Combining Incompatible Spatial Data

Carol A. GOTWAY and Linda J. YOUNG

Global positioning systems (GPSs) and geographical information systems (GISs) have been widely used to collect and synthesize spatial data from a variety of sources. New advances in satellite imagery and remote sensing now permit scientists to access spatial data at several different resolutions. The Internet facilitates fast and easy data acquisition. In any one study, several different types of data may be collected at differing scales and resolutions, at different spatial locations, and in different dimensions. Many statistical issues are associated with combining such data for modeling and inference. This article gives an overview of these issues and the approaches for integrating such disparate data, drawing on work from geography, ecology, agriculture, geology, and statistics. Emphasis is on state-of-the-art statistical solutions to this complex and important problem.

KEY WORDS: Change of support; Data assimilation; Ecological inference; Modifiable areal unit problem; Multiscale processes; Spatially-misaligned data.

1. INTRODUCTION

One of the most challenging and fascinating areas in spatial statistics is the synthesis of spatial data collected at different spatial scales. Advances in satellite imagery and remote sensing now permit scientists to access spatial data at several different resolutions. For example, data from Landsat thematic mapper (TM) scenes have a resolution of 30 m^2 (the size of the area on the ground represented by each unit in the image), the low-cost advanced very-high-resolution radiometer (AVHRR) has a much coarser resolution (2.2–16 km^2), and aerial photographs can be digitized to a very fine resolution (1 m^2 or smaller). The advent of the Internet has led to an explosion in the number of readily available datasets. Studies working with disparate data at multiple scales have become common, causing increased concern over issues surrounding scale. Recent advances in geographic information systems (GISs) make it possible to use spatial data from different resolutions and combine them with other types of disparate spatial information (e.g., networks, roads, utility pipelines, census tract information). Users of spatial information are frequently faced with the problem of how best to integrate such information.

The overall problem of "incompatible" spatial data has been encountered in several fields of study, and numerous terms have been introduced to describe one or more facets of the problem as well as various solutions to it. These terms include the ecological inference problem, the modifiable areal unit problem, spatial data transformations, the scaling problem, inference between incompatible zonal systems, block kriging, pycnophylactic geographic interpolation, the polygonal overlay problem, areal interpolation, inference with spatially misaligned data, contour reaggregation, multiscale and multiresolution modeling, and the change of support problem. In this article, we give a multidisciplinary history of this problem, with a description of early proposed solutions. We then review more recent statistical approaches, emphasizing the underlying assumptions and evaluating the progress made toward combining incompatible spatial data.

2. A HISTORICAL PERSPECTIVE

The choice of an appropriate scale for the study of spatial processes is an extremely important one because mechanisms vital to the spatial dynamics of a process at one scale may be unimportant or inoperative at another. Moreover, relationships between variables at one scale may be obscured or distorted when viewed from another scale. This is particularly true in the study of human, animal, and plant populations and has led many researchers in agriculture, geography, sociology, statistics, ecology, and the earth and environmental sciences to consider scale issues in detail (Fairfield Smith 1938; Yule and Kendall 1950; Robinson 1950; Selvin 1958; Matheron 1963; Hannan 1971; Moellering and Tobler 1972; Ghil, Cohn, Tavantzia, Bube, and Isaacson 1981; Morganstern 1982; Openshaw 1984; Turner, O'Neill, Gardner, and Milne 1989; Fotheringham and Wong 1991; Daley 1991; Richardson 1992; Levin 1992; Cressie 1993a, 1996; Jelinski and Wu 1996; King 1997).

2.1 The Modifiable Areal Unit Problem

In many instances, spatial aggregation is necessary to create meaningful units for analysis. This latter aspect was perhaps best described by Yule and Kendall (1950, p. 312), who stated that "geographical areas chosen for the calculation of crop yields are modifiable units and necessarily so. Since it is impossible (or at any rate agriculturally impractical) to grow wheat and potatoes on the same piece of ground simultaneously we must, to give our investigation any meaning, consider an area containing both wheat and potatoes and this area is modifiable at choice." Geographers have long had an appreciation for the problems associated with the use of modifiable units. These problems led Openshaw and Taylor (1979) to coin the term *modifiable areal unit problem*, now often referred to simply as the MAUP.

Gehlke and Biehl (1934) first documented that statistical inference could change with scale using two census tract variables. They found that when *contiguous* census tracts were

Carol A. Gotway is Senior Mathematical Statistician, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA 30333 (E-mail: *cdg7@cd.gov*). Linda J. Young is Professor, Department of Biometry, University of Nebraska, Lincoln, NE 68583 (E-mail: *ljyoung@unlnotes.unl.edu*). The authors gratefully acknowledge all of the authors whose work is represented and discussed in this article, who have made outstanding contributions to a persistent and perplexing problem. They especially thank Brad Carlin, Noel Cressie, Hsin-Cheng Huang, Chris Wikle, and Jeremy Aldworth, who were willing to share their work and ideas, some of which has not yet been published. Finally, they thank Noel Cressie for his review of this article during its infancy stages. His comments and suggestions, and his willingness to review a crude draft of a lengthy manuscript, are much appreciated.

grouped to form larger areas, the magnitude of the correlation coefficient between the two variables increased. However, when *random* census tracts were grouped, the correlations were unaffected by the group size. The impact of different grouping methods, such as random, by proximity, or by values of either the dependent or independent variable, was then studied by Yule and Kendall (1950), Blalock (1964), and Clark and Avery (1976). When random grouping was used, there was no systematic effect on the correlation coefficient or on the estimated slope parameter from simple linear regression. When the grouping was based on values of the dependent variable, both the correlation coefficient and the estimated slope parameter increased with scale. If the grouping criterion was based on values of the independent variable, then the correlation coefficient increased with increasing aggregation, but grouping had no systematic effect on the estimated slope parameter. Finally, when proximal units were grouped, the estimates of both the correlation coefficient and the slope parameter increased with the level of grouping, although the correlation coefficient increased only up to a certain level of aggregation, after which the correlations began to decrease. To better understand the nature of the problem and to suggest efficient groupings for geographical data, Openshaw and Taylor (1979) considered 99 counties in Iowa and constructed all possible groupings of these counties into larger districts. Their results were somewhat startling. When the correlation between the percentage of Republican voters and the percentage of elderly voters was considered, 12 districts could be contrived to produce correlations ranging from $-.97$ to $+.99$. Moreover, no obvious relationship seemed to exist between the spatial characteristics of the districts and the variation in the resulting correlation coefficients.

These studies illustrate that the MAUP is two interrelated problems. The first problem concerns the different inferences obtained when the same set of data is grouped into increasingly larger areal units. Often referred to as the *scale effect* or *aggregation effect*, this has received the most attention by statisticians. The second problem, often termed the *grouping effect* or the *zoning effect*, considers the variability in results due to alternative formations of the areal units leading to differences in unit shape at the same or similar scales (Openshaw and Taylor 1979, 1981; Openshaw 1984; Wong 1996). Both issues can be, and often are, present in a single analysis.

Theoretical reasons for the increase in correlations as the level of aggregation increases have been provided in the works of Robinson (1950), Prais and Aitchison (1954), Robinson (1956), and Cramer (1964) in the context of the simple linear regression model. Prais and Aitchison (1954) developed the use of *grouping matrices* to measure the grouping effects and suggested weighted least squares for unbiased estimation of the slope coefficient. This idea was extended by Williams (1976), Haitovsky (1984), and Arbia (1986) to more complicated grouping arrangements and the use of generalized least squares. However, the loss of efficiency resulting from aggregation cannot be removed simply by using a weighted analysis. Thus correlation coefficients and significance tests based on aggregated data tend to be inflated. Areal weighting, first suggested by Robinson (1956), is an effective solution to the MAUP only in very specialized situations (Thomas and Anderson 1965), and in general will not solve the problem caused by the zoning effect. The effects of the MAUP go beyond such simple statistics as the variance and the correlation coefficient discussed here. Inferential problems also occur in multivariate regression analysis (Fotheringham and Wong 1991), Poisson regression (Amrhein and Flowerdew 1992), hierarchical random coefficient models (Goldstein 1995; Steel and Holt 1996), spatial interaction models (Putnam and Chung 1989), spatial autocorrelation statistics (Jelinski and Wu 1996), and undoubtedly in many other statistical models and estimation procedures.

2.2 The Ecological Fallacy

In epidemiology, the term *ecological inference* refers to the process of deducing individual behavior from aggregate data. This term is from Robinson (1950), who noted that in *ecological correlation,* the statistical object is a group of persons. He concluded that ecological and individual correlations are almost certainly not equal, leading him and others to question the results of numerous studies in which conclusions on individual behavior had been drawn from grouped data. Thus the *ecological fallacy* occurs when analyses based on grouped data lead to conclusions different from those based on individual data (Selvin 1958). The resulting bias is often called "ecological bias" (Richardson 1992; Greenland and Robins 1994). This is comprised of two components: *aggregation bias* due to the grouping of individuals and *specification bias* due to the differential distribution of confounding variables created by grouping (Morganstern 1982). These are analogous to the scale and zoning effects in the MAUP, and the ecological inference problem can be viewed as a special case of the MAUP and not the converse, as advocated by King (1997).

Ecological bias has been well documented in the literature (Robinson 1950; Richardson, Stucker, and Hemon 1987; Piantadosi, Byar and Green 1988; Greenland and Morganstern 1989; Walter 1991; Richardson 1992; Klein and Freedman 1993), as have "inference rules" and "solutions" (Goodman 1959; Firebaugh 1978; Richardson 1992; King 1997). Cleave, Brown, and Payne (1995) and King (1997) have provided critical discussions of some of these in the context of contingency table analysis where individuals are divided into areal units (e.g., census tracts, voting districts), and then cross-classified by other variables (e.g., sex, race). Often the cause of specification bias is the failure to incorporate relevant spatial information about individuals (e.g., Klein and Freedman 1993). This type of analysis is not of primary interest here. Instead, our focus is on spatial data and methods for combining spatial data from disparate sources.

2.3 Sources of the Modifiable Areal Unit Problem and Ecological Bias

The smoothing effect that results from averaging is the underlying cause of both the scale problem in the MAUP and aggregation bias in ecological studies. As heterogeneity among units is reduced through aggregation, the uniqueness of each unit and the dissimilarity among units is also reduced. There is, however, another mitigating factor: spatial autocorrelation. The decrease in variance is moderated by the positive autocorrelation among the original observations, and exacerbated by negative autocorrelation. Arbia (1986) and Cressie

(1993b) have illustrated the effect of autocorrelation on the variance of the sample mean. When areal units are similar to begin with, the aggregation process results in much less information loss than when the units are highly dissimilar. To further compound the problem, the aggregation process itself induces positive spatial autocorrelation, particularly if the aggregation process allows overlapping units (e.g., moving averages). Cross-correlations with variables in neighboring units also can affect the results (Wong 1996).

The smoothing effect and resulting alterations in the spatial autocorrelation of the units are also sources of the zoning effect. The MAUP does not exist, or at least its effects are much less pronounced, when aggregation of areal units is performed in a noncontiguous or spatially random fashion. Only when contiguous units are combined, altering the spatial autocorrelation among the units, is the zoning effect of MAUP most apparent. Because the variation among the original areal units is not uniform over the entire region, merging smaller units is analogous to smoothing different combinations of spatial neighbors. Depending on the similarity of the neighbors, different zoning rules may lead to different analytical results. Thus, given the plethora of ways of combining even a relatively small number of spatial units, it is not hard to see how one could produce "a million or so correlation coefficients" as Openshaw and Taylor (1979) found.

The complex facets of the smoothing effect can be very difficult to sort out in any given application, leading to a vast literature on this problem. Some information is lost by aggregating and going to increasingly larger scales. Systematic effects, caused by either the aggregation aspect or the zoning aspect of the MAUP, depend on the spatial relationships among the original data values, the statistics being calculated, and the way in which the units are aggregated. Thus the effects of the MAUP and ecological bias that may occur in any particular application are often difficult or impossible to ascertain. But this fundamental understanding of the problem makes it clear that any solution must find a way to (1) account for or circumvent the loss of information due to aggregation and (2) relate the variation among the aggregated units to the variation among the original units composing each aggregate. Although difficult, this can be done, as described in Sections 3, 4, and 5.

2.4 The General Change of Support Problem

The different types of spatial data (point, line, area, surface), occurring naturally or as a result of the measurement process, potentially allow many ways of integrating these different types of spatial data. Arbia (1989) uses the term *spatial data transformations* to refer to situations in which the spatial process of interest is inherently of one form but the data observed are of another form, resulting in a "transformation" of the original process of interest. For example, sometimes the data are just not available at the desired scale of interest. Meteorologic processes occur over a continuum, but only point observations along such a surface can be recorded. Individual-level inference may be desired, but to ensure data confidentiality, only aggregate data are made available. These situations and all of Arbia's spatial data transformations are special cases of what is called the *change of support problem* (COSP) in geostatistics. The term "support" has come to mean simply the size or volume associated with each data value, but the complete specification of this term also includes the geometrical size, shape, and spatial orientation of the regions associated with the measurements (see, e.g., Olea 1991). Changing the support of a variable (typically by averaging or aggregation) creates a new variable. This new variable is related to the original one, but has different statistical and spatial properties. The problem of how the spatial variation in one variable associated with a given support relates to that of the other variable with a different support is the COSP. Table 1, modified from Arbia (1989), delineates some common COSPs.

Both the ecological inference problem and the MAUP are just specific COSPs. Many other terms have also been introduced to describe particular COSPs and solutions to particular COSPs including the scaling problem, inference between incompatible zonal systems, block kriging, pycnophylactic geographic interpolation, the polygonal overlay problem, areal interpolation, inference with spatially misaligned data, contour reaggregation, and multiscale and multiresolution modeling. Many of these are discussed in more detail in subsequent sections.

COSPs may result when studying a single spatial variable or when trying to relate two spatial variables of different supports. For example, consider a retrospective epidemiologic study designed to measure the effect of air quality on mortality or morbidity. Often in such studies, only aggregate health data are available, reflecting cost, time, and confidentiality considerations. This is an area-to-point COSP, because aggregate data must be used to make inferences about individuals. Another COSP problem results when trying to link the exposure data to the health outcome information, because the two variables have inherently different scales. This serves to illustrate a more fundamental problem not alleviated by even the most sophisticated measurement process. Disease is specific to an individual, but air quality varies over a continuum—how can these two different types of data be related in a way that permits valid inference?

Table 1. Examples of COSPs

We observe or analyze	*But the nature of the process is*	*Examples*
Point	Point	Point kriging; prediction of undersampled variables
Area	Point	Ecological inference; quadrat counts
Point	Line	Contouring
Point	Area	Use of areal centroids; spatial smoothing; block kriging
Area	Area	The MAUP; areal interpolation; incompatible/misaligned zones
Point	Surface	Trend surface analysis; environmental monitoring; exposure assessment
Area	Surface	Remote sensing; multiresolution images; image analysis

3. GEOSTATISTICAL SOLUTIONS TO THE CHANGE OF SUPPORT PROBLEM

Many of the statistical solutions to the COSP can be traced back to Krige's "regression effect" and subsequent corrections used in mining blocks of ore in the 1950s (Krige 1951). These were more formally developed into the beginning of the field of geostatistics by Matheron (1963). Point kriging is one solution to the point-to-point COSP, but geostatistics was really invented for more general COSPs. The basic geostatistical concepts of support and change of support have been presented by Clark (1979) and Armstrong (1999). More general solutions to change of support problems have been discussed by Journel and Huijbregts (1978), Chiles and Delfiner (1999), and Cressie (1993a, 1996).

3.1 Block Kriging

Consider the process $\{Z(s) : s \in D \subset \Re^d\}$, where $Z(s)$ represents the value of the random variable at a known location s and s varies continuously over a spatial domain D. Assume that $Z(s)$ has mean $\mu(s)$ and covariance function $\text{cov}(Z(u), Z(v)) = C(u, v)$ for u, v in D. Suppose that instead of observing a realization of this process, data $Z(B_1), Z(B_2), \ldots, Z(B_n)$ are collected, where

$$Z(B_i) = \frac{1}{|B_i|} \int_{B_i} Z(s)\,ds \tag{1}$$

and $|B_i|$ is the volume of $B_i \subset D$, $i = 1, 2, \ldots, n$. In geostatistics, B_i is called the spatial *support* of $Z(B_i)$. Assume that for $B \subset D$, inference is to be made on $Z(B)$, $B \neq B_i$, $i = 1, 2, \ldots, n$. The COSP is concerned with drawing inference on $Z(B)$ from data $Z(B_1), Z(B_2), \ldots, Z(B_n)$. This can be done if we can derive the distribution of $Z(B)$ (or at least its first two moments in the linear case) from the data $Z(B_1), Z(B_2), \ldots, Z(B_n)$.

The moments of $Z(B)$ can be derived from the moments of the underlying process. Thus, if $E(Z(s)) = \mathbf{x}(s)'\boldsymbol{\beta}$, where $\mathbf{x}(s) = (x_1(s), x_2(s), \ldots, x_p(s))'$ is a $p \times 1$ vector of explanatory variables, then $E(Z(B)) = \mathbf{x}(B)'\boldsymbol{\beta}$, where $\mathbf{x}(B) = (x_1(B), x_2(B), \ldots, x_p(B))'$, and $x_j(B) = \frac{1}{|B|}\int_B x_j(u)du$, $j = 1, 2, \ldots\ldots, p$ (see Cressie 1996). Note that the regression coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ are invariant to the change of support if it is also reflected in the explanatory variables (Arbia 1989; Cressie 1996). The $\text{cov}(Z(B_i), Z(B_j))$ may be written as

$$\begin{aligned}\text{cov}(Z(B_i), Z(B_j)) &= \overline{C}(B_i, B_j) \\ &= \int_{B_j}\int_{B_i} C(u, v)\,du\,dv/|B_i||B_j| \end{aligned} \tag{2}$$

(see Journel Huijbregts 1978; Cressie 1993b, 1996). Because "block support" covariances (i.e., those pertaining to support B) can be expressed in terms of "point support" covariances (i.e., those pertaining to support s), (2) forms a basis for solutions to the COSP. It is important to note that the behavior of the covariances depends not only on the point support covariance, but also on the specific blocks (and not just their volumes) considered.

The solution to most COSPs requires spatial prediction. A common inferential problem is the prediction of $Z(B)$ from point samples $Z(s_1), Z(s_2), \ldots, Z(s_n)$. In linear geostatistics, this predictor is called the *universal block kriging* predictor and is given by $\widehat{Z}(B) = \sum_{i=1}^n \lambda_i Z(s_i)$, where optimal weights $\{\lambda_i\}$ are obtained by solving (Journel and Huijbregts 1978; Chiles and Delfiner 1999)

$$\sum_{k=1}^n \lambda_k C(s_i, s_k) - \sum_{j=1}^p m_j x_j(s_i) = \overline{C}(B, s_i) \quad i = 1, \ldots, n$$

and

$$\sum_{i=1}^n \lambda_i x_j(s_i) = x_j(B) \quad j = 1, \ldots, p. \tag{3}$$

These equations result from minimizing prediction mean-squared error (PMSE) subject to unbiasedness constraints as in best linear unbiased prediction and universal kriging. The m_j are Lagrange multipliers from the constrained minimization, and $\overline{C}(B, s_i)$ is the point-to-block covariance given by

$$\overline{C}(B, s_i) = \text{cov}(Z(B), Z(s_i)) = \int_B C(u, v)\,du\,dv/|B|. \tag{4}$$

The prediction PMSEs look similar to those associated with the universal (point) kriging predictor as given by Chiles and Delfiner (1999).

The covariance function, $C(\cdot)$ (here a point-to-point covariance), is assumed known for theoretical derivations, but is then estimated and modeled with a valid positive definite function based on the data. In practice, integrals are computed by discretizing B into points, $\{u_j'\}$, so that (4) is approximated using $\overline{C}(B, s_i) \approx 1/N \sum_{j=1}^N C(u_j', s_i)$. Thus, given observations at locations with point support, block kriging can be used to predict the average value of the process at a larger scale, accounting not for only the size, but also for the shape and orientation of the blocks, and hence is a solution to the point-to-area COSP. Carroll, Day, Cressie, and Carroll (1995) used these ideas to combine both ground-based point observations with areal block measurements into a single analysis that accounted for the differing supports.

Now consider data with support smaller than the block for which prediction is desired but larger than point support. A practical application of this problem is one where an attribute is measured for census tracts or zip codes and prediction of this attribute at the county level is desired. Thus $Z(A_1), \ldots, Z(A_n)$ are observed and $Z(B)$, with $|A_i| < |B|$, is to be predicted. Note that $Z(A_i)$ is still related to the point support covariance through a relationship analogous to (1). The optimal linear predictor of $Z(B)$ based on data $\{Z(A_i)\}$ is $\widehat{Z}(B) = \sum_{i=1}^n \lambda_i Z(A_i)$, where the optimal weights $\{\lambda_i\}$ are solutions to the equations obtained by replacing $x_j(s_i)$ with $x_j(A_i)$ and replacing the point-to-point covariances $C(s_i, s_j)$ and the point-to-block covariances $\overline{C}(B, s_i)$ in (3) with

$$\begin{aligned}\overline{C}(A_i, A_j) &= \text{cov}(Z(A_i), Z(A_j)) \\ &= \int_{A_j}\int_{A_i} C(u, v)\,du\,dv/|A_i||A_j|, \end{aligned} \tag{5}$$

and

$$\overline{C}(B, A_i) = \operatorname{cov}(Z(B), Z(A_i)) = \int_B \int_{A_i} C(u, v)\, du\, dv/|B||A_i|.$$

Because data on any support can be built from data with point support, these relationships can be used for both the case where $|A_i| < |B|$ (aggregation) and the case where $|B| < |A_i|$ (disaggregation). However, unlike in the previous case, where we observed point support data and hence could estimate the point support covariance function, $C(u, v)$, in practice this function cannot be estimated directly from aggregate data. Cressie (1993b) suggested a practical approach to building a point-level model in which a parametric form for $C(u, v)$ is assumed and the theoretical covariance on the right side of (5) is equated to the empirical covariance function of $Z(A)$, estimated and modeled from available data. Once the parameters of the point-to-point covariance function are estimated, the foregoing methods can be used to aggregate and disaggregate data with any support. Gelfand, Zhu, and Carlin (2001) eloquently implemented this idea using Bayesian hierarchical models and Gibbs sampling.

In the bivariate COSP based on a linear regression situation where $Y(s) = X(u) + \epsilon(s)$ and the explanatory covariate $X(u)$ is measured at different locations than the response $Y(s)$, kriging and block kriging can be used to form part of a solution to the COSP. First, kriging can be used to predict $X(s)$. Point kriging can be used if both s and u have point support, and block kriging can be used if s and/or u have nonpoint support. Kriging variances for both point and block kriging can also be easily obtained. This converts the COSP into a spatial "errors-in-variables" problem, $Y(s) = \widehat{X}(s) + \epsilon'(s)$, in which current methods and software in this area (e.g., Littell, Milliken, Stroup, and Wolfinger 1996) can then be used. Alternative approaches are described in subsequent sections.

Block kriging is routinely used to solve the point-to-area COSP in geostatistics. Like kriging, it is relatively flexible and applicable to a wide variety of problems, and there are many choices for covariance models and parameter estimation. However, it does require inversion of large matrices and thus can be computationally prohibitive with large datasets. Its use for other COSPs, although theoretically possible, has been limited, largely because of a lack of need in mining applications and confusion with ordinary kriging.

3.2 Cokriging

Multivariate spatial prediction, or *cokriging*, was developed to improve the prediction of an "undersampled" spatial variable by exploiting its spatial correlation with a related spatial variable that is more easily and extensively measured. Consider predicting a spatial variable of interest, $Z(s_0)$, from data $Z(s_1), Z(s_2), \ldots, Z(s_n)$, and $X(u_1), X(u_1), \ldots, X(u_m)$. The spatial locations denoted by $\{u_j\}$ need not coincide with the spatial locations denoted by $\{s_i\}$. For simplicity, assume that both processes $Z(\cdot)$ and $X(\cdot)$ have constant, but unknown means, although a more general "universal" cokriging predictor could be considered (Chiles and Delfiner 1999). The optimal (ordinary cokriging) predictor is a linear combination of all available data,

$$\widehat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) + \sum_{j=1}^{m} \omega_j X(u_j),$$

where the weights $\{\omega_j\}$ and $\{\lambda_i\}$ are determined by

$$\begin{aligned}
&\sum_{i=1}^{n} \lambda_i C_{ZZ}(s_i, s_j) + \sum_{i=1}^{m} \omega_i C_{XZ}(u_i, s_j) + m_1 \\
&\qquad = C_{ZZ}(s_0, s_j), \quad j = 1, 2, \ldots, n, \\
&\sum_{i=1}^{n} \lambda_i C_{ZX}(s_i, u_j) + \sum_{i=1}^{m} \omega_i C_{XX}(u_i, u_j) + m_2 \\
&\qquad = C_{ZX}(s_0, u_j), \quad j = 1, 2, \ldots, m, \\
&\sum_{i=1}^{n} \lambda_i = 1, \qquad \text{and} \qquad \sum_{i=1}^{m} \omega_i = 0. \qquad (6)
\end{aligned}$$

As with ordinary kriging, these equations result from minimizing the PMSE subject to unbiasedness conditions. Here m_1 and m_2 are Lagrange multipliers from the constrained minimization, $C_{XX}(u_i, u_j) = \operatorname{cov}(X(u_i), X(u_j))$ and $C_{ZZ}(s_i, s_j) = \operatorname{cov}(Z(s_i), Z(s_j))$ are the covariance functions of the X and Z processes (called *autocovariance* functions), and $C_{ZX}(s_i, u_j) = \operatorname{cov}(Z(s_i), X(u_j))$ is the cross-covariance function between the two processes. (The full details of cokriging and the prediction standard error for the cokriging predictor can be found in, e.g., Isaaks and Srivastava 1989 and Wackernagel 1995.) These equations are valid regardless of the support of the data, but valid inference procedures for the autocovariances and cross-covariances that take into account differing supports are crucial for their use in COSP problems. When both Z and X are of point support, there are many approaches to estimation and modeling of the autocovariance and cross-covariance functions, including coregionalization models (Isaaks and Srivastava 1989; Wackernagel 1995), the use of "pseudo" or variance-based cross-variograms (Clark, Basinger, and Harper 1989; Myers 1991; Papritz, Kunsch, and Webster 1993; Cressie and Wikle 1998), Bayesian hierarchical models (Le and Zidek 1992), and moving average representations (Ver Hoef and Barry 1998). Once a consistent model for the point covariance functions is developed, block cokriging [i.e., prediction of $Z(B)$ from point data $Z(s)$ and $X(s)$] can be done by replacing the cross-covariances in (6) by their block-averaged counterparts (Myers 1984). Thus cokriging can be used for as a solution to both the point-point and point-block COSPs, although there is still great debate on how best to define, estimate, and model the second-order cross-variable relationships.

An interesting solution to the area-to-point COSP (the ecological inference problem) can be considered as a special case of cokriging in which data are available on only one of the spatial variables. This solution in the context of prediction from binomial data was considered by McNeill (1991) in mapping the spatial distribution of bird species and then by Oliver, Lajaunie, Webster, Muir, and Mann (1993) and Webster, Oliver, Muir, and Mann (1994) in predicting the risk of a rare disease. In this context, let $\{\pi(s) : s \in D \subset \Re^2\}$ denote a latent risk process, and assume that this process has

mean μ_π, and covariance function $C_{\pi\pi}(\cdot)$. Further assume that, conditional on this process, the observed frequencies, $R(u_i)$, associated with an area centered at location u_i, are independent binomial random variables with means $\pi(u_i)$ and variances $\pi((u_i)(1-\pi(u_i))/n(u_i)$, where $n(u_i)$ is the population in the ith area. In the cokriging context discussed earlier, with only one variable observed, the cokriging predictor and equations reduce to

$$\hat{\pi}(s_0) = \sum_{i=1}^{n} \lambda_i R(u_i),$$

where the weights are obtained from

$$\sum_{j=1}^{n} \lambda_j C_{RR}(u_i, u_j) + m = C_{\pi R}(s_0, u_i), \quad i = 1, \ldots, n$$

and

$$\sum_{i=1}^{n} \lambda_i = 1.$$

To use these equations, the cross-covariance function between the risk process and the observed frequencies must be inferred from the data. Given the model assumptions, standard conditioning arguments can be used to show that (see, e.g., McNeill 1991)

$$C_{\pi R}(s_0, u_i) = C_{\pi\pi}(s_0, u_i) \tag{7}$$

and

$$C_{RR}(u_i, u_j) = \left\{1 - \frac{\delta_{ij}}{n(u_i)}\right\} C_{\pi\pi}(u_i, u_j) + \frac{\delta_{ij}}{n(u_i)} \mu_\pi (1 - \mu_\pi),$$

where $\delta_{ij} = 1$ when $i = j$ and 0 otherwise. These equations relate the moments of the data process R to those of the latent process π, allowing estimation and inference on π from the frequency data. However, assuming that the covariance function of the aggregated data process is the same as that of the underlying risk process assumed to be of point support may not be realistic, particularly for physical processes whose mechanisms vary with scale.

3.3 Nonlinear Geostatistics and the Change of Support Problem

In many cases, $E(Z(B)|\mathbf{Z})$ is not linear in the data $\mathbf{Z}$; in others, prediction of a nonlinear function of $Z(B)$ is of interest. These problems require more information about the conditional distribution of $Z(B)$ given the data, $F_B(z|\mathbf{Z}) = P(Z(B) \le z|\mathbf{Z})$, than that used for linear prediction. Moreover, in many cases, such as mining and environmental remediation, the quantity $\Pr(Z(B) > z|\mathbf{Z})$ has meaning in its own right (e.g., proportion of high-grade blocks available in mining evaluation or the risk of contamination in a volume of soil). Nonlinear geostatistics offers solutions to COSPs that arise in this context.

3.3.1 The Multi-Gaussian Approach. The multi-Gaussian approach (Verly 1983) to nonlinear prediction in the point-to-block COSP assumes that available point data $Z(s_1), \ldots, Z(s_n)$ can be transformed to Gaussian variables, $\{Y(s)\}$, by $Z(s) = \phi(Y(s))$. The block B is discretized into points $\{u'_j, j = 1, \ldots, N\}$, and $Z(B)$ is approximated as

$$Z(B) \approx \frac{1}{N} \sum_{j=1}^{N} Z(u'_j). \tag{8}$$

Then

$$F_B(z|\mathbf{Z}) \approx P\left(\frac{1}{N}\sum_{j=1}^{N} Z(u'_j) < z | Z(s_1), Z(s_2), \ldots Z(s_n)\right)$$
$$= P\left(\sum_{j=1}^{N} \phi(Y(u'_j)) < Nz | Y(s_1), Y(s_2), \ldots Y(s_n)\right).$$

This probability is estimated through simulation. The vector $\mathbf{Y}(u) = (Y(u_1), \ldots Y(u_N))'$, is simulated from the conditional distribution of $\mathbf{Y}(u)|\mathbf{Y}(s)$. Because $\mathbf{Y}$ is Gaussian, this conditional distribution can be obtained by kriging, and simulation is straightforward. Then $F_B(z|\mathbf{Z})$ is estimated as the proportion of vectors satisfying $\sum_{j=1}^{N} \phi(Y(u'_j)) < Nz$.

If instead of point support data, data $Z(A_1), \ldots Z(A_n)$, $|A_i| < |B|$, are available, then this approach can still be used provided that an approximation similar to that of (8) remains valid. More general COSP models based on the multi-Gaussian approximation may be possible by building models from data based on point support as described in Section 3.1, or by using geostatistical simulation approaches similar to those described in the next section (Goovaerts 1997).

3.3.2 The Use of Indicator Data. When both the available data and the value to be predicted have point support, the quantity of interest is $F_{s_0}(z|\mathbf{Z}) = P(Z(s_0) \le z|\mathbf{Z})$. This distribution can be estimated by kriging the indicator $I(Z(s_0) \le z)$ from indicator data $I(Z(s_1) \le z), \ldots, I(Z(s_n) \le z)$ (Journel 1983), where

$$I(Z(s) \le z) = \begin{cases} 1 & \text{if } Z(s) \le z \\ 0 & \text{otherwise.} \end{cases}$$

Indicator kriging gives the optimal predictor, $E(I(Z(s_0) < z)|\mathbf{Z})$, which for indicator data is an estimate of $F_{s_0}(z|\mathbf{Z}) = P(Z(s_0) < z|\mathbf{Z})$. Because some information is lost by using indicator functions, *indicator cokriging* (Journel 1983) that uses k sets of indicators corresponding to various threshold levels, z_k, has been suggested as a better alternative.

For nonlinear prediction in the point-to-block COSP, it is tempting to use block kriging, described in Section 3.1, with the indicator data. However, this will yield a predictor of

$$I^*(B) = \frac{1}{|B|} \int_B I(Z(s) \le z)\, ds,$$

which is the proportion of B consisting of points where $Z(s)$ is at or below z. This quantity is clearly not the same as

$$I(B) = \begin{cases} 1 & \text{if } Z(B) \le z \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

which would provide an estimate of $P(Z(B) \le z|\mathbf{Z})$, the probability that the average value of $Z(\cdot)$ is at or below z. This latter quantity is the one of interest in COSPs. The problem arises with any nonlinear function of $Z(s)$, because the mean of block support data will not be the same as the block average of the point support data. This is also true in the more general COSP based on data with supports A_i that differ from support B.

Goovaerts (1997) suggested a solution to nonlinear block prediction based on simulation. The block is discretized, and data $Z(u'_j)$ are simulated at each discretized node. Simulated block values are then obtained via (8). Based on these simulated block values, block indicator values are constructed using (9). $P(Z(B) \le z|\mathbf{Z})$ is then estimated as the average of these block indicator values. Goovaerts (1997) recommended lower and upper triangular decomposition for the simulation of the Z values, but any conditional simulation technique (i.e., one that forces the realizations to honor the available data) could be used.

3.3.3 Isofactorial Models. Isofactorial models were first adapted to COSPs by Matheron (1984), who used them to build a joint distributions (point-block, block-block) from specified marginals. They have the general form

$$G_{i,j}(dz_i, dz_j) = \sum_{m=0}^{\infty} T_m(i,j)\chi_m(z_i)\chi_m(z_j)G(dz_i)G(dz_j),$$

where the orthonormal polynomials, $\chi_m(z)$, have nice statistical properties: $\chi_0(z) = 1, E(\chi_m(Z_i)) = 0, \text{var}(\chi_m(Z_i)) = 1$, and $\text{cov}(\chi_m(Z_i)), \chi_p(Z_j)) = 0$. The coefficients $T_m(i,j)$ are the covariances of the polynomials of the same order, that is, $T_m(i,j) = \text{cov}(\chi_m(Z_i), \chi_m(Z_j))$. The exact form of the polynomials is determined by the marginal distribution, $G(dz)$. For example, if $G(dz)$ is Gaussian, then Hermite polynomials are used because, for the Gaussian distribution, these are known to have the desired orthonormal properties.

The advantage of using isofactorial models in the prediction of nonlinear functions is the orthogonality of the polynomials. This property allows prediction of each polynomial via a separate kriging system and greatly reduces the computations required. In kriging the polynomials, the covariances needed for the kriging equations are given by the $T_m(i,j)$. These are inferred from assumptions pertaining to the bivariate distribution of the pairs $(Z(s_i), Z(s_j))$. For example, if $(Z(s_i), Z(s_j))$ is bivariate Gaussian with correlation function $\rho(\|i-j\|)$, then $T_m(i,j) = [\rho(\|i-j\|)]^m$.

As an example, suppose that all pairs $(Z(s), Z(u))$ are bivariate normal and we want to predict $I(B)$ in (9). This function can be expanded in terms of Hermite polynomials (see, e.g., Rivoirard 1994) as

$$I(B) = G(z) + \sum_{m=1}^{\infty} \frac{1}{\sqrt{m}} H_{m-1}(z)g(z)H_m(Z(B)),$$

where G is the cumulative Gaussian distribution function and g is the Gaussian density. Then the disjunctive kriging predictor of $I(B)$ is obtained by replacing each $H_m(Z(B))$ with its predictor obtained by kriging based on

$$\sum_{i=1}^{n} \lambda_{mi}[\text{cov}(Z(s_i), Z(s_j))]^m = [\text{cov}(Z(s_i), Z(B))]^m, \quad j = 1, \ldots n. \quad (10)$$

However, to actually implement the kriging, valid models for $T_m(i,j)$ (point-to-point), $T_m(B,j)$ (point-to-block), and $T_m(B,B)$ (block-to-block), needed as the covariances in (10) and for the prediction standard errors, must be constructed simultaneously and any parameters estimated from the data. This has been done only in special cases, such as, using what is called the *discrete Gaussian model*. Details for this model and other isofactorial models [e.g., when $Y(s)$ has a gamma distribution] have been given by Rivoirard (1994) and Chiles and Delfiner (1999).

As an alternative, Cressie (1993a) proposed *constrained kriging*, which uses $f(\lambda'\mathbf{Z})$ to predict $f(Z(B))$. The weights are chosen to minimize the PMSE of $\lambda'\mathbf{Z}$ subject to both an unbiasedness constraint as in ordinary kriging and also to a variance constraint, $\text{var}(\lambda'\mathbf{Z}) = \text{var}(Z(B))$. The function $f(\lambda'\mathbf{Z})$ is easily obtained from the data (see Cressie 1993a for the equations and distributional properties) with no more assumptions than those made for ordinary block kriging. The extra constraint forces this predictor to be more variable, compensating for the smoothness of the ordinary kriging predictor. Simulations of Cressie (1993a) and Aldworth and Cressie (1999) indicate that accurate nonlinear predictions of aggregate data can be made using this approach. An extension of this, *covariance-matching constrained kriging*, has been shown to have even better PMSE properties (Aldworth and Cressie 2002).

4. MULTISCALE MODELING

Prediction across scales is key to understanding many complex physical and biological processes (e.g., Daley 1992; May 1994; Turner, Dale, and Gardner 1989; Levin 1992; Bissonette 1997). Studies at several scales are often needed to achieve this understanding, and attention has recently focused on statistical methods for such multiscale processes.

4.1 Use Scale-Independent Statistics

Perhaps King (1997) best emphasized the need for scale-independent statistics when, with regard to the MAUP, he wrote

> Unfortunately, the statistics used to study these issues have not been aggregation-invariant (or "scale-invariant"). If a researcher wishes to have statistics that are invariant to the areal units chosen, then there is no reason to choose correlation coefficients, which depend heavily on the definition of available areal units. Solving the MAUP only requires developing statistics that are invariant to the level of aggregation (p. 250).

Tobler (1989) had this same idea when he suggested that methods of spatial analysis should be independent of the spatial coordinates used; the problem is not in the choice of units, but with the choice of models and methods used in the analysis. Thus, instead of the Pearson correlation coefficient, Tobler (1989) recommended using the *cross-coherence function* (the spectral equivalent of the cross-variogram). Equally as important is the choice of model on which to base inference.

Both Tobler (1989) and Amrhein and Flowerdew (1992) described models that can be "upscaled" but show no aggregation effects.

Tobler (1989) further suggested choosing models whose parameters change in a predictable manner across scales to solve the aggregation aspect of the MAUP. This is similar to the recommendations made by Cressie (1996, 1998) and to the ideas of Fotheringham (1989), who suggested focusing on rates of change across scales and using the fractal dimension as a scale-independent measure of a spatial relationship. Fractals have been effective multiscale models in several disciplines (e.g., Palmer 1988; Milne 1988; Sugihara and May 1990; Emerson, Lam, and Quattrochi 1999). Other statistical methods with a similar goal include spectral analysis (Renshaw and Ford 1984; Nielsen, Wendroth, and Parlange 1995), entropy decomposition analysis (Theil 1972; Batty 1976; Phipps 1991; Johnson and Patil 1998), nested analysis of variance (Greig-Smith 1952; Moellering and Tobler 1972; Oliver and Webster 1986; Bellehumeur and Legendre 1998; Ver Hoef and Cressie 1993), local image variation graphs in remote sensing (Woodcock and Strahler 1987), geostatistical methods (Legendre and Fortin 1989; Bell et al.1993; Ver Hoef, Cressie, and Glenn-Lewin 1993; Goovaerts 1998), and Markov transition models (Patil and Taillie 1999).

4.2 Multiscale Spatial Tree Models

To describe spatial processes operating at multiple resolutions, Basseville et al. (1992) and Chou, Willsky, and Nikoukah (1994) developed a scale-recursive algorithm based on a multilevel tree. Each level of the tree corresponds to a different spatial scale, with the finest scale at the lowest level of the tree. Let s represent any node on the tree, let s_0 denote the node at the coarsest scale (the "root node") of the tree, and let T denote the collection of all nodes of the tree. Nodes at the very finest scale are referred to as the leaves of the tree, and a node at one scale that is related to nodes at the next finest scale via the branches of the tree is called a parent node, denoted by p_s. A simple tree structure is shown in Figure 1.

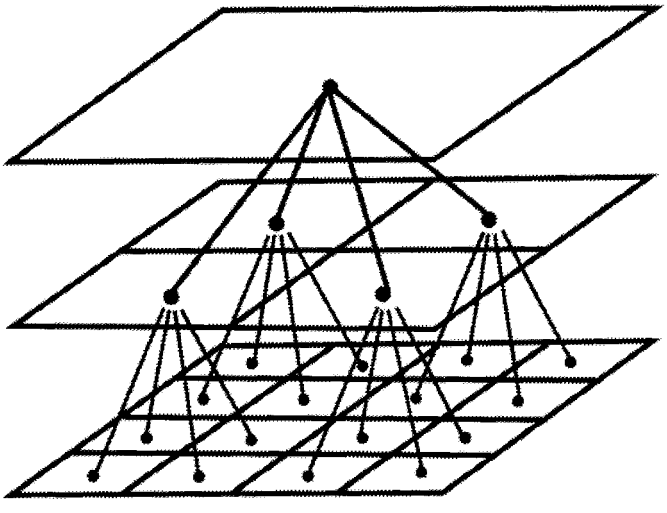

Figure 1. A Tree Structure for Multiscale Processes.

The goal is to predict an unobservable spatial process $\{\mathbf{X}(s), s \in T\}$, called the state process, from a noisy measurement process $\{\mathbf{Z}(s), s \in T\}$, from which data are observed at some nodes of the tree. The measurement process is assumed to be linearly related to the state process via the measurement equation

$$\mathbf{Z}(s) = K(s)\mathbf{X}(s) + \boldsymbol{\epsilon}(s), \tag{11}$$

where $\mathbf{Z}(\mathbf{s})$ is a n x 1 vector of measurements at node s; $\mathbf{X}(s)$ is a m x 1, zero-mean state vector that we would like to predict; $\boldsymbol{\epsilon}(s)$ is a white noise process independent of $\mathbf{X}(s)$, with known covariance matrix $R(s)$ that reflects measurement error in the observations; and $K(s)$ is an n x m deterministic selection matrix that relates the measurements to the state vector. The selection matrix specifies the components of the state vector that are measured and how each of these corresponds to the measurements at node s.

The state vector is not observable, but it is assumed to be related to its parent through the state equation

$$\mathbf{X}(s) = \Phi(s)\mathbf{X}(p_s) + \boldsymbol{\eta}(s), \tag{12}$$

where $\boldsymbol{\eta}(s)$ is a white noise process with covariance matrix $Q(s)$ that is independent of both $\boldsymbol{\epsilon}(s)$ and $\mathbf{X}(p_s)$. In addition to this "downtree" model, a corresponding "uptree" model can be derived. Assuming that $\mathbf{X}(s)$ follows a multivariate Gaussian distribution and using properties of conditional Gaussian distributions together with (12) gives $E(\mathbf{X}(p_s)|\mathbf{X}(s)) = P_{p_s}\Phi'(s)P_s^{-1}\mathbf{X}(s)$, where P_s is the covariance matrix of $\mathbf{X}(s)$. Then the uptree model can be written as

$$\mathbf{X}(p_s) = F(s)\mathbf{X}(s) + \boldsymbol{\omega}(s), \tag{13}$$

where $F(s) = P_{p_s}\Phi'(s)P_s^{-1}$, $\boldsymbol{\omega}(s) = \mathbf{X}(p_s) - P_{p_s}\Phi'(s)P_s^{-1}\mathbf{X}(s)$, and $W(s) \equiv E(\boldsymbol{\omega}(s)\boldsymbol{\omega}'(s)) = P_{p_s}(I - \Phi'(s)P_s^{-1}\Phi(s)P_{p_s})$. If P_0, the prior covariance of $\mathbf{X}(s_0)$ at the root node, is specified, then, from (12), P_s can be calculated recursively as $P_s = \Phi(s)P_{p_s}\Phi'(s) + Q(s)$.

Based on this model, Chou et al. (1994) generalized the Kalman filter to produce optimal predictions of the state vector in two steps. The first step, called an uptree filtering step, proceeds upward from the leaves of the tree to the root, successively computing the optimal predictor of $\mathbf{X}(s)$ and an associated PMSE based on the data at this node and at all nodes on the subtree below s. An additional Kalman filtering algorithm is also used in a "merge step" that combines predictions at the offspring nodes into a single prediction for use in updating prior information for the next prediction. The second step, called a downtree smoothing step, proceeds downward from the root of the tree, giving the optimal predictor of the state vector and an associated PMSE based on all available data. In this way, the algorithm can use data at multiple spatial scales. The algorithm is computationally efficient because it involves only local calculations by making the assumption of conditional independence: Conditional on any node of the tree, each of the subtrees connected to it is assumed to be conditionally independent. Thus computations involving the nodes of each subtree can be processed separately, allowing the order of computations to be proportional to the number of nodes at

the finest scale of the tree. The details of this computational algorithm and potential generalizations have been described in detail by Chou et al. (1994) and Fieguth, Karl, Willsky, and Wunsch (1995), and a Bayesian derivation was given by Huang and Cressie (2000).

To use the foregoing models in practical situations, one must specify the order of the model (the dimension of $\mathbf{X}(s)$) and the system matrices $K(s)$, $\Phi(s)$, $Q(s)$, and $R(s)$. In many practical applications, simple choices for these matrices can lead to flexible and powerful models. For example, Fieguth et al. (1995) considered predicting ocean surface height at a fine resolution from ocean surface measurements at a much coarser resolution. In this case, $\mathbf{X}(s)$ is a scalar representing ocean surface height at node s, and the selection matrix is an indicator function that takes the value 1 if the state vector was measured at a particular node. The matrices $\Phi(s)$ and $Q(s)$, now scalars in this example, were chosen from a class of spectral scaling rules commonly used to describe the ocean surface, and $R(s)$ reflected the known accuracy of the satellite platform used to collect the measurements. In another application, Gabrosek, Huang, and Cressie (1999) analyzed satellite data collected using NASA's total ozone mapping spectrometer associated with five different resolutions. In this analysis, Gabrosek et al. 1999 specified $\text{var}(\epsilon(s)) = \sigma_\epsilon^2$ and $\text{var}(\eta(s)) = \sigma_k^2$ for $k = 1, \ldots, 5$ levels and then used maximum likelihood to estimate these variance components.

The transition matrix, $\Phi(s)$, plays an important role in relating the process at different scales, and the state equation implicitly relates the covariance of the variable at one scale to that at another through $\text{cov}(\mathbf{X}(s), \mathbf{X}(b_s)) = \text{cov}(\Phi(s)\mathbf{X}(b_s) + \boldsymbol{\eta}(s), \mathbf{X}(b_s)) = \Phi(s)P(b_s) + Q(s)$. In view of this, Fieguth et al. (1995) modeled Q(s) as a function that decreases geometrically with scale, and many such models are conceivable. Huang, Cressie, and Gabrosek (2000) developed *heterogeneous* tree models that also allow the variance of the spatial tree process to change with scale. The basic algorithm used for prediction in multiscale tree models provides implicit forms for the covariance of $\mathbf{X}(s)$ at any particular scale and also cross-covariances between $\mathbf{X}(s)$ at different scales. However, because two neighboring points may have different parents, the correlation function of the state process on the finest scales may show unrealistic "blocky" artifacts (Basseville et al. 1992; Huang and Cressie 2000) that are reflected in the predicted surfaces. To alleviate this problem, Huang and Cressie (2000) extended the multiscale spatial tree models to more general graphical models, such as graphical Markov models, and also developed a generalized Kalman filter for these models. Models that do not satisfy the "pycnophylactic" property of Tobler (1979), which forces the average of all the offspring measurements of a parent node to equal their parent measurement, can also result in inconsistencies. Huang et al. (2000) extended the work of Huang and Cressie (2000) to incorporate such "mass balance" properties.

State-space models and their multilevel tree specification in particular combine model flexibility with the fast computational algorithms needed for modeling complex, high-order processes. However, they do not yet explicity account for changes in *support* that occur with changes in resolution. Although the data may not be explicitly aggregated, the resolution of the images associates each pixel with an area on the ground, and this area becomes larger and larger as the resolution of the image decreases. Thus the data observed at each scale, and their relationship to data at different scales, reflect aggregate properties of the support over which they are defined. Future research in this area will likely focus on the development of explicit models for these change-of-support relationships.

4.3 Bayesian Hierarchical Models for Multiscale Processes

The conditional specification of spatial tree models lends itself easily to a Bayesian approach to multiscale modeling. Instead of the tree structure that is specified node by node, let $\mathbf{Z}$ be a general $n \times 1$ spatial measurement vector and let $\mathbf{X}$ be the $m \times 1$ state vector. Corresponding to the state-space model described earlier, let the conditional distribution of $\mathbf{Z}$ given $\mathbf{X}$ be $[\mathbf{Z} \mid \mathbf{X}] \sim N_n(K\mathbf{X}, R)$, with $\mathbf{X} \sim N_m(\mathbf{m}, Q)$, where R, $\mathbf{m}$ and Q are known but general and the matrix K is a specified $n \times m$ matrix that relates the measurement process to the state vector. Then the posterior distribution of $\mathbf{X}$, $f(\mathbf{X} \mid \mathbf{Z})$, is $N_m(\mathbf{m}^*, Q^*)$ with

$$\mathbf{m}^* = (Q^{-1} + K'R^{-1}K)^{-1}(K'R^{-1}\mathbf{Z} - Q^{-1}\mathbf{m}) \qquad (14)$$

and

$$Q^* = (Q^{-1} + K'R^{-1}K)^{-1}, \qquad (15)$$

and so $\mathbf{m}^* = E(\mathbf{X}|\mathbf{Z})$ is the optimal (minimum PMSE) predictor of $\mathbf{X}$. This is just the kriging predictor of $\mathbf{X}$ from noisy measurements $\mathbf{Z}$ used in Kalman filtering (see, e.g., Brown 1983; Meinhold and Singpurwalla 1983; Cressie 1993b) and is a basic model for *data assimilation* in atmospheric science that unites both observational data and results from deterministic weather forecasting models (Daley 1991).

Relaxing the assumptions in this basic model makes the model more realistic, but also more complex. For example, assume that $R = \sigma_\epsilon^2 D$, where D is known but now σ_ϵ^2 is unknown. (Often D is taken to be an identity matrix, reflecting a situation in which the measurement errors are uncorrelated, but their variance in unknown.) A Bayesian approach is useful for estimating σ_ϵ^2 and accounting for its uncertainty in the prediction of $\mathbf{X}$. It is often assumed that $\sigma_\epsilon^2 \sim IG(a_\epsilon, b_\epsilon)$, where a_ϵ and b_ϵ are specified constants such that σ_ϵ^{-2} has a gamma($a_\epsilon/2$, $b_\epsilon/2$) distribution (see, e.g., O'Hagan 1994) and the prior distributions on X and σ^2 are independent. (A similar development, considered in Kitanidis 1986, Handcock Stein 1993, and Gaudard, Karson, Linder, and Sinha 1999, assumed a prior for $\mathbf{X}$ that also depends on σ^2, a situation not realistic here. The analytical derivations and computational details included in their work may be useful in developing other multiscale models and in other spatial applications.) With this model, $f(\mathbf{X} \mid \mathbf{Z})$, and $f(\sigma_\epsilon^2 \mid \mathbf{Z})$ required for inference on $\mathbf{X}$ and σ_ϵ^2 are more difficult to obtain, but Gibbs sampling is straightforward. The full conditional for $\mathbf{X}$, $f(\mathbf{X} \mid \sigma_\epsilon^2, \mathbf{Z})$, is the same as that given earlier [Gaussian with parameters in (14) and (15)] with $\mathrm{R} = \sigma_\epsilon^2 D$, and the full conditional distribution for σ_ϵ^2 is $IG(a_\epsilon + s, b_\epsilon + n)$, with $s = (\mathbf{Z} - K\mathbf{X})'D^{-1}(\mathbf{Z} - K\mathbf{X})$. This model can be extended to allow a generalized linear model for the measurement process

(Diggle, Tawn, and Moyeed 1998). Another useful extension is to model the covariance matrix of $\mathbf{X}$ as a function of spatial dependence parameters, for example, $\text{cov}(\mathbf{X}(\mathbf{s}_i), \mathbf{X}(\mathbf{s}_j)) = Q(i,j) = \exp(-\alpha_1 \|\mathbf{s}_i - \mathbf{s}_j\|^{\alpha_2})$, with $\alpha_1 > 0$, and $0 < \alpha_2 < 2$. Other covariance functions could be used, (see, e.g., Handcock and Stein 1993; Gaudard et al. 1999). The parameters α_1 and α_2 can also be given prior specifications. Independent uniform priors could be specified (e.g., Diggle et al. 1998), or instead a truncated Gaussian prior could be used for α_1 (e.g., Royle, Berliner, Wikle, and Milliff 1997). The full conditional distributions for these parameters are usually not tractable, and so another algorithm (e.g., Metropolis–Hastings) must be used to sample from these distributions. In many applications, $\mathbf{X}$ describes a physical process, and for such cases it may be useful to model $\mathbf{m} = H\boldsymbol{\beta}$. This second stage of the hierarchy, $\mathbf{X} \sim N_m(H\boldsymbol{\beta}, Q(\boldsymbol{\alpha}))$, has been called the *process model* (Berliner, Royle, Wikle, and Milliff 1999), with the goal of this modeling step being a description of the true, unobserved process through physically motivated conditional distributions. Royle et al. (1997) and Berliner et al. (1999) used this stage to relate a surface wind field ($\mathbf{X}$) to the gradient of a pressure field ($\boldsymbol{\beta}$). The components of the matrix $\mathbf{H}$ were based on differential equations that describe atmospheric dynamics involving pressures and winds. Wikle, Berliner, and Cressie (1998) and Wikle, Milliff, Nychka, and Berliner (2001) used a similar formulation to incorporate temporal variation as well as physical process constraints in describing the dynamics of a tropical wind process. A prior distribution for $\boldsymbol{\beta}$ may also be specified to reflect information and uncertainty about this parameter.

When the measured locations are different from the state locations, K maps the measurements to the nearest state locations (Royle et al. 1997). Thus, if both $\mathbf{Z}$ and $\mathbf{X}$ have point support, then the aforementioned formulation is a solution to the point-point COSP. Wikle et al. (2001) showed how this model can be extended to combine spatial data at different scales. In their development of a space-time model of tropical ocean surface winds, high-resolution, satellite-derived wind estimates were observed at grid locations A_i, and lower-resolution wind data provided by major weather centers were observed at grid locations C_k. Prediction was desired at grid locations B_j, with $|A| < |B| < |C|$. For the conditional measurement equations, Wikle et al. (2001) assumed

$$\left[\mathbf{Z}_A \mid K_A, \mathbf{X}, \sigma_A^2\right] \sim N(K_A\mathbf{X}, \sigma_A^2 I)$$

and

$$\left[\mathbf{Z}_C \mid K_C, \mathbf{X}, \sigma_C^2\right] \sim N(K_C\mathbf{X}, \sigma_C^2 I).$$

Because the data $\mathbf{Z}_A$ were at a finer resolution than the desired prediction grid, K_A was taken to be an incidence matrix that mapped the conditional mean of these observations to the nearest prediction grid location. The data $\mathbf{Z}_C$ were at a coarser resolution than the desired prediction grid, so K_C operated by assuming that the measured data were smoothed versions of the true process. Each observed point was a weighted average of the nine closest prediction points within a distance D. The variances of the measurement errors were allowed to differ, reflecting the different instrumentations used for measurement. Additional hierarchical levels, similar to those discussed earlier, were used to describe the dynamics (both temporal and mechanistic) of the wind processes.

Thus, one general Bayesian hierarchical model for a multiscale process has the following basic form:

$$\begin{aligned}
\left[\mathbf{Z}_A \mid \mathbf{X}, \sigma_A^2\right] &\sim N(K_A\mathbf{X}, \sigma_A^2 D_A), \\
&\qquad D_A \text{ known or the identity matrix;} \\
\left[\mathbf{Z}_C \mid \mathbf{X}, \sigma_C^2\right] &\sim N(K_C\mathbf{X}, \sigma_C^2 D_C), \\
&\qquad D_C \text{ known or the identity matrix;} \\
\left[\mathbf{X} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}\right] &\sim N(H\boldsymbol{\beta}, Q(\boldsymbol{\alpha})), \quad Q(i,j) \\
&\qquad = \exp(-\alpha_1 \|\mathbf{s}_i - \mathbf{s}_j\|^{\alpha_2}),\ \alpha_1 > 0, 0 < \alpha_2 < 2; \\
\boldsymbol{\beta} &\sim N(\boldsymbol{\beta}_0, \Sigma_{\boldsymbol{\beta}}); \\
\sigma_A^2 &\sim IG(a_A, b_A); \\
\sigma_C^2 &\sim IG(a_C, b_C); \\
\alpha_1 &\sim U(a_1, b_1);
\end{aligned}$$

and

$$\alpha_2 \sim U(a_2, b_2),$$

where $\boldsymbol{\beta}_0$, $\Sigma_{\boldsymbol{\beta}}$, a_A, b_A, a_C, b_C, a_1, b_1, a_2, and b_2 are specified. In actual applications, additional hierarchical structures may be used, different covariance models and prior distributions may be chosen, and some simplifications may be required for implementation. [See Royle et al. 1997; Diggle et al. 1998; Wikle et al. 1998; Wikle et al. 2001 for applications and details on choices for the hyperparameters and details of the Markov chain Monte Carlo (MCMC) sampling used for inference.]

An advantage of this type of specification is that the joint distribution of $\mathbf{Z}$ and $\mathbf{X}$ (and, in particular, their cross-covariance matrix) does not have to be specified and modeled. Nevertheless, assumptions about the covariance between $\mathbf{X}$ and $\mathbf{Z}$ are being made. Consider just the first two stages of the models described earlier, $\mathbf{Z}|\mathbf{X} \sim N(K\mathbf{X}, \sigma^2 I)$ and $\mathbf{X} \sim N(H\boldsymbol{\beta}, Q)$. If K is an incidence matrix, then the implicit assumptions are that $\text{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = \sigma^2 + \text{cov}(X(\mathbf{u}_k), X(\mathbf{u}_l))$, where $(\mathbf{u}_k, \mathbf{u}_l)$ is the node closest to $(\mathbf{s}_i, \mathbf{s}_j)$, and $\text{cov}(Z(\mathbf{s}), X(\mathbf{u})) = \text{cov}(X(\mathbf{s}), X(\mathbf{u}))$. Although more levels in the hierarchy are usually considered, and some of these may also allow spatial dependence, assumptions of diagonal covariances and sparse mapping matrices may result in simple approximations to potentially complex multiscale relationships. As with the multiscale tree models described earlier, these hierarchical spatial models do not account for changes in support that result from changes in resolution. It may be possible to formulate COSP models by relating both processes to an underlying process with point support, $\epsilon(\mathbf{s})$. Then, if $\mathbf{Z}$ is measured with support A and $\mathbf{X}$ is measured with support B,

$$[\mathbf{Z} \mid \mathbf{X}, \sigma^2] \sim N(W_A\mathbf{X}, \Sigma_\epsilon^A + \sigma_A^2 I)$$

and

$$\mathbf{X} \sim N(\boldsymbol{\mu}(B), \Sigma_\epsilon^B)$$

with Σ_ϵ^S having (i, j)th element

$$\int_{S_i}\int_{S_j} C_\epsilon(u, v)\, du\, dv/|S_i||S_j|$$

and W_A having (i, j)th element $|A_i \cap B_j|/|A_i|$ to ensure proper relationships between the means of $\epsilon(\mathbf{s})$, $Z(A_i)$, and $X(B_j)$.

Suppose instead that specification of the joint distribution is not problematic. Assume that $Z(\mathbf{s})$ is a Gaussian process and that $\mathbf{Z}_s = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))'$ has mean $\boldsymbol{\mu}_s$ and covariance matrix Σ_{ss}. Then

$$\begin{bmatrix}\mathbf{Z}_s \\ \mathbf{Z}_B\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_s & (\boldsymbol{\beta}) \\ \boldsymbol{\mu}_B & (\boldsymbol{\beta})\end{bmatrix}, \begin{bmatrix}\Sigma_{ss}(\boldsymbol{\alpha}) & \Sigma_{sB}(\boldsymbol{\alpha}) \\ \Sigma_{Bs}(\boldsymbol{\alpha}) & \Sigma_{BB}(\boldsymbol{\alpha})\end{bmatrix}\right),$$

where $\boldsymbol{\mu}_B$ and the elements of $\Sigma_{sB}(\boldsymbol{\alpha})$, $\Sigma_{Bs}(\boldsymbol{\alpha})$, and $\Sigma_{BB}(\boldsymbol{\alpha})$ are obtained by integrating the analogous moments of the point-support process as described in Section 3.1. Using a Bayesian analysis, spatial prediction requires either $f(\mathbf{Z}_{s_0}|\mathbf{Z}_s)$ for the prediction of points from points, $f(\mathbf{Z}_B|\mathbf{Z}_s)$ for the prediction of blocks from points, and $f(\mathbf{Z}_{B_0}|\mathbf{Z}_B)$ for the prediction of blocks from block data. The analytical form of these distributions is easily obtained from the properties of the multivariate Gaussian distribution. Placing prior distributions on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ and then using a MCMC sampling method (e.g., Gibbs sampling) gives $\boldsymbol{\beta}^g$ and $\boldsymbol{\alpha}^g$, which effectively allows all of these predictive distributions to be completely specified. Gelfand et al. (2001) used Gibbs and other MCMC sampling methods to estimate $\boldsymbol{\beta}$ and α and obtain block-block, block-point, and point-block conditional distributions required for solutions to the corresponding COSPs.

5. MAP OVERLAY OPERATIONS AND THE CHANGE OF SUPPORT PROBLEM

Comparing data from different sources, studying the change in a variable over time, or evaluating the relationship between two or more variables are perplexing problems if the areal units are not the same for all variables during each collection period. The units for which data exist are often called *source* units, and those for which data are desired are called *target* units (Markoff and Shapiro 1973). The process of superimposing source and target units has become known in the geographic literature as the *polygon overlay problem.*

5.1 Probabilistic Potential Mapping in a Geographical Information System

Consider an areal event of interest A (e.g., a mineral deposit or a chemical spill) occurring within a domain S and associated explanatory maps ("evidential themes" in GIS terminology) $E_i(j)$, with i indexing different variables and j indexing discrete class states within each map. The goal is to predict $P(A|E_i(j))$. Although both the source units and the target units are pixels within S, A is a polygonal attribute, so that linking the themes to A by pixel (as logistic regression would require) is not easily accomplished. A simple approach that can be used within a GIS was proposed by Bonham-Carter, Agterberg, and Wright (1988) and is implemented by the program Arc-WofE in ArcView GIS (Raines, Bonham-Carter, and Kemp 2000).

This approach determines a *weight of evidence* for each class value of each map, $W_i(j)$, that reflects the degree of spatial association between A and $E_i(j)$. The weights are calculated from logarithms of conditional probabilities, $P(E_i(j)|A)/P(E_i(j))$, that are related to $P(A|E_i(j))$ by Bayes's theorem and assumptions of conditional independence of $E_i(j)$ given A. The probabilities comprising the weights are estimated from area measurements expressed as unit cell counts. Given the weights of evidence and a prior estimate of $P(A)$, a posterior probability map reflecting the probability that a unit cell contains an event A given all explanatory information can be drawn. Uncertainties associated with variability in the estimated weights and missing data are reflected in the uncertainties of the posterior probabilities and also can be mapped. Additional details and examples have been given by Bonham-Carter (1994). Chung and Fabbri (1999) compared several approaches to estimating the posterior probabilities including crude estimation using observed relative frequencies from historical data, Bayes estimation based on the assumption of conditional independence, and a general multivariate linear regression model. In the regression, the conditional probability of the event given the explanatory variables (based on discretizing A into pixels) was regressed on the bivariate conditional probabilities inferred from the historical data. The regression analysis, with and without expert opinion, tended to perform better than the other methods in the validation case studies considered by Chung and Fabbri (1999). It is interesting that Chung and Fabbri (1999) seemed to believe that avoiding the assumption of conditional independence was a positive attribute of this regression model. Although the models advocated by Bonham-Carter et al. (1988) and Chung and Fabbri (1999) may seem rather simple statistically, they are some of the first real attempts at probabilistic modeling within an interactive GIS framework.

5.2 Pixel Aggregation and Areal Weighting

When areal source units are nested within areal target units, three methods of aggregating spatial source data to produce estimates on the target units are averaging, central-pixel resampling, and median. The averaging method assigns the average value of the original units to the newly aggregated unit. The central-pixel resampling method uses the value associated with the central pixel of the aggregated unit. The median value of the original units is associated with the aggregated unit for the median method. Based on a simulation study using various underlying spatial correlation structures and Gaussian errors, Bian and Butler (1999) concluded that the averaging method results in aggregated data that have a more predictable statistical and spatial behavior than the other two methods. The median method was shown to have similar properties, but the central-pixel method was the least predictable. Bian and Butler (1999) also noted that aggregating within the range of spatial autocorrelation can reduce the errors induced by averaging dissimilar units.

Often the source and target units overlap. In this case, the value for a target unit is often taken to be a weighted average of the values for the source units that intersect it, a process called *areal interpolation* (Goodchild and Lam 1980). It is based on a familiar solution to the point-to-area COSP

proposed by Thiessen (1911). Thiessen polygons are still frequently used for interpolation problems in geology and hydrology, particularly in very small sample cases where kriging cannot be used. Two primary forms of areal interpolation have been used: polygon overlay methods and smoothing methods. In polygon overlay methods, the weights are equal to the proportion of the target unit comprised by each source unit (Markoff and Shapiro 1973; Goodchild and Lam 1980). This is often referred to as *proportional allocation* or *areal weighting*. The implicit assumption is that the variable of interest is evenly distributed in the source unit.

5.3 Spatial Smoothing Methods

With these methods, a smooth surface is fitted to data for the source units and used to interpolate values at the nodes of a fine grid. The interpolated values are then summed or averaged over the target units to obtain areal estimates for these units. Let $Z(A_1), \ldots, Z(A_n)$ denote the aeral data (counts or totals) observed in regions $A_i \subset D$, and suppose that there exists an underlying smooth density, $\lambda(\mathbf{s}), \mathbf{s} = (x, y) \subset D, \lambda(\mathbf{s}) \geq 0$. Tobler (1979) suggested choosing $\lambda(x, y)$ to minimize

$$\int\int \left[\left(\frac{\partial \lambda}{\partial x} \right)^2 + \left(\frac{\partial \lambda}{\partial y} \right)^2 \right] dx\, dy$$

subject to the constraints $\lambda(\mathbf{s}) \geq 0$ and

$$\int_{A_i} \lambda(\mathbf{s}) d\mathbf{s} = |A|. \tag{16}$$

Tobler (1979) called the constraint in (16) the *pycnophylactic property*, which ensures that the density process aggregates to the observed data for each region. The solution to this partial differential equation requires specification of boundary conditions. These will affect the smoothness properties of the surface, particularly near the edges of the domain. Tobler (1979) used finite difference methods to solve this constrained minimization and suggested forcing either the surface or its gradient to be 0 at the boundary of the domain. Dyn, Wahba, and Wong (1979) suggested using a different type of spline that allows both the surface and its gradient on the boundary to be determined by the data.

Brillinger (1990) considered a different optimization criterion based on a locally weighted analysis with weights,

$$w_i(x, y) = 1/|A_i| \int\int_{A_i} W(x-u, y-v)\, du\, dv,$$

where $W(\cdot)$ is a specified kernel function. Given the weight function, a locally weighted estimate of the underlying density was obtained by maximizing the weighted log-likelihood of the data. For areal data, taking $Z(A_i) \sim \text{Poisson}(N_i\lambda)$, a locally weighted estimate of λ at (x, y) is

$$\hat{\lambda}(x, y) = \sum_i w_i(x, y) Z(A_i) / N_i \sum_i w_i(x, y).$$

Brillinger (1990) extended this approach to include explanatory covariates (using the common Poisson-lognormal model), and Brillinger (1994) used the delta method to obtain an estimate of the uncertainty associated with estimated density surface at each point. Similar ideas have been used to estimate relative risks of the form $\lambda(\mathbf{s}) = g(\mathbf{s})/f(\mathbf{s})$, where $g(\cdot)$ is the intensity of a disease process and $f(\cdot)$ is the population density (see Müller, Stadtmüller, and Tabnak 1997).

The weight function in locally weighted smoothing and the use of the pycnophylactic property in the Laplacian smoother of Tobler (1979) force the density estimates to account for the differing supports of the regions A_i. Thus these smoothing methods are one solution to the area-point COSP, although the locally weighted smoothers may not satisfy the pycnophylactic property. Averaging the density estimates over different domains is a solution to the area-area COSP (i.e., the MAUP). However, both approaches assume independent data and thus ignore any spatial correlations in the areal data. More sophisticated smoothing algorithms may be able to account for this correlation, as well as any measurement error in the aeral data.

5.4 Areal Regression Models

Flowerdew and Green (1989, 1992, 1994) used explanatory variables collected on the target units to improve estimates from areal interpolation. Consider the simple example given by Flowerdew and Green (1989) in which the variable of interest is a count variable observed on n source regions. Of interest is an estimate for the count associated with each of several irregularly shaped target zones, each of which overlaps some of the source units (Fig. 2). Flowerdew and Green called these *incompatible zones*. A binary covariate is measured on each target zone, and the intensity of the count process is assumed to be either λ_1 or λ_2, depending on whether the value of the covariate is 1 or 2. Flowerdew and Green (1989) estimated the parameters λ_1 and λ_2 by regressing (using Poisson regression) the count for each source unit, Y_i, on A_{1i} and A_{2i}, where A_{1i} is the area of source unit i overlapped by a target unit with a covariate value of 1 and A_{2i} is defined similarly. The estimated count for any particular subunit formed by division of a source unit by the boundaries of the target units was then taken to be the area of the subunit times the estimated λ_i associated with the covariate value i for that unit. To force the estimates to satisfy the pycnophylactic property, the source zone estimates were scaled by a factor equal to the ratio of the observed value to the fitted value. Estimated counts for each target unit were then obtained by adding the estimates of each subunit composing the target unit.

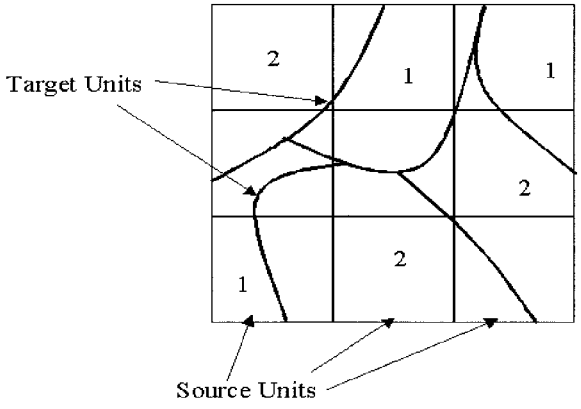

Figure 2. Incompatible Zones. (Adapted from Flowerdew and Green 1989.)

Flowerdew and Green (1992, 1994) extended their earlier work to include several different types of response variables (e.g., binomial, Gaussian) and continuous covariate information. The EM algorithm was used to estimate the necessary parameters, but any software implementing generalized linear model (GLM) methodology may be adapted to this problem. In particular, GLM software that allows the areal data to be spatially correlated (not considered in Flowerdew and Green's work) might be used to improve the results. If the estimates of λ_1 and λ_2 are substantially different and the relationship between the count variable and the binary covariate is strong, then this approach will be more accurate than the traditional polygonal overlay method.

5.5 Bayesian Areal Regression Models

Mugglin and Carlin (1998) extended the work of Flowerdew and Green using a Bayesian approach to areal interpolation. They initially considered the case in which an outcome variable of interest, Y, was measured on $i = 1, 2$ regions (the source units) and prediction of this variable for region 3 (the target unit) was desired. A binary covariate, X, was measured on j subunits, $j = 1, \ldots N_i$, completely nested within both target and source units (Fig. 3).

To develop a predictor of Y_3, Mugglin and Carlin (1998) assumed the existence of latent count variables defined on each subregion. These were assigned independent Poisson distributions with parameters λ_1 or λ_2, depending on the value of the covariate in the subregion. Then the count in each source zone was considered to be the sum of the counts in the subregions composing the source zone. So, following Flowerdew and Green (1989),

$$Y_i \mid \lambda_1, \lambda_2 \sim \text{Poisson}(\lambda_1 A_{1i} + \lambda_2 A_{2i}),$$

where A_{1i} is the area of region i with a covariate value of 1 and A_{2i} is defined similarly. But instead of estimating λ_1 and λ_2 using Poisson regression, Mugglin and Carlin (1998) developed a Bayesian hierarchical model. Vague, independent gamma$(2, 60)$ prior distributions were chosen for λ_1 and λ_2 so that the prior mean corresponded to the expected average count of each subregion, and Metropolis updating was used to sample from the joint posterior distribution and make inferences on λ_1 and λ_2. Mugglin and Carlin (1998) also recognized that *prediction* of Y_3 (the count in the target zone), not estimation, was the appropriate inferential tool for this problem. In obtaining the predictive distribution of $Y_3 \mid Y_1, Y_2$, they showed that the count variables $Y_{3\cap1}$ and $Y_{3\cap2}$ (delineating the parts of Y_3 that also lie in regions 1 and 2) were both binomial variables that could not exceed Y_1 and Y_2. This ensured Tobler's pycnophylactic property. Additional details and supporting theorems were given by Mugglin and Carlin (1998). They then applied these ideas to a more general regression situation based on leukemia counts in Tompkins County, New York. Disease counts were available at the census tract level, and prediction of disease counts at the census block group level was of interest. Demographic factors, including population totals, and distance of the block group centroid from an hazardous waste site, were used as covariates for the predictions. Mugglin, Carlin, Zhu, and Conlon (1999) extended these ideas to smoothing models that allow for spatial heterogeneity and clustering, and Mugglin, Carlin, and Gelfand (2000) considered a more general misalignment problem. This latter work considered two misaligned grids, B and C, that have different spatial supports. The B cells were the source units, and a response variable of interest, Y_i, was observed for each B_i. Covariates, X_j, were observed on each target unit C_j. The misalignment of the B's and C's prevents the use of standard regression methods for inference on Y. To solve this misalignment problem, Mugglin et al. (2000) developed a Bayesian hierarchical model by conditioning on latent variables associated with smaller areas (called *atoms*) delineated by the intersection of the two grids. The latent variables were assumed to have conditionally independent Poisson distributions, the means of which were taken to be functions of the areas of each atom and random effects that accounted for grid cell effects. In addition, a spatial Markov random field prior specification was introduced to model the spatial association among the Y_i's. This latter extension overcame one of the main criticisms of the methods described in Sections 5.3 and 5.4. Although the details of the model building and MCMC implementation are complex, the underlying idea was to build a model at the atom level, thus producing a common grid on which X and Y could be related. Mugglin et al. (2000) illustrated a more general model that included a covariate, W, associated with each B cell, and noted that both X and W could be vectors of covariates and not simply scalars. These perhaps could be used to account for large-scale spatial nonstationarity in the Y process. Also, different distributions for the aggregated count measurements could be used, although the user would need to make the necessary theoretical derivations and confront the implementational issues not addressed by Mugglin et al. (2000).

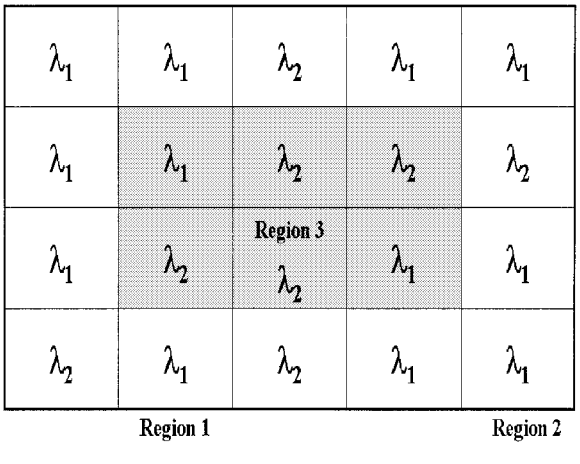

Figure 3. The Nested Misalignment Problem. (Adapted from Mugglin and Carlin 1998.)

As noted earlier, the main cause of bias in ecological inference is the loss of relevant information on the individuals composing each target and source zone. Flowerdew and Green (1989) and Mugglin and Carlin (1998) recovered some of this information by including relevant covariate information available on some of the source zones. Best, Ickstadt, and Wolpert (2000) considered a slightly different approach using a marked point process that directly allowed the use of individual-level

outcomes (disease cases) and covariate information (e.g., race, sex) instead of areal data to make the desired individual-level inferences. They recommended the use of Poisson regression with an idenity link, as opposed to the traditional logarithmic link function, to make the analysis scale independent (i.e., the regression parameter is the same for both individual and areal data) and reduce ecological bias. To solve the problem of incompatible spatial data (disease prevalence data and air pollution measurements from monitors), Best et al. (2000) assigned the case data to the centroid of the postal code of the home address of the case and estimated the air pollution concentration at this centroid from point measurements obtained from air pollution monitors. A Bayesian hierarchical point process model using the Poisson/gamma model developed by Wolpert and Ickstadt (1998) was then used to make inferences about the relationship between disease and pollution at the postal centroids. This is similar to a solution suggested in Section 3.1, although the approach of Best et al. (2000) did not account for locational errors in assigning the cases to the centroids or for uncertainty in the estimated air pollution measurements.

These models illustrate the power and flexibility of Bayesian hierarchical models in map overlay problems. However, their utility comes from several key assumptions that are largely unverifiable from the data. The creation of latent variables defined on the atoms, together with the prior specifications and the sampling methods used for inference with MCMC methods, effectively amounts to assumed knowledge of the distributions of both X and Y variables at a common level of support. As with the multiscale tree models described in Section 4.2, implicit assumptions are made about the cross-correlation between the two variables that are difficult to verify. The use of proportional allocation in modeling conditional means does not account for the shape of the units needed for complete solutions to COSPs. Future refinements to existing solutions to the spatial misalignment problem may be able to address these issues.

6. DISCUSSION

This article has presented a comprehensive review of statistical methods for combining incompatible spatial data. The multidisciplinary history provided conveys the complexity of the problems encountered in combining disparate spatial data and the widespread interest in solutions that have been developing over many decades. This is now an active area of statistical research, and many powerful and novel methods have recently been developed.

Although the problems that arise in combining incompatible spatial data have been given many different names, they all can be considered COSPs. Spatial support is much more than the area or volume associated with the data; it also includes the shape and orientation of the spatial units being considered. The central issue in COSPs is determination of the relationships between data at various scales or levels of aggregation. A common solution strategy for COSPs is to build a model from point support data or from data with small areal support, (even if no observations are taken at this level of support) and then find a way to (optimally) estimate important parameters and make valid inference. Of course, it is the latter part of this paradigm that makes viable solutions so difficult, and some assumptions must be made to obtain any solution. Thus concern shifts to the validity of these assumptions. Traditional geostatistical solutions cleanly delineate complex covariance structures, but then often rely on contrived parametric models to describe them. A hierarchical specification circumvents this problem by allowing the scientist to break down a complex problem into more tractable pieces. Are these pieces really more tractable, or are they more tractable only because we are somehow more willing to make simplifying assumptions conditionally, hoping that the structure induced through the various hierarchies will be sufficiently complex to provide a satisfactory model? Ecologists have adopted the approach of trying to explain how variables at one scale change as they go to another scale. Statisticians assume that they have this knowledge either through the choice of parametric covariance models or through a hierarchical specification and prior information. Much progress has been made on combining incompatible spatial data, but the assumptions made when doing so often ignore the differing supports. Thus these assumptions are much more than mathematical and computational assumptions, and they can (often surreptitiously) result in unusual or unrealistic cross-scale relationships. Further development of data-driven diagnostic and validation tools will help assess the impact and validity of many of these assumptions.

Increasing use of GIS systems makes the COSP relevant and solutions crucial. Substantial progress has been made in the development of valid methods for combining incompatible spatial data, much of it only recently. Hopefully, these methods will be quickly incorporated into an interactive GIS framework that will help ensure that valid statistical methods are used for spatial analysis.

[Received September 2000. Revised August 2001.]

REFERENCES

Aldworth, J., and Cressie, N. (1999), "Sampling Designs and Prediction Methods for Gaussian Spatial Processes," in *Multivariate Analyses, Design of Experiments, and Survey Sampling*, ed. S. Ghosh, New York: Marcel Dekker, pp. 1–54.

——— (2002), "Prediction of Nonlinear Spatial Functionals," *Journal of Statistical Planning and Inference*, In Press.

Amrhein, C. G., and Flowerdew, R. (1992), "The Effect of Data Aggregation on a Poisson Regression Model of Canadian Migration," *Environment and Planning*, Ser. A, 24, 1381–1391.

Arbia, G. (1986), "The Modifiable Areal Unit Problem and the Spatial Autocorrelation Problem: Towards a Joint Approach," *Metron*, 44, 391–407.

——— (1989), "Statistical Effect of Data Transformations: A Proposed General Framework," in *The Accuracy of Spatial Data Bases*, eds. M. Goodchild and S. Gopal, London: Taylor and Francis, pp. 249–259.

Armstrong, M. (1999), *Basic Linear Geostatistics*, New York: Springer-Verlag.

Basseville, M., Benveniste, A., Chou, K. C., Golden, S. A., Nikoukhah, R., and Willsky, A. S. (1992), "Modeling and Estimation of Multiresolution Stochastic Processes," *IEEE Transactions on Information Theory*, 38, 766–784.

Batty, M. (1976), "Entropy in Spatial Aggregation," *Geographical Analysis*, 8, 1–21.

Bell, G., Lechowitz, M. J., Appenzeller, A., Chandler, M., DeBlois, E., Jackson, L., Mackenzie, B., Preziosi, R., Scallenberg, M., and Tinker, N. (1993), "The Spatial Structure of the Physical Environment," *Oecologia*, 96, 114–121.

Bellehumeur, C., and Legendre, P. (1998), "Multiscale Sources of Variation in Ecological Variables: Modeling Spatial Dispersion, Elaborating Sampling Designs," *Landscape Ecology*, 13, 15–25.

Berliner, L. M., Royle, J. A., Wikle, C. K., and Milliff, R. F. (1999), "Bayesian Methods in Atmospheric Sciences" (with discussion), in *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, New York: Oxford University Press, pp. 83–100.

Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000), "Spatial Poisson Regression for Health and Exposure Data Measured at Disparate Resolutions," *Journal of the American Statistical Association*, 95, 1076–1088.

Bian, L., and Butler, R. (1999), "Comparing Effects of Aggregation Methods on Statistical and Spatial Properties of Simulated Spatial Data," *Photogrammetric Engineering & Remote Sensing*, 65, 73–84.

Bissonette, J. A. (1997), "Scale-Sensitive Ecological Properties: Historical Context, Current Meaning," in *Wildlife and Lanscape Ecology: Effects of Pattern and Scale*, ed. J. A. Bissonette, New York: Springer-Verlag, pp. 3–31.

Blalock, H. M. (1964), *Causal Inferences in Nonexperimental Research*, Chapel Hill, NC: University of North Carolina Press.

Bonham-Carter, G. F. (1994), *Geographic Information Systems for Geoscientists*, New York: Pergamon Press/Elsevier Science.

Bonham-Carter, G. F., Agterberg, F. P., and Wright, D. F. (1988), "Integration of Geological Datasets for Gold Exploration in Nova Scotia," *Photogrammetric Engineering & Remote Sensing*, 54, 1585–1592.

Brillinger, D. R. (1990), "Spatial-Temporal Modeling of Spatially Aggregate Birth Data," *Survey Methodology*, 16, 255–269.

——— (1994), "Examples of Scientific Problems and Data Analyses in Demography, Neurophysiology, and Seismology," *Journal of Computational and Graphical Statistics*, 3, 1–22.

Brown, R. G. (1983), *Random Signal Analysis and Kalman Filtering* (rev. ed.), New York: Wiley.

Carroll, S. S., Day, G. N., Cressie, N., and Carroll, T. R. (1995), "Spatial Modeling of Snow Water Equivalent Using Airborne and Ground-Based Snow Data," *Environmetrics*, 6, 127–139.

Chiles, J. P., and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, New York: Wiley.

Chou, K. C., Willsky, A. S., and Nikoukhah, R. (1994), "Multiscale Systems, Kalman Filters, and Riccati Equations," *IEEE Transactions on Automatic Control*, 39, 479–492.

Chung, C.-J. F., and Fabbri, A. G. (1999), "Probabilistic Prediction Models for Landslide Hazard Mapping," *Photogrammetric Engineering & Remote Sensing*, 65, 1389–1400.

Clark, I. (1979), *Practical Geostatistics*, Essex, U.K.: Applied Science.

Clark, I., Basinger, K. L., and Harper, W. V. (1989), "MUCK: A Novel Approach to Cokriging," in *Proceedings of the Conference on Geostatistical, Sensitivity, and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modeling*, ed. B. Buxton, Columbus, OH: Battelle Press, pp. 473–493.

Clark, W. A. V., and Avery, K. L. (1976), "The Effects of Data Aggregation in Statistical Analysis," *Geographical Analysis*, 8, 428–438.

Cleave, N., Brown, P. J., and Payne, C. D. (1995), "Evaluation of Methods for Ecological Inference," *Journal of the Royal Statistical Society*, Ser. B, 159, 55–72.

Cramer, J. S. (1964), "Efficient Grouping, Regression and Correlation in Engel Curve Analysis," *Journal of the American Statistical Association*, 59, 233–250.

Cressie, N. (1993a), "Aggregation in Geostatistical Problems," in *Geostatistics Troia '92*, ed. A. Soares, Dordrecht: Kluwer Academic, pp. 25–35.

——— (1993b), *Statistics for Spatial Data*, New York: Wiley.

——— (1996), "Change of Support and the Modifiable Areal Unit Problem," *Geographical Systems*, 3, 159–180.

——— (1998), "Aggregation and Interaction Issues in Statistical Modeling of Spatiotemporal Processes," *Geoderma*, 85, 133–140.

Cressie, N., and Wikle, C. K. (1998), "The Variance-Based Cross-Variogram: You Can Add Apples and Oranges," *Mathematical Geology*, 30, 789–799.

Daley, R. (1991), "*Atmospheric Data Analysis*," Cambridge, U.K.: Cambridge University Press.

——— (1992), "Estimating Model-Error Covariances for Application to Atmospheric Data Assimilation," *Monthly Weather Review*, 120, 1735–1746.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-Based Geostatistics" (with discussion), *Applied Statistics*, 47, 299–350.

Dyn, N., Wahba, G., and Wong, W.-H. (1979), discussion of "Smooth Pycnophylactic Interpolation for Geographical Regions" by W. Tobler, *Journal of the American Statistical Association*, 74, 530–535.

Emerson, C. W., Lam, N. S.-N., and Quattrochi, D. A. (1999), "Multi-Scale Fractal Analysis of Image Texture and Pattern," *Photogrammetric Engineering & Remote Sensing*, 65, 51–61.

Fairfield Smith, H. (1938), "An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops," *Journal of Agricultural Science*, 28, 1–23.

Fieguth, P. W., Karl, W. C., Willsky, A. S., and Wunsch, C. (1995), "Multiresolution Optimal Interpolation and Statistical Analysis of TOPEX/POSEIDON Satellite Altimetry," *IEEE Transactions on Geoscience and Remote Sensing*, 33, 280–292.

Firebaugh, G. (1978), "A Rule for Inferring Individual-Level Relationships From Aggregate Data," *American Sociological Review*, 43, 557–572.

Flowerdew, R., and Green, M. (1989), "Statistical Methods for Inference Between Incompatible Zonal Systems," in *The Accuracy of Spatial Data Bases*, eds. M. Goodchild and S. Gopal, London: Taylor and Francis, pp. 239–247.

——— (1992), "Developments in Areal Interpolating Methods and GIS," *Annals of Regional Science*, 26, 67–78.

——— (1994), "Areal Interpolation and Types of Data," in *Spatial Analysis and GIS*, eds. S. Fotheringham and P. Rogerson, London: Taylor and Francis, pp. 121–145.

Fotheringham, A. S. (1989), "Scale-Independent Spatial Analysis," in *The Accuracy of Spatial Data Bases*, eds. M. Goodchild and S. Gopal, London: Taylor and Francis, pp. 221–228.

Fotheringham, A. S., and Wong, D. W. S. (1991), "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis," *Environment and Planning*, Ser. A, 23, 1025–1044.

Gabrosek, J., Huang, H.-C., and Cressie, N. (1999), "Spatio-Temporal Prediction of Level 3 Data for NASA's Earth Observing System," in *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*, eds. K. Lowell and A. Jaton, Chelsea, MI: Ann Arbor Press.

Gaudard, M., Karson, M., Linder, E., and Sinha, D. (1999), "Bayesian Spatial Prediction," *Environmental and Ecological Statistics*, 6, 147–171.

Gehlke, C. E., and Biehl, K. (1934), "Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material," *Journal of the American Statistical Association*, supp 29, 169–170.

Gelfand, A. E., Zhu, L., and Carlin, B. P. (2001), "On the Change of Support Problem for Spatio-Temporal Data," *Biostatistics*, 2, 31–45.

Ghil, M., Cohn, S., Tavantzia, J., Bube, K., and Isaacson, E. (1981), "Applications of Estimation Theory to Numerical Prediction Assimilation," in *Dynamic Meteorology: Data Assimilation Methods*, eds. L. Bengtsson, M. Ghiul, and E. Kallen, London: Springer-Verlag, pp. 139–224.

Goldstein, H. (1995), "Multilevel Statistical Models," (2nd ed.), Kendall's Library of Statistics Vol. 3, London: Edward Arnold.

Goodchild, M. F., and Lam, N. S.-N. (1980), "Areal Interpolation: A Variant of the Traditional Spatial Problem," *Geo-Processing*, 1, 297–312.

Goodman, L. A. (1959), "Some Alternatives to Ecological Correlation," *American Journal of Sociology*, 64, 610–625.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, New York: Oxford University Press.

——— (1998), "Geostatistical Tools for Characterizing the Spatial Variability of Microbiological and Physico-Chemical Soil Properties," *Biology and Fertility of Soils*, 27, 315–334.

Greenland, S., and Morganstern, H. (1989), "Ecological Bias, Confounding, and Effect Modification," *International Journal of Epidemiology*, 18, 269–274.

Greenland, S., and Robins, J. (1994), "Ecologic Studies—Biases, Misconceptions, and Counterexamples," *American Journal of Epidemiology*, 139, 747–760.

Greig-Smith, P. (1952), "The Use of Random and Contiguous Quadrats in the Study of Structure in Plant Communities," *Annals of Botany, New Series*, 16, 293–316.

Haitovsky, Y. (1984), *Regression Estimation From Grouped Observations*, New York: Harner Press.

Handcock, M. G., and Stein, M. L. (1993), "A Bayesian Analysis of Kriging," *Technometrics*, 35, 403–410.

Hannan, M. T. (1971), *Aggregation and Disaggregation in Sociology*, Lexington, MA: D.C. Heath.

Huang, H.-C., and Cressie, N. (2000), "Multiscale Graphical Modeling in Space: Applications to Command and Control," in *Proceedings of the Spatial Statistics Workshop*, ed. M. Moore, New York: Springer-Verlag, pp. —.

Huang, H.-C., Cressie, N., and Gabrosek, J. (2000), "Fast Resolution-Consistent Spatial Prediction of Global Processes From Satellite Data," Technical Report C-2000-2, Institute of Statistical Science, Academia Sinica, available at *http://www.stat.sinica.edu.tw/hchuang/paper.html*.

Isaaks, E. H., and Srivastava, R. M. (1989), *An Introduction to Applied Geostatistics*, New York: Oxford University Press.

Jelinski, D. E., and Wu, J. (1996), "The Modifiable Areal Unit Problem and Implications for Landscape Ecology," *Landscape Ecology*, 11, 129–140.

Johnson, G. D., and Patil, G. P. (1998), "Quantitative Multiresolution Characterization of Landscape Patterns for Assessing the Status of Ecosystem Health in Watershed Management Areas," *Ecosystem Health*, 4, 177–187.

Journel, A. G. (1983), "Nonparametric Estimation of Spatial Distributions," *Journal of the International Association for Mathematical Geology*, 15, 445–468.

Journel, A. G., and Huijbregts, C. J. (1978), *Mining Geostatistics*, London: Academic Press.

King, G. (1997), *A Solution to the Ecological Inference Problem*, Princeton, NJ: Princeton University Press.
Kitanidis, P. K. (1986), "Parameter Uncertainty in Estimation of Spatial Functions: Bayesian Analysis," *Water Resources Research*, 22, 499–507.
Klein, S. P., and Freedman, D. A. (1993), "Ecological Regression in Voting Rights Cases," *Chance*, 6, 38–43.
Krige, D. G. (1951), "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand," *Journal of Chemical, Metallurgical, and Mining Society of South Africa*, 52, 119–139.
Le, N. D., and Zidek, J. V. (1992), "Interpolation With Uncertain Spatial Covariances: A Bayesian Alternative to Kriging," *Journal of Multivariate Analysis*, 43, 351–374.
Legendre, P., and Fortin, M. (1989), "Spatial Pattern and Ecological Analysis," *Vegetation*, 80, 107–138.
Levin, S. A. (1992), "The Problem of Pattern and Scale in Ecology," *Ecology*, 73, 1943–1967.
Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996), *The SAS System for Linear Models*, Cary, NC: SAS Institute.
Markoff, J., and Shapiro, G. (1973), "The Linkage of Data Describing Overlapping Geographical Units," *Historical Methods Newsletter*, 7, 34–46.
Matheron, G. (1963), "Principles of Geostatistics," *Economic Geology*, 58, 1246–1266.
——— (1984), "Isofactorial Models and Change of Support," in *Geostatistics for Natural Resources Characterization*, eds. G. Verly, M. David, A. Journel, and A. Marechal, Dordrecht: Reidel, pp. 449–467.
May, R. M. (1994), "The Effects of Spatial Scale on Ecological Questions and Answers," in *Large-Scale Ecology and Conservation Biology (35th Symposium of the British Ecological Society of Conservation Biology)*, eds. P. Edwards, R. May, and N. Webb, London: Blackwell Scientific, pp. 1–17.
McNeill, L. (1991), "Interpolation and Smoothing of Binomial Data for the Southern African Bird Atlas Project," *South African Journal of Statistics*, 25, 129–136.
Meinhold, J., and Singpurwalla, N. D. (1983), "Understanding the Kalman Filter," *The American Statistician*, 37, 123–127.
Milne, B. T. (1988), "Measuring the Fractal Dimension of Landscapes," *Applied Mathematical Computing*, 27, 67–79.
Moellering, H., and Tobler, W. (1972), "Geographical variances," *Geographical Analysis*, 4, 34–50.
Morganstern, H. (1982), "Uses of Ecologic Analysis in Epidemiologic Research," *American Journal of Public Health*, 72, 1336–1344.
Mugglin, A. S., and Carlin, B. P. (1998), "Hierarchical Modeling in Geographic Information Systems: Population Interpolation Over Incompatible Zones," *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 117–130.
Mugglin, A. S., Carlin, B. P., and Gelfand, A. E. (2000), "Fully Model-Based Approaches for Spatially Misaligned Data," *Journal of the American Statistical Association*, 95, 877–887.
Mugglin, A. S., Carlin, B. P., Zhu, L., and Conlon, E. (1999), "Bayesian Areal Interpolation, Estimation, and Smoothing: An Inferential Approach for Geographic Information Systems," *Environment and Planning*, Ser. A, 31, 1337–1352.
Müller, H.-G., Stadtmüller, U., and Tabnak, F. (1997), "Spatial Smoothing of Geographically Aggregated Data, With Application to the Construction of Incidence Maps," *Journal of the American Statistical Association*, 92, 61–71.
Myers, D. E. (1984), "Cokriging—New Developments," in *Geostatistics for Natural Resources Characterization*, eds. G. Verly, M. David, A. G. Journel, and A. Marechal, Dordrecht: Reidel, pp. 295–305.
——— (1991), "Pseudo–Cross-Variograms, Positive Definiteness, and Cokriging," *Mathematical Geology*, 23, 805–816.
Nielsen, D. R., Wendroth, O., and Parlange, M. B. (1995), "Opportunities for Examining On-Farm Soil Variability," in *Site-Specific Management for Agricultural Systems*, eds. P. Robert, R. H. Rust, and W. E. Larson, Madison, WI: American Society of Agronomy, pp. 95–132.
O'Hagan, A. (1994), *Bayesian Inference*, London: Edward Arnold.
Olea, R. A. (1991), *Geostatistical Glossary and Multilingual Dictionary*, New York: Oxford University Press.
Oliver, M. A., Lajaunie, C., Webster, R., Muir, K. R., and Mann, J. R. (1993), "Estimating the Risk of Childhood Cancer," in *Geostatistics Troia '92*, ed. A. Soares, Dordrecht: Kluwer Academic Publishers, pp. 899–910.
Oliver, M. A., and Webster, R. (1986), "Combining Nested and Linear Sampling for Determining the Scale and Form of Spatial Variation of Regionalized Variables," *Geographical Analysis*, 18, 227–242.
Openshaw, S. (1984), *The Modifiable Areal Unit Problem*, Norwich, U.K.: Geobooks.
Openshaw, S., and Taylor, P. (1979), "A Million or so Correlation Coefficients," in *Statistical Methods in the Spatial Sciences*, ed. N. Wrigley, London: Pion, pp. 127–144.
——— (1981), "The Modifiable Areal Unit Problem," in *Quantitative Geography: A British View*, eds. N. Wrigley and R. Bennett, London: Routledge & Kegan Paul, pp. 60–69.
Palmer, M. W. (1988), "Fractal Geometry: A Tool for Describing Spatial Patterns of Plant Communities," *Vegetation*, 75, 91–102.
Papritz, A., Kunsch, H. R., and Webster, R. (1993), "On the Pseudo-Cross-Variogram," *Mathematical Geology*, 25, 1015–1026.
Patil, G. P., and Taillie, C. (1999), "Quantitative Characterization of Hierarchically Scaled Landscape Patterns," in *Proceedings of the 25th Biennial Meeting of the International Statistical Institute*, Vol. 58, Helsinki: Statistics Finland, pp. 89–92.
Phipps, M. (1991), "Entropy and Community Pattern Analysis," *Journal of Theoretical Biology*, 93, 253–273.
Piantadosi, S., Byar, D. P., and Green, S. B. (1988), "The Ecological Fallacy," *American Journal of Epidemiology*, 127, 893–904.
Prais, S. J., and Aitchison, J. (1954), "The Grouping of Observations in Regression Analysis," *Revue de l'Institut International de Statistique*, 1, 1–22.
Putnam, S. H., and Chung, S.-H. (1989), "Effects of Spatial System on Spatial Interaction Models. 1: The Spatial System Definition Problem," *Environment and Planning*, Ser. A, 21, 27–46.
Raines, G. L., Bonham-Carter, G. F., and Kemp, L. (2000), "Predictive Probabilistic Modeling Using ArcView GIS," *ArcUser*, April-June, 46–48.
Renshaw, E., and Ford, E. D. (1984), "The Description of Spatial Pattern Using Two-Dimensional Spectral Analysis," *Vegetation*, 56, 75–85.
Richardson, S. (1992), "Statistical Methods for Geographical Correlation Studies," in *Geographical and Environmental Epidemiology: Methods for Small Area Studies*, eds. P. Elliott, J. Cuzick, D. English, and R. Stern, New York: Oxford University Press, pp. 181–204.
Richardson, S., Stucker, I., and Hemon, D. (1987), "Comparison of Relative Risks Obtained in Ecological and Individual Studies: Some Methodological Considerations," *International Journal of Epidemiology*, 16, 111–120.
Rivoirard, J. (1994), *Introduction to Disjunctive Kriging and Nonlinear Geostatistics*, Oxford, U.K.: Clarendon Press.
Robinson, A. H. (1956), "The Necessity of Weighting Values in Correlation Analysis of Areal Data," *Annals of the Association of American Geographers*, 46, 233–236.
Robinson, W. S. (1950), "Ecological Correlations and the Behavior of Individuals," *American Sociological Review*, 15, 351–357.
Royle, J. A., Berliner, L. M., Wikle, C. K., and Milliff, R. (1997), "A Hierarchical Spatial Model for Constructing Wind Fields From Scatterometer Data in the Labrador Sea," Technical Report 97-30, Ohio State University.
Selvin, H. C. (1958), "Durkheim's 'Suicide' and Problems of Empirical Research," *American Journal of Sociology*, 63, 607–619.
Steel, D. G., and Holt, D. (1996), "Analyzing and Adjusting Aggregation Effects: The Ecological Fallacy Revisited," *International Statistical Review*, 64, 39–60.
Sugihara, G., and May, R. M. (1990), "Application of Fractals in Ecology," *Trends in Ecology and Evolution*, 5, 79–86.
Theil, H. (1972), *Econometrics and Information Theory*, Amsterdam: North-Holland.
Thiessen, A. H. (1911), "Precipitation for Large Areas," *Monthly Weather Review*, 39, 1082–1084.
Thomas, E. N., and Anderson, D. L. (1965), "Additional Comments on Weighting Values in Correlation Analysis of Areal Data," *Annals of the Association of American Geographers*, 55, 492–505.
Tobler, W. (1979), "Smooth Pycnophylactic Interpolation for Geographical Regions" (with discussion), *Journal of the American Statistical Association*, 74, 519–536.
——— (1989), "Frame Independent Spatial Analysis," in *The Accuracy of Spatial Data Bases*, eds. M. Goodchild and S. Gopal, London: Taylor and Francis, pp. 115–122.
Turner, M. G., Dale, V. H., and Gardner, R. H. (1989), "Predicting Across Scales: Theory Development and Testing," *Landscape Ecology*, 3, 245–252.
Turner, M. G., O'Neill, R. V., Gardner, R. H., and Milne, B. T. (1989), "Effects of Changing Spatial Scale on the Analysis of Landscape Pattern," *Landscape Ecology*, 3, 153–162.
Ver Hoef, J. M., and Barry, R. M. (1998), "Constructing and Fitting Models for Cokriging and Multivariate Spatial Prediction," *Journal of Statistical Planning and Inference*, 69, 275–294.
Ver Hoef, J. M., and Cressie, N. (1993), "Spatial Statistics: Analysis of Field Experiments," in *Design and Analysis of Ecological Experiments*, eds. S. M. Scheiner and J. Gurevitch, New York: Chapman and Hall, pp. 319–341.
Ver Hoef, J. M., Cressie, N. A. C., and Glenn-Lewin, D. C. (1993), "Spatial Models for Spatial Statistics: Some Unification," *Journal of Vegetation Science*, 4, 441–452.

Verly, G. (1983), "The Multi-Gaussian Approach and Its Applications to the Estimation of Local Reserves," *Journal of the International Association for Mathematical Geology*, 15, 259–286.

Wackernagel, H. (1995), "*Multivariate Geostatistics—An Introduction With Applications*, Berlin: Springer Verlag.

Walter, S. D. (1991), "The Ecologic Method in the Study of Environmental Health II: Methodologic Issues and Feasibility," *Environmental Health Perspectives*, 94, 67–73.

Webster, R., Oliver, M. A., Muir, K. R., and Mann, J. R. (1994), "Kriging the Local Risk of a Rare Disease From a Register of Diagnoses," *Geographical Analysis*, 26, 168–185.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998), "Hierarchical Bayesian Space-Time Models," *Environmental and Ecological Statistics*, 5, 117–154.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001), "Spatio-Temporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds," *Journal of the American Statistical Association*, 96, 382–397.

Williams, I. N. (1976), "Optimistic Theory Validation From Spatially Grouped Regression," *Transactions of the Martin Centre for Architectural and Urban Studies*, 1, 113–145.

Wolpert, R. L., and Ickstadt, K. (1998), "Poisson/Gamma Random Field Models for Spatial Statistics," *Biometrika*, 85, 251–267.

Wong, D. W. S. (1996), "Aggregation Effects in Geo-Referenced Data," in *Advanced Spatial Statistics*, ed. D. Griffiths, Boca Raton, FL: CRC Press, pp. 83–106.

Woodcock, C. E., and Strahler, A. H. (1987), "The Factor of Scale in Remote Sensing," *Remote Sensing and the Environment*, 21, 311–332.

Yule, G. U., and Kendall, M. G. (1950), *An Introduction to the Theory of Statistics* (14th ed.), London: Griffin.