urban Planning,
The University of Michigan
Ann Arbor, MI 48109

Daniel A. Griffith, Ph.D. Geography.
Professor and Chairman,
Department of Geography and Program in Interdisciplinary Statistics
144 Eggers Hall
Syracuse University
Syracuse, New York 13244-1160

John D. Nystuen, Ph.D. Geography.
Professor,
Geography and Urban Planning,
College of Architecture and Urban Planning,
The University of Michigan
Ann Arbor, MI 48109

Chapter contributors other than editors:

Dr. Daniel G. Brown
Department of Geography
Michigan State University
East Lansing, Michigan 48824

Dr. Ayse Can
Fannie Mae
Office of Housing Research
3900 Wisconsin Avenue, NW
Washington, D.C. 20016-2819

Mr. Michael Feng
Department of Geography
144 Eggers Hall
Syracuse University
Syracuse, New York 13244-1160

Dr. Bin Li
Department of Geography
P.O. Box 8067
University of Miami
Coral Gables, Florida 33124-2060

Dr. Daniel Long
Northern Agricultural Research Center
Star Route 36, Box 43
Montana State University
Havre, Montana 59501

Dr. W. Scott Overton
Department of Statistics
44 Kidder Hall
Oregon State University
Corvallis, Oregon 97331-4606

Dr. Stephen V. Stehman
Faculty of Forestry
322 Bray Hall
SUNY/CES&F
Syracuse, New York 13210

Ms. Irina Ren Vasiliev
Department of Geography
State University College
Geneseo, New York 14454

Dr. David Wong
Department of Geography and Earth Systems Science
George Mason University
Fairfax, Virginia 22030

Acknowledgments

One of the nice parts of writing a book with many co-editors and co-authors is having the opportunity to get to know an academically diverse set of interesting individuals. The editors extend their heartiest, and primary, thanks to all of the authors and hope one day to have the good fortune to meet each of them in person. They have all been enjoyable to work with and hopefully that constructive spirit of interaction is evident in the fruit it has borne. The editors in Ann Arbor also wish to extend particular congratulations to Daniel A. Griffith; Dan has worked endlessly and painstakingly to make this work come to life!

In addition to the authors, we thank, bearing in mind the unfortunate possibility that we might inadvertently have omitted someone, the following individuals and organizations.

Community Systems Foundation (CSF) of Ann Arbor has generously provided access to computing resources for this project. Their continuing support of handbook projects has been indispensable. William E. Arlinghaus provided complimentary professional copyright procurement assistance when needed. His efforts in these matters is greatly appreciated. Syracuse University generously extended copyright permission to include segments of computer programs written by Griffith.

Wayne Yuhasz at CRC Press was our original editor. He was delightful to work with. Nora Konopka, also of CRC Press, has served tirelessly as a friendly liaison during the transition in change in editorship at CRC. How fortunate we are, now, to have Robert Stern as the new editor to help see us through the final stages of manuscript preparation. His assistance with, and concern for, a variety of matters is greatly appreciated. We also salute the many others at CRC who have been so helpful.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION: THE NEED FOR SPATIAL STATISTICS

Daniel A. Griffith

Geographic information and analysis (GIA) is a critical, emerging scientific discipline. [Cook *et al.*, 1994] The establishment in the late 1980s of the National Center for Geographic Information and Analysis (NCGIA), with funds from the National Science Foundation, attests to its importance. Data that are tied to position on the Earth's surface, that are spatial or geo-referenced data, often serve as the empirical backbone of much of the research that is presently done in this general context. The statistical analysis of spatial data forms the subject matter of "spatial statistics." Indeed, in writing about his Cornell Theory Center supercomputer project, Durrett notes [1994, p. 4] that

> *[f]or a half century, the literature ... has been dominated by models in which spatial location is ignored and each individual [site] is assumed to interact equally with all the others. Such models provide an acceptable approximation in many contexts, but there is a growing list of examples of phenomena that must be treated by models that are spatially explicit ... .*

Others echo this need for models that are spatially explicit: the *Chorley Report* [1987] released in Great Britain; reports of the National Research Council in the United States entitled Spatial Data Needs and Renewing U.S. Mathematics [1990a, b]; Warnecke's survey of state activities [1990, 1991]; and, International Business Machine's (IBM's) feature article in 1991.

I. COMPONENTS OF GEOGRAPHIC INFORMATION AND ANALYSIS

The academic subject matter of Geographic Information and Analysis comprises three principal components: geographic information systems (GISs), spatial statistics, and classical spatial analysis. From a broad perspective, geographic information systems are a form of applied computer science, spatial statistics is a form of specialized applied multivariate statistics, and classical spatial analysis is a form of quantitative geography.

GISs constitute a powerful new technology that can address many information needs of decision makers working with geographically

0-8493-0132-7/95/$0.00+$.50

(geo-) referenced data--data that are tagged, or identified, by locational coordinates. Often this tagging is for coordinates on the Earth's surface; today many tags are created with the aid of the satellite-based technology of the global positioning system (GPS). GISs are unique combinations of computer hardware and software--including high-resolution graphic displays, large-capacity electronic storage devices, efficient strategies for data organization, high-volume communication channels, specialized algorithms for data integration and reliability analysis, and specialized query computer languages. These components, together with massive amounts of highly complex geo-referenced data, are organized efficiently (through a sequence of electronic interfaces) to store, inventory, manage, search, manipulate, display (instantaneously), and analyze information contained in a geo-referenced database. The goal is to combine tabular attribute data with computerized maps in an enlightening way, achieving this goal by having a large storage capacity, a rapid response time, and a wide repertoire of analytical functions. Together, these support a dramatic mode of scientific visualization.

Generally speaking, spatial statistics is concerned with the statistical analysis of geo-referenced data. In 1991 a National Research Council report characterized this subject area as (p. vii)

> *one of the most rapidly growing areas of statistics, rife with fascinating research opportunities.*

Yet, despite these opportunities, many statisticians remain unaware of them and most students in the United States are never exposed to course work in spatial statistics--this handbook attempts to bridge that gap. *Spatial statistics differs from classical statistics in that the observations analyzed are not independent; this single assumption violation is the crux of the difference.* Cressie [1991, p. 3] characterizes this problem area as follows:

> *Independence is a very convenient assumption that makes much of mathematical-statistical theory tractable. However, models that involve statistical dependence are often more realistic.*

Moreover, observations are correlated strictly due to their relative locational positions (referred to as spatial autocorrelation), resulting in spill-over of information from one location to another (locational information). This spill-over causes redundant information to be present in data values. The redundancy increases as the degree of locational dependence increases. This duplication of information produces complications in the statistical analysis of geo-referenced data that remain dormant in the statistical analysis of traditional data composed of independent observations. That is, invoking an assumption of independent observations suppresses potential data complexities. The net result is that classical statistics applied to geo-referenced data fail to capture locational information, raising questions of estimator sufficiency, bias, efficiency, and consistency. These four cardinal statistical properties might also be coupled with two others: robustness and minimum variance. Geo-referenced data are highly complex, with spatial dependence introducing further complications. Examples of studies that demonstrate changes in statistical inferences when a traditional ordinary least squares (OLS) regression model estimation is replaced with a spatial statistical one include explaining the Huk rebellion [Cliff and Ord, 1981, p. 237], predicting county wage rates [Anselin, 1988, pp. 191, 193], and estimating mean density of coffee production [Griffith, 1989].

Classical spatial analysis has been treated conceptually for about a century, and algebraically for several decades. It has played a central role in the quantitative scientific tradition in geography. Spatial analysis involves spatial operations research (such as minimum route selection), logical overlaying (identifying areal units possessing joint categorical attributes), triangulated irregular networking (TIN; which in a sense forms the basis of spatial statistics), and buffering (distance bands around points or lines), among others. Most of these procedures currently are available as functions in GIS tool kits; they automate what once were tedious manual tasks.

II. BACKGROUND: THE IMPORTANCE OF LOCATIONAL INFORMATION

Scholarly awareness of complications attributable to locational information latent in spatial data, especially in terms of their impact on the validity of traditional statistical analyses, has emerged recently among scientists, catapulting spatial statistics into the forefront of much data analysis discussion. In fact, the analysis of spatial data has become a major preoccupation of numerous statisticians only rather recently. Accordingly, increasing attention has begun to focus on the general field of spatial and geo-statistics (and spatial econometrics). For instance, the announced goals in the solicitation of proposals for the NCGIA [NSF, 1987] included the objective of promoting advances in spatial statistics within the context of GISs. And, the Board of Mathematical Sciences of the National Research Council [1990b] has targeted spatial statistics as one of twenty-seven topics of national concern in mathematics (its rank is 17). Similar evidence has been made available by the British scientific community, particularly through that country's Regional Research Laboratories initiative. (For example, as a cooperative effort, the Department of Mathematics at Lancaster University and the North West Regional Research Laboratory in Great Britain initiated a project to integrate statistical and GIS software, while the U.K. Economic and Social Research Council (ESRC) funded an "experts" workshop, held at the University of Sheffield in March of 1991, on this same theme. [see Goodchild, Haining, and Wise, 1992]

In 1989 a symposium entitled "Spatial statistics: past, present, and future" was hosted by the Department of Geography, Syracuse University. Reported findings of this symposium [Griffith, 1990] include

- (1) there is a need for a MINITAB or SAS for spatial statistics [Ripley, p. 56; Haining, p. 101; Doreian, p. 105; Wartenberg, p. 153; Upton, p. 158];
- (2) there is a need for many more additional relevant empirical applications of spatial statistical techniques [Martin, p. 27; Richardson, p. 130; Upton, p. 354; Wartenberg, p. 393]; and,
- (3) as the issue of computational intensity subsides, and GIS software becomes increasingly user-friendly, more ubiquitously available, and a source for implementing spatial statistical techniques, the danger of malpractice by the non-specialist practitioner grows [Anselin, p. 73; Martin, p. 124].

Similar sentiments are echoed in Cressie [1991; pp. 657, 699], while a review of the literature demonstrates that little spatial statistics and spatial econometrics technology has been adopted in scientific research [see Anselin and Griffith, 1987; Anselin and Hudak, 1992], highlighting the existence of a dissemination problem. In response to this first point, Griffith [1989, 1993c] has developed both MINITAB and SAS macros for undertaking spatial statistical analyses, whereas Anselin [1992] has developed SPACESTAT for executing spatial econometrics.

Therefore, although a scholarly awareness of spatial statistics currently exists in various fields (especially those in the geosciences), important research needs have been defined by leading researchers in the field, and although both curricular developments and dissemination endeavors are underway, insufficient synthesis and consulting materials are available. The literature is piecemeal, specialized, and diverse. Its content consists of books that tend to be introductory [Goodchild, 1986; Griffith, 1987; Odland, 1988] or advanced [Ripley, 1988]; little exists at the intermediate level. One book covers pattern models [Ahuja and Schachter, 1983], another surface partitionings [Okabe, Boots, and Sugihara, 1992]. Some books treat mostly point pattern analysis [Cliff and Ord, 1981], others geo-statistics [Cressie, 1991], spatial autoregressive models [Griffith, 1988], or spatial econometrics [Paelinck and Klaasen, 1979; Anselin, 1988]. Some books are extremely theoretical [Bartlett, 1975; Matérn, 1986], while others are very applications-oriented [Upton and Fingleton, 1985; Isaaks and Srivastave, 1989; Haining, 1990]. This handbook fills the gap identified here, and in so doing explicitly addresses the two points of furnishing additional relevant empirical applications of spatial statistical techniques, and providing guidance to help non-specialist practitioners avoid improper spatial statistical practice. The style, nature, and scope of this volume is designed to pique the curiosity of graduate students and spatial scientists alike.

III. BACKGROUND: STATISTICAL ESTIMATOR PROPERTIES

Complications mentioned in the preceding discussion can be referenced to the statistical properties of estimator sufficiency, unbiasedness, efficiency, and consistency. An estimator is sufficient if, when reducing the sample data to its corresponding summary statistic(s), it does not foster a loss of information pertinent to the population to which an inference is to be drawn. Exactly all of the information relevant to a population that is contained in a sample is condensed into a sufficient statistic. This definition means that all the knowledge about a parameter than can be gleaned from both the individual sample values and their ordering *must* be gained from the value of the estimator alone. Griffith [1988] uses Neyman's factorization theorem, and the theorem on completeness for the exponential family [Lindgren, 1976] to show that in the presence of locational information the mean and standard deviation are not sufficient statistics. Rather, a four-dimensional statistic is needed that incorporates both the geographic arrangement of observations and the nature and degree of their spatial dependence. *These are the estimators found in spatial statistics, the ones that should be employed in the statistical analysis of geo-referenced data.*

An estimator is unbiased if the mean of the sampling distribution generated by it equals the parameter that it is supposed to estimate. On average, the estimator value is equivalent to its corresponding population parameter value. Many common statistics involving simple linear combinations of geo-referenced data, such as the arithmetic mean and regression coefficients, are unbiased. But ones involving more complicated arithmetic operations, such as the variance and correlation coefficients, tend not to be. For example, Cordy and Griffith [1993] discuss how, in some cases, the OLS regression coefficient estimators provide a reasonable alternative to their spatial statistical counterparts, while the usual variance estimators are severely biased when regression errors are spatially autocorrelated. The main problem with using OLS in the presence of spatial autocorrelation is that the usual standard error estimator tends to underestimate the true standard error. This result indicates that the geographic arrangement of observations and the nature and degree of prevailing spatial dependence affects levels of statistical significance, and hence the precision of any single set of sample estimates, as well as prediction error. This consequence raises questions about OLS results reported in many existing studies involving the analysis of geo-referenced data.

One of two candidate estimators is relatively efficient if its sampling distribution has the smaller variance of the two, both of which are unbiased, making it the more reliable measure. An unbiased estimator that attains the lower bound established in the Cramér-Rao inequality is an efficient estimator; efficiency, then, may be defined as the ratio of the Cramér-Rao lower bound to the actual variance of an unbiased estimator (in general this lower bound is not attainable). This statistical trait coupled with the Cramér-Rao inequality implies that the more efficient an estimator is, the more information it provides about a target population. In the limit, then, a population constant would have zero variance, and its companion estimator would yield perfect information about this parameter. Cordy and Griffith [1993] found that, in general, the need to take spatial autocorrelation into account in variance estimation for geo-referenced data tends to negate advantages due to computational simplicity affiliated with the use of OLS estimation. For the common case where the spatial autocorrelation parameter needs to be estimated, some of the potential gains in efficiency by employing spatial statistical estimators are not realized. Sometimes if either this spatial autocorrelation parameter is negligible or the sample size is small, traditional statistical estimators can be more efficient that their spatial statistical counterparts (likely due to edge effects). Frequently, however, the spatial autocorrelation parameter value is positive, and moderate in magnitude (roughly indicating juxtaposed correlations falling into the range of 0.3-0.5).

An estimator is consistent, which is a large sample or limiting property, if for large n it takes on values that are very close to the value of its corresponding parameter. This will occur only if both the variance of and the bias of an estimator tend to zero as n tends to infinity. It suggests that when the sample size is sufficiently large, there is near certainty that the error made with a consistent estimator will be less than any small preassigned positive constant. The larger the sample size the better the inference one could expect to make.

Moreover, if an estimator is unbiased and its variance goes to zero as n increases to infinity, then it is a consistent estimator of its respective parameter. Consistency is a neglected topic with regard to spatial statistics. Ord [1975] broached the topic, noting that least squares and one popular spatial statistical estimator are inconsistent, and that at least certain maximum likelihood estimators are consistent. Mardia and Marshall [1984] stress that when geo-referenced data involve a single map being observed, which often is the case, consistency of estimators is not obvious. They maintain that only weak consistency can be established in general. Traditional estimators tend to display statistical consistency when the size of a geographic region is increased without limit, although their performance deteriorates as the degree of spatial dependence in, and areal unit articulation of, a geo-referenced data set increase. In contrast, when more and more geo-referenced samples are drawn on a continuous variable in a region of fixed size (i.e., the sampling density increases), many of the commonly used estimators tend to become inconsistent. This more realistic latter situation is pertinent to investigations of almost any geo-referenced data set, as well as changing geographic scales (resolution) of analysis, where the nature of large samples causes the average distance between sampled points to decrease, and the degree of observed spatial dependence to increase. One complicating factor is that a lower limit exists for impacts of scale change due to the discreteness of punctiform geographic distributions, which probably prevents perfect positive spatial autocorrelation from being attainable for many geographically distributed phenomenon.

IV. OVERVIEW OF THE TOPICS

The principal objective of this *Handbook* is to illustrate how to implement spatial statistical techniques, and to show clearly basic differences between spatial and conventional statistical analyses using real-world data sets. Its contents are organized into ten chapters, outlined in the Table of Contents, illustrating the differences between spatial and classical statistics. In the discussion that follows, a view of the topics of the chapters is taken from a different vantage point to enhance overall perspective. (The reference database is STARS from the World Bank; however, a variety of additional databases are employed, too.)

The critical research frontier issues in spatial statistics, particularly with regard to implementation of its techniques, are represented by the ten topics that compose these chapters. Moreover, interlaced with methodological recommendations, cautions, and guidance, and coupled with exemplary illustrations, are treatments of pressing contemporary spatial statistical issues.

A. DEFINING SPATIAL WEIGHTS MATRICES

Stetzer [1982] was one of the first researchers to systematically address the issue of how a spatial weights matrix should be specified. Often a spatial scientist needs to know the answer to questions asking

- (1) which spatial autoregressive model and accompanying spatial weights matrix performs best, irrespective of sample size and mis-specification;
- (2) about the consequences of over-specifying the weights matrix (i.e., including false spatial dependence adjacencies); and,
- (3) about the consequences of under-specifying the weights matrix (i.e., omitting true spatial dependence adjacencies).

Griffith [1993b], and Griffith and Csillag [1993] indicate that a comprehensive treatment of the spatial weights matrix specification is warranted as part of any quantitative geographic research project. This treatment should deal with edge effects, questions of internal partitioning of a geographic landscape, regional shape, and the nature and degree of the prevailing spatial dependence.

B. IMPLEMENTING SPATIAL STATISTICS WITH SUPERCOMPUTERS

Massively large geo-referenced data sets are becoming the norm, rather than the exception, in today's world of high resolution imagery and high performance computing. The volume of data in a standard GIS database, accompanied by the numerically intensive requirements of many spatial statistics calculations, suggests that an integration of sophisticated spatial analysis and GIS requires the current capabilities of supercomputers in order to circumvent computational bottlenecks. Implementation of spatial statistical models in these situations frequently requires that supercomputing play a central role. Li [1993] has found that two-dimensional geographic distributions map nicely onto the connection machine, with its substantial number of CPUs. In addition, Griffith [1990], and Griffith and Sone [1992, 1993], have found that spatial statistical procedures can be easily implemented with vector supercomputers (the Cray 2 and IBM 3090). The sum total of these outcomes is that numerically intensive spatial statistical calculations can be performed in a very reasonable time, with the time increase accompanying increasing sample size remaining quiet manageable. This finding has particular relevance to the newest novelty to appear in the desktop computing environment, namely workstations like the IBM RISC/System 6000 machine.

C. SPATIAL STATISTICS FOR REGULARLY SPACED DATA

Classical experimental design is based on the three concepts of randomization, blocking, and replication. Randomization strives to neutralize the effects of spatial correlation and renders valid tests for the hypothesis of equal treatment effects. Grondona and Cressie [1991] have shown that resorting to a spatial statistical model can yield more efficient estimators of the treatment contrasts than classical statistical approaches. In so doing, such a spatial statistical analysis gives a more complete understanding of the phenomena influencing crop yield. For example, by focusing on the influence of one variable (using partial derivatives) on the effects of spatial autocorrelation on variance estimates, detection of treatment differences is enhanced.

Meanwhile, Griffith [1993c] has found that the special nature of the uniform structuring of a regular square tessellation surface partitioning allows measures of spatial autocorrelation to be computed, and spatial statistical model parameters to be estimated, without having to explicitly construct a geographic weights matrix. The regular two-dimensional arrangement of the geo-referenced data allows standard time series LAG functions to be used in order to identify pixels lying immediately to the east and south of a given areal unit. Invoking this function necessitates the addition of a missing values column to the map, since the extreme eastern pixels have no values immediately to their east. By including a sequential numbering scheme the data can then be reordered by sorting on this numbering variable. The consequence is that a LAG function can then be used to identify values within the boundary that are adjacent to each pixel. Regular square tessellation data require far simpler computer code for undertaking spatial statistical analyses. This simplicity is less demanding of computer memory, allowing substantially larger problems to be analyzed.

D. AGGREGATION EFFECTS IN GEO-REFERENCED DATA

Geographic scale can vary between something that is quite coarse, a category heading that many surface partitions fall under, to something that is quite fine (high resolution). Customarily (as in conventional univariate statistics), some aggregation of geo-referenced data is necessary in order to unmask pattern from detail, although too much aggregation also obscures pattern; aggregation into a single large areal unit blends regional distinctions, camouflaging them into oblivion. One theme in the quantitative geographic literature acknowledging complications attributable to changes in scale is known as the modifiable areal unit problem (MAUP; see Amrhein, 1994; and, Chapter 5 in this book).

A spatial scientist needs to properly analyze the variablility of data over space at an appropriate scale. Thus, knowing which linear statistics are, and which are not, sensitive to variations in geographic scale and zoning (surface partitioning) systems is exceedingly important. [See, for example, Green, 1993] This variation arises from three major sources: natural variability (stochastic error), measurement error, and sampling error. Each source of variability adds uncertainty to analysis results and confounds understanding of the nature and degree of relationships among geographic distributions. This uncertainty can be complicated by geographic aggregation, which propagates and convolutes these various sources of error. In the end, a spatial scientist routinely seeks methods to disentangle aggregation effects that constrain the drawing of inferences based upon geo-referenced data from one geographic scale to others.

E. SPATIAL SAMPLING

GISs have revolutionized geo-referenced data handling, in general, and the visualization of spatial data, in particular. But for this technological advance to be informative, data entered into a GIS database must be collected in a meaningful fashion. One way to ensure meaningfulness is to implement a proper sampling design for collecting the geo-referenced data. Another is to comply with a proper data analysis. And a third is to extract appropriate interpretations from the data analysis.

Published theoretical and applied works concerning spatial sampling designs span more than fifty years and embrace many disciplines. Initial applications of statistical sampling theory to problems involving spatially autocorrelated variables appeared in the late 1930s. Meanwhile, discussions addressing the relative efficiency of random sampling versus systematic sampling of data distributed in two dimensions appeared nearly fifty years ago. The general conclusion reached was that systematic sampling is superior to random sampling of geo-referenced data, with the optimal sampling network design being a superimposed equilateral triangular grid. To date, almost without exception, research on sampling network design has focused on improved estimation of averages for predefined regions; the goal has been to obtain a minimum variance estimate of this mean from a fixed number of sample points.

The problem of spatial sampling designs has been revisited by Stehman and Overton (see, for example, 1989; and, Chapter 3 in this book). To begin, they emphasize that systematic sampling has many advantages over unrestricted random sampling. One practical advantage is ease of implementation. One theoretical advantage is increased precision of estimators. One geo-referenced data context advantage is the ability to furnish information on spatial or temporal patterns in a target population. In fact, geo-referenced data properties of estimators based upon a systematic sample are determined by the natural geographic ordering of a response variable. But, although systematic sampling is widely accepted as a practical sampling design, obtaining an unbiased estimation of variance with it has continued to be problematic. In practice, though, a standard recommendation is to treat the systematic sample as an unrestricted random sample, assuming that the response variable occurs randomly along the underlying natural ordering. In this context the standard formula unfortunately overestimates the true variance in most circumstances. Evaluation results obtained by Overton and Stehman [1990], using the criteria of precision and suitability of variance estimation, argues for use of the triangular network tessellation-stratified sampling design.

F. SPATIAL STATISTICS AND GIS

Presently there is both evidence and prevailing expert opinion indicating that spatial statistical techniques need to be converted into GIS functions. [See, for example, Griffith, 1993a] Access to standard GIS functions for managing, transforming, and displaying spatial input data and visualizing model output and residuals enhances a spatial statistical analysis. Candidate techniques for inclusion in a GIS toolbox include indices of spatial autocorrelation and selected spatial autoregressive models, which, for instance, can be used to enhance satellite digital data classification procedures. [Brown and Walsh, 1993] These procedures should be augmented with an elementary multiple linear regression function, a spatial weights generator, and a Moran Coefficient function to test for spatial autocorrelation in regression residuals, spatial correlogram and semivariogram functions. The descriptive functions can be used to examine data for spatial patterns, which may suggest causal processes or, in some cases, reveal systematic errors in the data. [Brown and Bara, 1994] An extremely useful geo-statistical function is the semivariogram, which can serve as a tool for suggesting optimum cell sizes in process models linked with raster GIS databases [Brown *et al.*, 1993] and for the correction of systematic errors in digital elevation models. [Brown and Bara, 1993] Moreover, spatial statistical procedures constitute an essential element of a complete battery of functions that should be available to the quantitative spatial scientist. When they are not embedded in GIS software, the scientist may well have no other recourse than to export geo-referenced data from the GIS package into a statistical or custom-designed software package.

G. VISUALIZATION OF SPATIAL DEPENDENCE

One of the hallmarks of GIS software is its ability to support and foster scientific visualization and computer mapping (cartographic representation, and display and analysis of geo-referenced data). Dealing with the relationship between spatial autocorrelation and scientific visualization dates back, at least, to Olsen's [1975] seminal piece. One important element in the overall look of a choropleth map, for instance, is the relationship between neighboring values as they appear on the map. Hence maps and spatial statistics should go hand-in-hand to help a spatial scientist understand how much of an effect neighboring attribute values have on each other. By so doing, the scientist will begin to accumulate the knowledge necessary for eventually acquiring an intuitive understanding of the effects of locational information simply from visual inspection.

H. EMPIRICAL APPLICATIONS

Applied work frequently is guided by example. The case dealing with dependence in geo-referenced forestry data endeavors to provide guidance for the use of spatial statistics in the analysis of forestry data, much of which is geo-referenced, and to demonstrate the value and utility of spatial analysis for natural resources problems (Chapter 7). Spatial scientists need to gain a better understanding of the geographic and attribute complexities latent in forestry data.

The example dealing with spatial dependence in geo-referenced urban data seeks to examine the geographic distribution of population in an urban area, which is arranged in such a way that it displays conspicuous patterns of spatial autocorrelation (Chapter 9). Anselin and Can [1986] already have investigated a number of different specifications of the negative exponential component of population density gradients, deciding upon the simultaneous autoregressive errors model to account for the presence of spatial autocorrelation.

Finally, spatial dependence in geo-referenced population data ventures to explore reaction and interaction processes in demographic fertility transition with spatial statistical procedures. Feng's study in Chapter 8 deals with these concerns in the context of China's current population policy.

V. SUMMARY

In summary, geo-referenced data are highly complex with spatial dependence introducing further complications. These complications are similar to those found in time series analysis. They are exacerbated by the multi-directional, two-dimensional nature of spatial dependence (time series entails dependencies that are unidirectional along a single dimension), and the far more complex geometric infrastructure involved (classical time series entails a regular linear geometry). The cost of such oversights can be considerable. Incorporation of such features can be achieved by following the guidance and prescriptions offered in this book.

REFERENCES

Ahuja, N., and Schachter, B. *Pattern Models.* Wiley, New York,1983.

Amrhein, C. Searching for the elusive aggregation effect: evidence from statistical simulations. *Environment and Planning A,* 26, 1994.

Anselin, L. *Spatial Econometrics.* Kluwer, Dordrecht, 1988.

Anselin, L. SPACESTAT TUTORIAL: A workbook for using SpaceStat in the analysis of spatial data. *Technical Software Series S-92-1*, NCGIA, Santa Barbara CA, 1992.

Anselin, L., and Can, A. Model comparison and model validation issues in empirical work on urban density functions. *Geographical Analysis,* 18, 179-197, 1986.

Anselin, L., and Griffith, D. Do spatial effects really matter in regression analysis? *Papers* of the Regional Science Association, 65, 11-34, 1987.

Anselin, L., and Hudak. Spatial econometrics in practice: a review of software options. *Regional Science and Urban Economics,* 22, 509-536, 1992.

Bartlett, M. *Statistical Analysis of Spatial Pattern.* Chapman and Hall, London, 1975.

Brown, D., Bian, L., and Walsh, S. Response of a distributed watershed erosion model to variations in input data aggregation levels. *Computers and Geosciences,* 19(4), 499- 509, 1993.

Brown, D. and Walsh, S. Spatial autocorrelation in remotely sensed and GIS data. *Proceedings* of the ACSM/ASPRS Annual Convention, New Orleans, LA, Vol. 3, 13-39, 1993.

Brown, D. and Bara, T. Recognition and reduction of systematic error in elevation and derivative surfaces from 7 1/2-minute DEMs. *Photogrammetric Engineering and Remote Sensing,* 60, 189-194, 1994.

Chorley, R. [chairman]. *Handling Geographic Information.* Her Majesty's Stationary Office, Report to the Select Committee on GIS, London, 1987.

Cliff, A., and Ord, K. *Spatial Processes.* Pion, London, 1981.

Cordy, C., and Griffith, D. Efficiency of least squares estimators in the presence of spatial autocorrelation. *Communications in Statistics--Simulation and Computation,* 22, 1161-1179, 1993.

Cressie, N. *Statistics for Spatial Data.* Wiley, New York, 1991.

Durrett, R. Stochastic spatial models, *Forefronts* (newsletter of the Cornell Theory Center), 9 (#4, Spring), 4-6, 1994.

Goodchild, M. *Spatial Autocorrelation.* CATMOG, Norwich, England, 1986.

Goodchild, M., Haining, R., and Wise, S. Integrating GIS and spatial data analysis: problems and possibilities. *International Journal of Geographical Information Systems*, 6, 407-423, 1992.

Green, M. Ecological fallacies and the modifiable areal unit problem. *Research Report No. 27,* North West Regional Research Laboratory, Lancaster University, UK, 1993.

Griffith, D. *Spatial Autocorrelation.* Association of American Geographers, Washington, D. C., 1987.

Griffith, D. *Advanced Spatial Statistics.* Kluwer, Dordrecht, 1988.

Griffith, D. Spatial regression Analysis on the PC: spatial statistics using MINITAB. *Discussion Paper #1*, Institute of Mathematical Geography, Ann Arbor, MI, 1989.

Griffith, D. (ed.). *Spatial Statistics: Past, Present, and Future.* Institute of Mathematical Geography, Ann Arbor, MI, 1990.

Griffith, D. Which spatial statistics techniques should be converted to GIS functions? in *Geographic Information Systems, Spatial Modelling and Policy Evaluation*, edited by M. Fischer and P. Nijkamp. Springer-Verlag, 103-114, Berlin, 1993a.

Griffith, D. Advanced spatial statistics for analysing and visualizing geo-referenced data. *International Journal of Geographical Information Systems*, 7, 107-123, 1993b.

Griffith, D. *Spatial Regression Analysis on the PC: Spatial Statistics Using SAS.* Association of American Geographers, Washington, D.C., 1993c.

Griffith, D., and Csillag, F. Exploring relationships between semi-variogram and spatial autoregressive models. *Papers in Regional Science*, 72, 283-295, 1993.

Griffith, D., and Sone, A. Trade-offs associated with computational simplifications for estimating spatial statistical models. *Working Paper*, l'Institut de Mathématiques Economiques, Université de Bourgogne, Dijon, France (with French Resumé), 1992.

Griffith, D., and Sone, A. Some trade-offs associated with computational simplifications for estimating spatial statistical/econometric models: preliminary results. *Discussion Paper* No. 103, Department of Geography, Syracuse University, 1993.

Grondona, M., and Cressie, N. Using spatial considerations in the analysis of experiments. *Technometrics*, 33, 381-392, 1991.

Haining, R. *Spatial Data Analysis in the Social and Environmental Sciences.* Cambridge University Press, Cambridge, England, 1990.

IBM Exploring new worlds with GIS. *Directions*, Summer/Fall, 12-19, 1991.

Isaaks, E. and Srivastava, R. *An Introduction to Applied Geostatistics.* Oxford University Press, Oxford, England, 1989.

Lindgren, B. *Statistical Theory*, 3rd ed. Macmillan, New York, 1976.

Mardia, K., and Marshall, R. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71, 135-146, 1984.

Matérn, B. *Spatial Variation*, 2nd ed. Springer-Verlag, Berlin, 1986.

National Research Council (Mapping Science Committee; Commission on Physical Sciences, Mathematics, and Resources). *Spatial Data Needs: The Future of the National Mapping Program*, National Academy Press, Washington, D.C., 1990a.

National Research Council (Board on Mathematical Sciences). *Renewing U.S. Mathematics: A Plan for the 1990s.* National Academy Press, Washington, D.C., 1990b.

National Research Council (Panel on Spatial Statistics and Image Processing). *Spatial Statistics and Digital Image Analysis.* National Academy Press, Washington, D.C., 1991.

National Science Foundation. *Solicitation: National Center for Geographic Information and Analysis.* Biological, Behavioral, and Social Sciences Directorate, Washington, D.C., 1987.

Odland, J. *Spatial Autocorrelation.* Sage, Beverly Hills, CA, 1988.
Okabe, A., Boots, B., and Sugihara, K. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams.* Wiley, New York, 1992.
Olsen, J. Autocorrelation and visual map complexity. *Annals*, Association of American Geographers, 65, 189-204, 1975.
Ord, K. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70, 120-126, 1975.
Overton S., and Stehman, S. Statistical properties of designs for sampling continuous functions in two dimensions using a triangular grid. *Technical Report* No. 143, Department of Statistics, Oregon State University, 1990.
Paelinck, J. and Klaassen, L. *Spatial Econometrics.* Saxon House, Farnborough, England, 1979.
Ripley, B. *Statistical Inference for Spatial Processes.* Cambridge University Press, Cambridge, England, 1988.
Stehman, S., and Overton, W. Variance estimation for fixed-configuration, systematic sampling. *Technical Report* No. 134, Department of Statistics, Oregon State University, 1989.
Stetzer, F. Specifying weights in spatial forecasting models: the results of some experiments. *Environment and Planning A*, 14, 571-584, 1982.
Upton, G., and Fingleton, B. *Spatial Data Analysis by Example*, vol. 1. Wiley, New York, 1985.
Warnecke, L. GIS in the states: applications abound. *GIS World*, 3, (# 3), 54-58, 1990.
Warnecke, L. *State Geographic Information Activities Compendium.* Council of State Governments, Lexington, KY, 1991.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com