A Geostatistical Approach to Linking Geographically Aggregated Data From Different Sources

Carol A. GOTWAY and Linda J. YOUNG

The widespread availability of digital spatial data and the capabilities of Geographic Information Systems (GIS) make it possible to easily synthesize spatial data from a variety of sources. More often than not, data have been collected at different geographic scales, and each of the scales may be different from the one of interest. Geographic information systems effortlessly handle these types of problems through raster and geoprocessing operations based on proportional allocation and centroid smoothing techniques. However, these techniques do not provide a measure of uncertainty in the estimates and lack the ability to incorporate important covariate information that may be used to improve the estimates. They also often ignore the different spatial supports (e.g., shape and orientation) of the data. On the other hand, statistical solutions to change-of-support problems are rather specific and difficult to implement. In this article, we present a general geostatistical framework for linking geographic data from different sources. This framework incorporates aggregation and disaggregation of spatial data, as well as prediction problems involving overlapping geographic units. It explicitly incorporates the supports of the data, can adjust for covariate values measured on different spatial units at different scales, provides a measure of uncertainty for the resulting predictions, and is computationally feasible within a GIS. The new framework we develop also includes a new approach for simultaneous estimation of mean and covariance functions from aggregated data using generalized estimating equations.

Key Words: Change-of-support; Disaggregation; Ecological inference; Generalized estimating equations; Modifiable areal unit problem.

1. INTRODUCTION

Many users of geographical data need to supplement their data with other important information collected on different geographical units. This often results in the need to combine data from geographical units at the same or similar scales. For example, crimes are often reported by police precinct or ZIP code, but population information needed to provide

Carol A. Gotway is Mathematical Statistician, National Center for Environmental Health, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333 (E-mail: *cdg7@cdc.gov*). Linda Young is Professor, Department of Statistics, University of Florida, Gainesville, FL.

Journal of Computational and Graphical Statistics, Volume 16, Number 1, Pages 115–135
DOI: 10.1198/106186007X179257

a suitable baseline measure for analysis and interpretation is aggregated over Census tracts. In other cases, we might have data at one scale (e.g., disease counts for each county), but we would like to use more specific covariates available at a smaller scale (e.g., Census tracts), or individual-level inference may be desired, but only aggregate data are available (e.g., voter characteristics compiled by voting districts). Even when detailed individual-level data are available, information on a suitable comparison group for spatial analysis can be limited. For example, cancer registries record information about individuals with cancer and increasingly this information includes geocoded address information (e.g., the longitude and latitude of each patient's address). However, when studying a particular type of cancer, the choice of a suitable control group from a registry database may be unclear, or the cost or effort in obtaining geocoded addresses for the controls is too large. In such cases, the disease data are often aggregated to Census block groups or tracts where population data are available. Finally, when linking health data to environmental data, we may need to combine geographically aggregated health outcome data with measurements from environmental monitors at point locations or with remotely sensed data at a variety of resolutions.

Many different statistical approaches have been proposed for combining incompatible spatial data. In reviewing these, Gotway and Young (2002) found that many approaches focus on just one of the scenarios described above. For example, Flowerdew and Green (1992, 1994) and Mugglin and Carlin (1998) considered the problem of disaggregating count data based on one set of areal units to provide predictions on another set of areal units. Methods here tend to center around the idea of first predicting data with point support or data supported by the intersection of two or more sets of spatial units (called "atoms" by Mugglin, Carlin, and Gelfand 2000) and then adding the point or atom-level predictions to produce predictions for larger areal units. However, this approach, and other approaches based on centroid smoothing (e.g., Bracken and Martin 1989), assume the data are independent and do not account for the differing supports of the spatial units involved. Some approaches to smoothing aggregated data do account for differing supports (Tobler 1979; Brillinger 1990, 1994; Müller, Stadtmüller, and Tabnak 1997) and the models used by Kelsall and Wakefield (2002) account for spatial dependence as well. However, all of these approaches are concerned solely with obtaining a smooth risk or density surface of point support from aggregated data. On the other hand, most geostatistical approaches to problems associated with combining incompatible spatial data have been primarily concerned with upscaling or aggregation from data with point support (e.g., block kriging) or focus on downscaling environmental variables summarized by averages, rather than on disaggregating regional counts (Huang, Cressie, and Gabrosek 2002; Wikle, Milliff, Nychka, and Berliner 2001; and Gelfand, Zhu, and Carlin 2001). Many of these consider only rectangular spatial units such as pixels from satellite data or grid blocks from deterministic process models. Most statistical approaches also assume a great deal of prior information (e.g., counts arise from a Poisson process, the data are multivariate Gaussian, knowledge of hierarchical formulation and cross-scale relationships, hyperprior distributions). Moreover, since these approaches are difficult to implement and automate, proportional allocation remains the commonly used approach in most other disciplines and it is often implemented seamlessly as part of the geoprocessing functions (e.g., dissolve, union, intersection) and raster calculations (e.g.,

with gridded data) within a geographic information system (GIS). We seek a compromise between these two extremes, one that provides a unified framework for the different types of spatial problems, is based on fewer assumptions, and is computationally feasible within a commercial GIS using current GIS technology and geoprocessing operations, yet also offers an alternative to proportional allocation that can include covariates and provide standard errors for the resulting predictions.

In this article, we develop a new approach for simultaneous estimation of mean and covariance functions from geographically aggregated data. This approach is combined with an adaptation of best linear unbiased prediction to create a general method for combining data collected on different geographical units. It can be used to provide a smooth surface of point support based on aggregated data. It can also be used to make predictions for one set of geographic units from another set of geographic units (either nested or overlapping) and there is no need to first make point or atom-level predictions and then add these back up to larger geographical units of interest. Thus, the new approach we present here can handle all of the situations described above. It also explicitly accounts for the differing supports of all units involved and allows the data associated with these units to be spatially correlated. The predictions satisfy what Tobler (1979) called the pycnophylactic or volume-preserving property for point-level predictions or what Huang et al. (2002) called "consistency in aggregation" or "mass balance" for areal predictions meaning that predictions for subunits within a given unit sum (counts) or average (spatial averages) to the unit total or spatial average. Unlike Tobler's method, covariates can be incorporated into the predictions and a measure of prediction uncertainty can be obtained. In Section 2, we give the development of this new geostatistical approach, beginning with an overview of best linear unbiased prediction often used for upscaling spatial processes in geostatistics. In Sections 2.2 and 2.3 we extend the ideas behind upscaling to more general change of support problems, and in Section 2.4 we show how the predictions from our more general approach maintain consistency in aggregation. Section 2.5 provides a new method for the simultaneous estimation of mean and covariance functions from aggregated data. Section 3 presents two examples combining the ideas in previous sections to disaggregate spatial data and combine spatial data on different geographical units. Section 4 provides a summary of our new approach and some suggestions for extentions and further development.

2. SPATIAL PREDICTION AND CHANGE OF SUPPORT

Consider a spatial process $\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \Re^2\}$, where $Z(\mathbf{s})$ is a random variable at a known location $\mathbf{s}$, and $\mathbf{s}$ varies continuously over a spatial domain D. We assume $Z(\mathbf{s})$ has mean $E(Z(\mathbf{s})) = \mathbf{x}(s)'\boldsymbol{\beta}$, where $\mathbf{x}(\mathbf{s}) = (x_1(\mathbf{s}), x_2(\mathbf{s}), \ldots, x_p(\mathbf{s}))'$ is a p x 1 vector of explanatory variables observed at location $\mathbf{s}$, $\boldsymbol{\beta}$ is a vector of fixed, but unknown parameters and $x_1(\mathbf{s}) \equiv 1$ as in linear regression. We also assume that $Z(\mathbf{s})$ has a stationary covariance function $\text{cov}(Z(\mathbf{u}), Z(\mathbf{v})) = C(\mathbf{u} - \mathbf{v})$ for $\mathbf{u}, \mathbf{v}$ in D.

2.1 UPSCALING WITH BLOCK KRIGING

A common inferential problem in environmental science is the prediction of the spatial average

$$Z(B) = \frac{1}{|B|}\int_B Z(\mathbf{s})d\mathbf{s}, \tag{2.1}$$

from point-support measurements $Z(\mathbf{s}_1), Z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_n)$. The set B is an areal region (in 2-D) or a volume (in 3-D) that forms the spatial *support* of $Z(B)$. In linear geostatistics, the best linear unbiased predictor of $Z(B)$ can be obtained by minimizing mean-squared prediction error subject to unbiasedness constraints. The derivation is analogous to that of kriging for predicting $Z(\mathbf{s}_0)$ at a point location $\mathbf{s}_0$. The resulting predictor is known in the geostatistical literature as the universal block kriging predictor and it is given by $\hat{Z}(B) = \sum_{i=1}^n w_i Z(\mathbf{s}_i)$, where the optimal weights $\{w_i\}$ are obtained by solving (Journel and Huijbregts 1978; Cressie 1993; Chilès and Delfiner 1999)

$$\begin{aligned}\sum_{k=1}^n w_k C(\mathbf{s}_i, \mathbf{s}_k) + \sum_{j=1}^p m_j x_j(\mathbf{s}_i) &= C(B, \mathbf{s}_i) \quad i = 1, \ldots, n \\ \sum_{i=1}^n w_i x_j(\mathbf{s}_i) &= x_j(B) \quad j = 1, \ldots, p.\end{aligned}$$

In these equations, the volume-level covariates satisfy (see Cressie 1996)

$$x_j(B) = \frac{1}{|B|}\int_B x_j(\mathbf{u})d\mathbf{u}, \quad j = 1, 2, \ldots, p, \tag{2.2}$$

the m_j are Lagrange multipliers from the constrained minimization and $C(B, \mathbf{s}_i)$ is the point-to-volume covariance given by

$$C(B, \mathbf{s}_i) = \text{cov}(Z(B), Z(\mathbf{s}_i)) = \int_B C(\mathbf{u}, \mathbf{v})d\mathbf{u}d\mathbf{v}/|B|. \tag{2.3}$$

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 for not only the size, but also the shape and orientation of the volume B.

A primary advantage of predicting $Z(B)$ in this fashion rather than predicting $Z(\mathbf{u}_j)$ for many points $\mathbf{u}_j$ in B and then averaging the resulting predictions lies in the ability to easily obtain a measure of uncertainty of the resulting predictor. The prediction mean-squared error associated with $\hat{Z}(B)$ is given by (Cressie 1993; Chilès and Delfiner 1999)

$$C(B, B) - \sum_{i=1}^n C(B, \mathbf{s}_i) - \sum_{j=1}^p m_j x_j(B),$$

where

$$C(B, B) = \frac{1}{|B|^2}\int_B\int_B C(\mathbf{u} - \mathbf{v})d\mathbf{u}d\mathbf{v}.$$

The point-to-point covariance function, $C(\mathbf{u} - \mathbf{v})$ is assumed known for theoretical derivations, but is then estimated and modeled with a valid positive definite function based on the point-support data. In practice, integrals are computed by discretizing B into points, $\{\mathbf{u}'_j\}$, so that Equation (2.3) is approximated using $C(B, \mathbf{s}_i) \approx 1/N \sum_{j=1}^N C(\mathbf{u}'_j, \mathbf{s}_i)$.

2.2 THE MORE GENERAL PREDICTION PROBLEM

The field of geostatistics has been concerned with change of support issues since its inception. However, the practical problems driving the initial development of geostatistics were those encountered in the mining industry, with a primary problem being the prediction of the average grade of a mining block from drill core samples. Although the change of support problem is a general one that refers to the problem of valid inference of a spatial variable based on one support from data at different supports, most change of support problems, and the solutions to them, have been concerned with "upscaling": the prediction of a variable whose support is larger than that of the observed data. The block kriging predictor is commonly used in upscaling spatial processes and is one solution to the point-to-volume change of support problem.

In this article, we take a broader view of change of support problems and extend the development in Section 2.1 to more general change of support problems. We also provide statistical methods for practical implementation of this new approach that can be used with a variety of spatial variables associated with many different types of spatial supports. In particular, we develop a new approach for the simultaneous estimation of mean and covariance functions from aggregated data.

Suppose that instead of observing a realization of the point-level process, $Z(\mathbf{s})$, we have data $Z(\mathbf{B}) = (Z(B_1), \ldots, Z(B_n))'$ and prediction of $Z(A)$ is of interest. Here the volumes A and B can be general, allowing several different types of change of support problems. For example, if the B_i are points instead of volumes, then this is the point-volume change of support problem described above. If A is a point, then the problem becomes one of spatial disaggregation requiring prediction of a spatial intensity function from aggregated data. In geographical applications, A and B_i may be different geographic regions at the same spatial scale or at different spatial scales. In such cases, A may be wholly contained in one of the B_i, or it may overlap two or more of the B units.

In this context, the best linear unbiased predictor has the form

$$\hat{Z}(A) = \mathbf{w}(A)'Z(\mathbf{B}), \quad \text{with } \mathbf{w}(A)' = (w_1(A), \ldots, w_n(A)), \tag{2.4}$$

where each weight $w_i(A)$ measures the influence of datum $Z(B_i)$ on the prediction of $Z(A)$. The optimal weights are given by

$$\mathbf{w}_u = (\Sigma_u)^{-1}\boldsymbol{\sigma}_u, \tag{2.5}$$

where $\mathbf{w}_u = (w_1(A), \ldots, w_n(A), \mathbf{m}')' \equiv (\mathbf{w}(A)'\,\mathbf{m}')'$ and $\mathbf{m}$ is a $(p \times 1)$ vector of Lagrange multipliers.

The $(n + p) \times (n + p)$ matrix Σ_u is based on the data and covariates on the B-units:

$$\Sigma_u = \begin{bmatrix} \Sigma_{BB} & \mathbf{x}(\mathbf{B}) \\ \mathbf{x}(\mathbf{B})' & \emptyset \end{bmatrix}. \tag{2.6}$$

The elements of the $(n \times n)$ matrix Σ_{BB} are the (B_i, B_j) covariances

$$C(B_i, B_j) = \operatorname{cov}(Z(B_i), Z(B_j)) = \frac{1}{|B_i|}\frac{1}{|B_j|}\int_{B_i}\int_{B_j} C(\mathbf{u} - \mathbf{v})d\mathbf{u}d\mathbf{v}. \tag{2.7}$$

The matrix $\mathbf{x}(\mathbf{B})$ has dimension $(n \times p)$ with ith row equal to $(x_1(B_i), x_2(B_i), \ldots, x_p(B_i))$, with $x_j(B)$ given in Equation (2.2). The $x_i(B)$ are explanatory covariates associated with the B units. These may measure average demographic characteristics, environmental conditions, or other fixed properties associated with the B units.

The matrix $\boldsymbol{\sigma}_u$ is comprised of information on the A units and their spatial relationships to the B-units and is equal to

$$\boldsymbol{\sigma}_u = \begin{bmatrix} \boldsymbol{\sigma}_{AB} \\ \\ \mathbf{x}(A) \end{bmatrix}, \tag{2.8}$$

where the elements of the $(n \times 1)$ vector $\boldsymbol{\sigma}_{AB}$ are the (A, B_i) covariances

$$C(A, B_i) = \text{cov}(Z(A), Z(B_i)) = \frac{1}{|A|}\frac{1}{|B_i|}\int_A \int_{B_i} C(\mathbf{u} - \mathbf{v}) d\mathbf{u} d\mathbf{v}. \tag{2.9}$$

The $(p \times 1)$ vector $\mathbf{x}(A)$ is comprised of the covariate information associated with A, $\mathbf{x}(A) = (x_1(A), x_2(A), \ldots, x_p(A))'$, where $x_j(A)$ satisfies the general relationship in Equation (2.2).

The corresponding prediction mean-squared errors are given by

$$\sigma_{AA} - \mathbf{w}_u' \boldsymbol{\sigma}_u, \tag{2.10}$$

with $\sigma_{AA} = \text{cov}(Z(A), Z(A))$, defined analogously to that in Equation (2.7).

As with traditional "point-to-point" methods of spatial prediction (kriging), measurement error in the data $Z(B_i)$ can be accommodated by assuming

$$Z(B) = \int_B Z(\mathbf{s}) d\mathbf{s} + \epsilon(B)$$

and using a filtered version of this predictor (Cressie 1993, p. 128).

If instead of predicting $Z(A)$ associated with a region A, we want to predict $Z(\mathbf{s})$ associated with a point $\mathbf{s}$ using data $Z(\mathbf{B})$, the analogous predictor is $\hat{Z}(\mathbf{s}) = \sum_{i=1}^n w_i(\mathbf{s}) Z(B_i)$. The weights $w_i(\mathbf{s})$ can be obtained using Equation (2.5) with $\boldsymbol{\sigma}_u$ replaced with $[\boldsymbol{\sigma}_{sB}' \mathbf{x}(\mathbf{s})']'$, where $\boldsymbol{\sigma}_{sB}$ is the $(n \times 1)$ vector with ith element $\text{cov}(\mathbf{s}, B_i)$ given in Equation (2.3).

The best linear unbiased predictors presented in this section can also be written in terms of the semivariogram. The relationship between the semivariogram associated with $Z(B)$ and that associated with the underlying process of point support $Z(\mathbf{s})$ is given by (Cressie 1993, p. 66)

$$\begin{aligned} 2\gamma(B_i, B_j) \equiv \text{var}(Z(B_i) - Z(B_j)) = & - \frac{1}{|B_i||B_i|}\int_{B_i}\int_{B_i} \gamma(\mathbf{s} - \mathbf{u})\, d\mathbf{s}\, d\mathbf{u} \\ & - \frac{1}{|B_j||B_j|}\int_{B_j}\int_{B_j} \gamma(\mathbf{s} - \mathbf{u})\, d\mathbf{s}\, d\mathbf{u} \\ & + \frac{2}{|B_i||B_j|}\int_{B_i}\int_{B_j} \gamma(\mathbf{s} - \mathbf{u})\, d\mathbf{s}\, d\mathbf{u}, \end{aligned} \tag{2.11}$$

where $\gamma(\mathbf{s} - \mathbf{u}) = (1/2)\text{var}(Z(\mathbf{s}) - Z(\mathbf{u}))$ is the semivariogram of the point-support process $\{Z(\mathbf{s})\}$.

The predictor in Equation (2.4) allows a comprehensive framework for spatial prediction involving spatial averages. Taking **B** to be spatial units of point support leads to upscaling and block kriging. If the A-units are nested within the B-units, the predictor can be used for downscaling. In other cases, both units may be on similar spatial scales with A units overlapping B units (e.g., ZIP code areas and Census tracts). This framework allows for completely general choices for the A and B units, provided each set of units forms a partition of the same fixed domain. Although the focus here thus far has been on spatial averages, best linear unbiased prediction often works quite well with a variety of different types of data. We adapt the same ideas for spatial count data in Section 2.3 and show that this approach can be used with a variety of change of support predictions problems and forms a basis for a more general, comprehensive solution.

2.3 CHANGE OF SUPPORT WITH COUNT DATA

The ideas described in Section 2.2 can be further modified for use with count data, for example to predict population data or species counts on different spatial units. In this case, we assume aggregate data $Z(B_1), Z(B_2), \ldots, Z(B_n)$ are collected where

$$Z(B_i) = \int_{B_i} Z(\mathbf{s})d\mathbf{s}. \tag{2.12}$$

For conciseness, and to emphasize the generality of our approach, we use the same notation for $Z(B_i)$ in Equation (2.12) and $Z(B)$ in Equation (2.1), although we emphasize that the two quantities are different. Then the expected value of $Z(A)$ over any region, say A, is now

$$E(Z(A)) = \sum_{l=1}^{p} x_l(A)\beta_l \equiv \mathbf{x}(A)'\boldsymbol{\beta}, \tag{2.13}$$

where $\mathbf{x}(A) \equiv (x_1(A), \ldots, x_p(A))'$ and with $x_l(A)$ now defined as

$$x_l(A) = \int_A x_l(\mathbf{s})d\mathbf{s}. \tag{2.14}$$

Thus, the matrices needed for prediction have the same form as those in Section 2.2, but their elements must be modified slightly to reflect the use of counts instead of spatial averages. Specifically, the elements of the $(n \times n)$ matrix Σ_{BB} are the (B_i, B_j) covariances given by

$$C(B_i, B_j) = \text{cov}(Z(B_i), Z(B_j)) = \int_{B_j}\int_{B_i} C(\mathbf{u} - \mathbf{v})d\mathbf{u}d\mathbf{v}, \tag{2.15}$$

(with the elements of $\boldsymbol{\sigma}_{AB}$ defined analogously) and the covariates forming the elements of $\mathbf{x}(\mathbf{B})$ and $\mathbf{x}(\mathbf{A})$ are defined according to Equation (2.14). Note that if $x_1(\mathbf{s}) \equiv 1$ corresponding to fitting an intercept in the regression equation for $Z(\mathbf{s})$, then the first of the unbiasedness conditions is

$$\sum_{i=1}^{n} w_i(A)|B_i| = |A|.$$

2.4 CONSISTENCY IN AGGREGATION

Because kriging-type predictors are exact interpolators, the geostatistical predictor presented in Section 2.2 is aggregation consistent. Thus, for $Z(B)$ in Equation (2.1), $\hat{Z}(A)$ in Equation (2.4) with weights given in Equation (2.5) satisfies the mass balance property of Huang et al. (2002): for $\{A_j : j = 1, \ldots J_i\}$ such that $A_j \bigcap A_k = \emptyset$ and $\bigcup_{j=1}^{J_i} A_j = B_i$

$$Z(B_i) = \sum_{j=1}^{J_i} |A_i| \hat{Z}(A_j)/|B_i|, \quad i = 1, \ldots, n. \tag{2.16}$$

To see this, it is helpful to reexpress the geostatistical predictor as (see, e.g., Cressie 1993)

$$\hat{Z}(A_j) = \hat{\mu} + \mathbf{c}'_{A_j B} \Sigma_{BB}^{-1} (Z(\mathbf{B}) - \hat{\boldsymbol{\mu}}),$$

where $\mathbf{c}_{A_j B} = (\text{cov}(A_j, B_1), \text{cov}(A_j, B_2), \ldots, \text{cov}(A_j, B_n))'$, with each covariance having the form of Equation (2.9), and $\hat{\mu}$ is the generalized least squares estimator of μ. Now,

$$\sum_{j=1}^{J_i} |A_i| \hat{Z}(A_j)/|B_i| = \hat{\mu} \sum_{j=1}^{J_i} |A_j|/|B_i| + \sum_{j=1}^{J_i} \frac{|A_j|}{|B_i|} \mathbf{c}'_{A_j B} \Sigma_{BB}^{-1} (Z(\mathbf{B}) - \hat{\boldsymbol{\mu}}). \tag{2.17}$$

Since $\sum_{j=1}^{J_i} |A_j| = |B_i|$, the first term is equal to $\hat{\mu}$. The first part of the second term has a kth row equal to

$$\begin{aligned} \sum_{j=1}^{J_i} \frac{|A_j|}{|B_i|} \text{cov}(A_j, B_k) &= \frac{1}{|B_i|} \sum_{j=1}^{J_i} \frac{|A_j|}{|B_k||A_j|} \int_{A_j} \int_{B_k} C(\mathbf{u} - \mathbf{v}) d\mathbf{u} d\mathbf{v} \\ &= \frac{1}{|B_k||B_i|} \int_{\bigcup_{j=1}^{J_i} A_j} \int_{B_k} C(\mathbf{u} - \mathbf{v}) d\mathbf{u} d\mathbf{v}, \end{aligned}$$

which is equal to $\text{cov}(B_i, B_k)$. Thus, Equation (2.17) is

$$\hat{\mu} + \mathbf{c}'_{B_i B} \Sigma_{BB}^{-1} (Z(\mathbf{B}) - \hat{\boldsymbol{\mu}}),$$

with $\mathbf{c}'_{B_i B}$ having elements $[[\text{cov}(B_i, B_k)]]$, for $k = 1, \ldots, n$. By definition, this is $\hat{Z}(B_i)$ and since kriging predictors are exact interpolators, this is equal to $Z(B_i)$.

A very similar derivation holds for counts using $Z(B)$ defined in Equation (2.12). In this case, the geostatistical predictor satisfies

$$\sum_{j=1}^{J_i} \hat{Z}(A_j) = Z(B_i) \text{ for } i = 1, \ldots, n.$$

When making point-level predictions from aggregated data, $\hat{Z}(\mathbf{s})$ satisfies what Tobler (1979) called the *pycnophylactic property* (or volume preserving property):

$$\int_{B_i} \hat{Z}(\mathbf{s}) d\mathbf{s} = Z(B_i), \quad \text{for} \quad i = 1, \ldots, n. \tag{2.18}$$

2.5 INFERRING SPATIAL DEPENDENCE

As in the derivation of the ordinary and universal kriging predictors, the above development assumes that the underlying point-point covariance function or semivariogram is known (see, e.g., Cressie 1993). In practice, one of these functions must be inferred from the data and the estimator used must lead to a valid semivariogram (i.e., one that is conditionally negative definite) or a valid covariance function (i.e., one that is positive definite). To ensure this validity, we follow traditional geostatistical methods and restrict our inference to parametric models, $\gamma(\mathbf{s}-\mathbf{u};\boldsymbol{\theta})$ or $C(\mathbf{u}-\mathbf{v};\boldsymbol{\theta})$, that are known to valid in $\Re^2$.

A generalized estimating equations (GEE) approach provides a flexible method for estimating $\boldsymbol{\theta}$ (McShane, Albert, and Palmatier 1997). Consider the squared differences

$$Y_{ij}^{(1)} = (Z(B_i) - Z(B_j))^2,$$

with $Z(B)$ defined in Equation (2.1). Note that if $E(Z(\mathbf{s})) \equiv \mu$, then $E(Z(B_i)-Z(B_j)) = 0$ and $E(Y_{ij}) = 2\gamma(B_i, B_j;\boldsymbol{\theta})$, where $2\gamma(B_i, B_j;\boldsymbol{\theta})$ is a parametric model for $2\gamma(B_i, B_j)$ given by Equation (2.11). Taking an identity working variance-covariance matrix for the $Y_{ij}^{(1)}$, the generalized estimating equations for $\boldsymbol{\theta}$ can be written as

$$U(\boldsymbol{\theta};\{Y_{ij}^{(1)}\}) = 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\partial\gamma(B_i, B_j;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\left(Y_{ij}^{(1)} - 2\gamma(B_i, B_j;\boldsymbol{\theta})\right) \equiv \mathbf{0}.$$

Unfortunately, solving these equations requires repeated evaluation of $2\gamma(\mathbf{h}_{ij},\boldsymbol{\theta})$, which in turn requires computation of three spatial integrals (Equation (2.11)). If instead we consider the cross products

$$Y_{ij}^{(2)} = (Z(B_i) - \mu_s)(Z(B_j) - \mu_s), \tag{2.19}$$

where $\mu_s = E(Z(\mathbf{s})) = E(Z(B))$, then the generalized estimating equations become

$$U(\boldsymbol{\theta};\{Y_{ij}^{(2)}\}) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\partial C(B_i, B_j;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\left(Y_{ij}^{(2)} - C(B_i, B_j;\boldsymbol{\theta})\right) \equiv \mathbf{0}, \tag{2.20}$$

where $C(B_i, B_j;\boldsymbol{\theta})$ is a parametric version of Equation (2.7). This approach yields an estimator of $\boldsymbol{\theta}$ that is a special case of the approach suggested by Mockus (1998) that finds $\hat{\boldsymbol{\theta}}$ as the vector minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_{ij}\left(Y_{ij}^{(2)} - \int_{B_j}\int_{B_i} C(\mathbf{u},\mathbf{v};\boldsymbol{\theta})d\mathbf{u}d\mathbf{v}\right)^p.$$

Taking $p = 2$ and $\xi_{ij} = 1$ corresponds to ordinary nonlinear least squares estimation that is computationally equivalent to solving the generalized estimating equations in (2.20). However, since μ_s is unknown, Equations (2.20) will not be unbiased estimating equations if we use an estimator, $\hat{\mu}_s$, to define the cross products in Equation (2.19). We can solve this problem and also treat the more general case where $\mu_s = \mu_s(\boldsymbol{\beta}) = \sum_{l=1}^{p} x_l(\mathbf{s})\beta_l \equiv \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}$

by using another set of estimating equations for $\boldsymbol{\beta}$. Thus, extending the ideas in McShane et al. (1997) with the models developed in Section 2.2, the first estimating equation for $\boldsymbol{\beta}$ is

$$\mathbf{x}(\mathbf{B})'[V_{BB}(\boldsymbol{\theta})]^{-1}(Z(\mathbf{B}) - \mathbf{x}(\mathbf{B})\boldsymbol{\beta}) = \mathbf{0}, \tag{2.21}$$

where $Z(\mathbf{B}) = (Z(B_1), \ldots, Z(B_n))'$, the elements of $V_{BB}(\boldsymbol{\theta})$ are $C(B_i, B_j; \boldsymbol{\theta})$ and $\mathbf{x}(\mathbf{B})$ is a $(n \times p)$ matrix with (i, j) element $1/|B_i| \int_{B_i} x_j(\mathbf{s})d\mathbf{s}$. The second estimating equation has the same form as that in (2.20), but is based on cross products of residuals

$$Y_{ij}^{(3)} = (Z(B_i) - \mathbf{x}(B_i)\boldsymbol{\beta})(Z(B_j) - \mathbf{x}(B_j)\boldsymbol{\beta}). \tag{2.22}$$

Taking an identity matrix as the working variance matrix of $Y_{ij}^{(3)}$, the second estimating equation for $\boldsymbol{\theta}$ is

$$U(\boldsymbol{\theta}; \{Y_{ij}^{(3)}\}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\partial C(B_i, B_j; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \left(Y_{ij}^{(3)} - C(B_i, B_j; \boldsymbol{\theta})\right) \equiv \mathbf{0}. \tag{2.23}$$

Consistent estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be obtained by solving (2.21) and (2.23) simultaneously by iteration. The estimator of $\boldsymbol{\beta}$ obtained from the first estimating Equation (2.21) is just a generalized least squares estimator. Thus, solving (2.21) and (2.23) simultaneously is an iteratively reweighted generalized least squares algorithm with covariance parameters estimated using Equation (2.23).

3. ILLUSTRATION AND COMPARISON

In this section, we present two examples chosen to illustrate the flexibility of geostatistical methods in general change of support problems. Both examples are based on a domain containing 25 contiguous counties in southeastern Georgia that comprise Georgia Health Care District 9 (GHCD9) (Figure 1). We focus on downscaling problems since the use of block kriging for upscaling and aggregation is fairly well known in other disciplines. The first example illustrates the general geostatistical approach using spatial averages. The second example involves predicting count data and allows us to compare the results of our approach to those obtained from commonly used methods (e.g., Tobler's pycnophylactic interpolation and proportional allocation).

3.1 PREDICTING LOW BIRTH WEIGHT RATES

Our research was initially motivated by our study of very low birth weight babies whose mothers resided in Georgia Health Care District 9 in 1999 (Rogers et al. 2000; Gotway and Young 2004). For confidentiality reasons, we could not present county-level rates for the very low birth weight babies, so we consider the number of low birth weight (LBW) babies, those that weigh less than 2,500 grams at birth, taken from the vital statistics database for Georgia for 1999. This example is more practical in many ways since vital statistics reported for each county are often available in the public domain and many observational studies in epidemiology and geography routinely use this type of data.

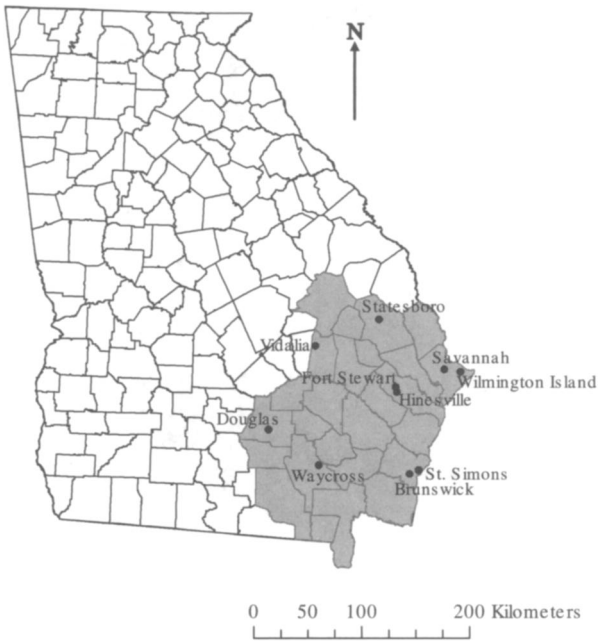

Figure 1. Counties comprising Georgia Health Care District 9.

Of interest in this article is the use of aggregated data on B_i units (here, low birth weight rates per county for each of $i = 1, \ldots, 25$ counties) to predict data on A_j units (here, low birth weight rates for each Census tract, for each of $j = 1, \ldots n_i$ tracts where n_i is the number of tracts in county i) where more specific covariate information can be obtained. Thus, $Z(B_i)$ in Equation (2.1) is the low birth weight rate observed for county i computed as the number of low birth weight babies born in 1999 per number of live births in 1999 per 100,000 people. The boundaries of the counties and the Census tracts were projected to Universal Transverse Mercator (UTM) coordinates based on the Mercator projection to UTM Zone 17N and referenced to the NAD83 Geodetic datum. The resulting spatial coordinates are given in meters east and meters north and all distances are in meters or kilometers. A map showing the county boundaries, the tract boundaries, and a proportional symbol indicating the relative magnitudes of the county-level LBW rates is shown in Figure 2.

In this example, we use the number of babies born to mothers living in each Census tract that are less than one year of age as a covariate in the prediction of LBW rates for each tract. While the number of live births per Census tract might be a better covariate, this type of information is not easily available at the Census-tract level. We consider a simple linear regression relationship between this covariate and the low birth weight data, $Z = \beta_1 + \beta_2 x_2$, where x_2 represents the population density of children less than 1 year of age. Then, the matrix of covariate values, $\mathbf{x}(\mathbf{B})$, comprising part of Σ_u in Equation (2.6) is constructed in the following way. The first column contains a vector of 1's, corresponding to taking $x_1(\mathbf{s}) = 1$ so that $x_1(B_i) = \int_{B_i} x_1(\mathbf{s})d\mathbf{s}/|B_i| = 1$. The second column contains $x_2(B_i) = \int_{B_i} x_2(\mathbf{s})d\mathbf{s}/|B_i|$, the average population density of children less than 1 year of

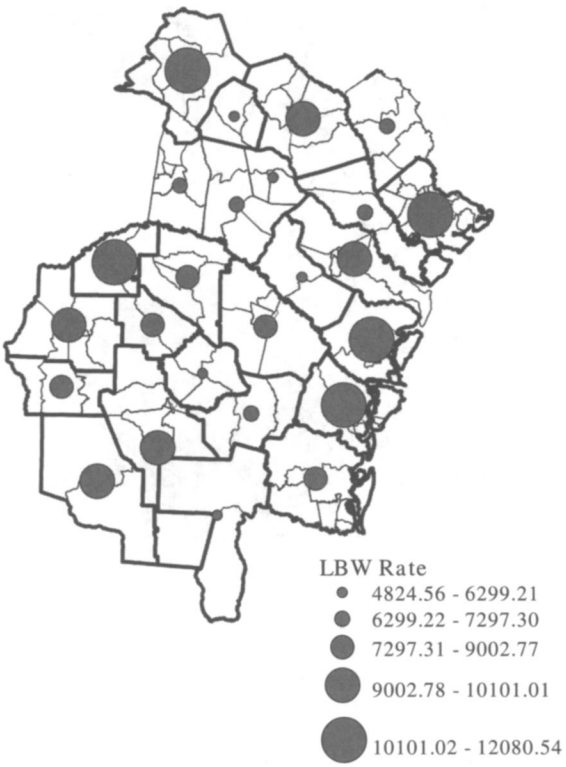

Figure 2. Low birth weight rates per year per 100,000 live births in GHCD9 in 1999. The thin lines are the boundaries of the Census tracts. Proportional symbols are located at county centroids.

age in county B_i. The vector $\mathbf{x}(A)$ in Equation (2.8) is $\mathbf{x}(A) = (1\ x_2(A))'$ with $x_2(A) = \int_A x_2(\mathbf{s})d\mathbf{s}/|A|$. Because Census units are nested and Census data are upwardly additive, $x_2(B) = \sum_i |A_i|x_2(A_i)/|B|$, for $\bigcup_i A_i = B$.

To model the small-scale spatial variability we assume that $\text{cov}(Z(\mathbf{s}), Z(\mathbf{u})) = C(\mathbf{s} - \mathbf{u}; \boldsymbol{\theta})$. To determine a plausible parametric covariance function for $C(\mathbf{s} - \mathbf{u}; \boldsymbol{\theta})$ and to estimate $\boldsymbol{\theta}$ from Equation (2.23), we combine the ideas of Mockus (1998) with those of Cressie (1985) on nonlinear least squares estimation for semivariogram models based on data with support. We assess the variability in the cross products $Y_{ij}^{(3)}$ (defined in Equation (2.22)) as a function of the support-adjusted distance between regions B_i and B_j computed as

$$d_{ij} = \frac{\int_{\mathbf{u}\in B_i}\int_{\mathbf{v}\in B_j} ||\mathbf{u} - \mathbf{v}||d\mathbf{u}d\mathbf{v}}{\int_{\mathbf{u}\in B_i}\int_{\mathbf{v}\in B_j} d\mathbf{u}d\mathbf{v}}, \tag{3.1}$$

where $||\mathbf{u} - \mathbf{v}||$ is the Euclidean distance between spatial locations $\mathbf{u}$ and $\mathbf{v}$. Based on the shape of the resulting empirical covariance function, we used a Gaussian covariance function to model the underlying point-point covariance function:

$$C(\mathbf{s} - \mathbf{u}; \boldsymbol{\theta}) = \tau^2 \exp(-||\mathbf{s} - \mathbf{u}||/a). \tag{3.2}$$

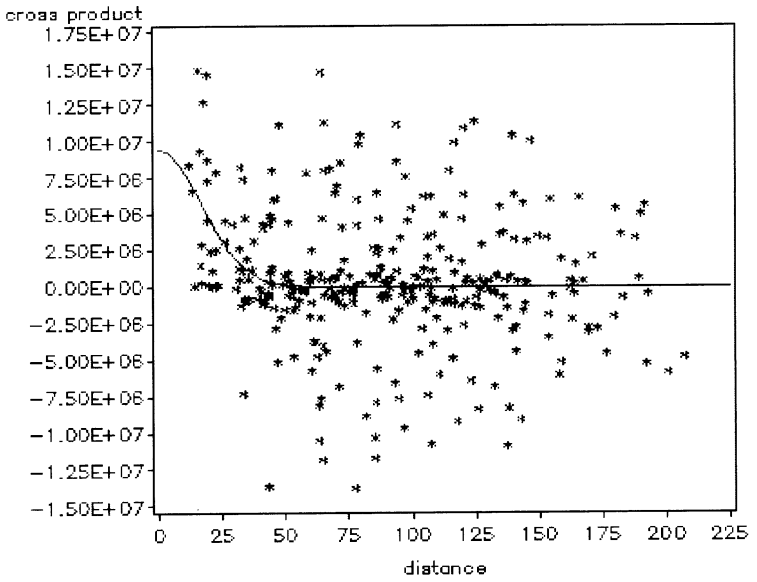

Figure 3. Empirical covariance function and fitted model based on the methods described in Section 2.5.

Estimates of $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ and $\boldsymbol{\theta} = (a, \tau^2)'$ were obtained using iteratively reweighted generalized least squares as defined by Equations (2.21) and (2.23). The spatial integrals in Equations (2.9) and (3.1) were computed using basic numerical integration. The GHCD9 domain was discretized using a very fine (1 km $\times$ 1 km) grid and the integrals were approximated using Riemann sums. We used a fine grid in order to adequately represent the support of the small tracts. The amount of computation required will depend on the support of the units involved and their relative sizes. In this case, the iteratively reweighted generalized least squares algorithm converged after six iterations and we obtained $\boldsymbol{\beta} = (8258.84, 1910.59)'$, $\hat{a} = 24.06$ km, and $\hat{\tau}^2 = 9398577$. The final empirical residual covariance function and the fitted model are both shown in Figure 3. There is a lot of variation in this empirical covariance function since it is based on only 25 counties and there is considerable variation in the LBW rates. Consequently, we compared our approach to one based on binning and averaging the covariance values using specified distance intervals. The estimates obtained were very similar.

Tract-level predictions were then obtained using the predictor in Equation (2.4) with optimal weights defined in Equation (2.5). These predictions are shown in Figure 4 and the corresponding root prediction mean-squared errors obtained by taking the square root of Equation (2.10) are shown in Figure 5.

Several features of interest can be seen in Figure 4. First, although Chatham county containing Savannah in the northeast section of GHCD9 had one of the highest low birth weight rates, the predicted rates for each tract are relatively low. This illustrates the effect of the mass balance constraint: since the number of tracts in this county is very large and the rates must average to the observed LBW rate, the predicted rate in each tract must

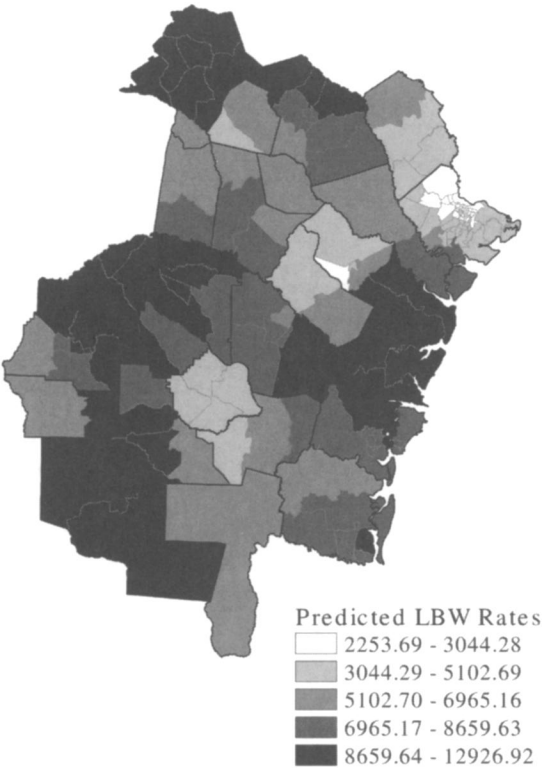

Figure 4. Predicted LBW rates for each tract using the geostatistical predictor.

be relatively low. Second, tracts with relatively high (low) rates tend to be connected to proximate tracts that also have high (low) rates. This is the effect of incorporating spatial autocorrelation in the geostatistical modeling technique and also results from smoothing the high (low) values. Finally, a tract within a county that has a relatively high LBW rate can have a relatively low predicted LBW rate if the tract is closer to a county with a relatively lower LBW rate than it is to the county in which it is contained. This is due to the effect of autocorrelation and the use of support averaged distances.

A comparison of Figures 4 and 5 indicates that, in general, as the predicted LBW rate increases, the RPMSE decreases. Thus, many of the tracts with relatively high predicted LBW rates in the left center part of the region tend to have relatively low RPMSEs. The converse tends to be true in areas like Savannah that have relatively low predicted LBW rates and high RPMSEs. However, this relationship is confounded by the observed LBW rates and the number of tracts per county: the average RPMSE is higher for counties that have high LBW rates and also large numbers of tracts per county.

3.2 Predicting ZIP Code Populations

To illustrate how the geostatistical approach works with count data on overlapping spatial units, we used the Census-derived 1999 county populations to predict populations for ZIP

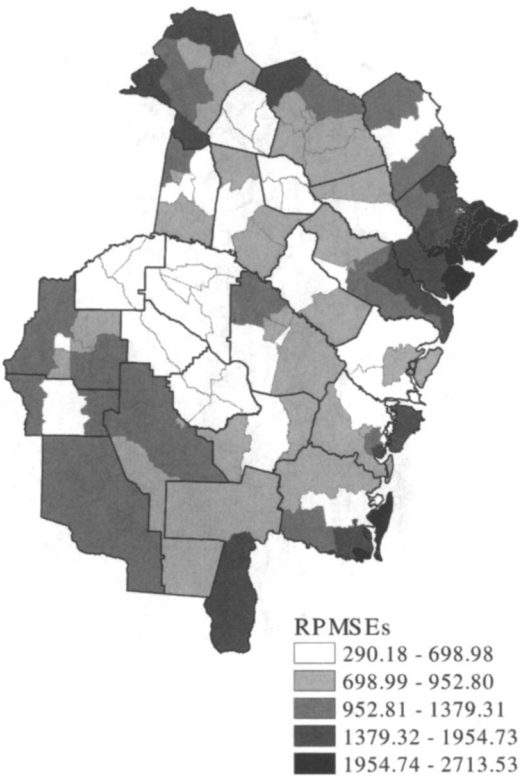

Figure 5. Root prediction mean-squared errors for LBW rate predictions from geostatistical prediction.

code areas. The county-level populations and boundaries, and the ZIP code boundaries are shown in Figure 6. We used the geostatistical approach to spatial prediction described in Section 2.3 to predict population counts for each ZIP code area in GHCD9. We chose to use population data since fairly good estimates of the population in each ZIP code can be obtained from Census data by adding Census block populations. Thus, we can compare the predictions obtained from the geostatistical approach using county-level data to estimates obtained by summing Census block populations within each ZIP code. We also compare our approach to the commonly used methods of proportional allocation and Tobler's pycnophylactic interpolation (Tobler 1979). In order to more directly compare all of the approaches, we did not use covariates when making predictions with the geostatistical method, but covariates like the number of street intersections or the total road length in each geographic unit might improve these predictions.

We assumed the data model described in Section 2.3, with $Z(B_i)$ in Equation (2.12) representing the county-level populations, $E(Z(\mathbf{s})) \equiv \mu$, and $C(B_i, B_j)$ given in Equation (2.15). We again assumed the same general form for the parametric covariance function, $C(\mathbf{s}-\mathbf{u}; \boldsymbol{\theta})$, as in Equation (3.2). We defined $Y_{ij}^{(3)}$ in Equation (2.22) based on the normalized variables $Z(B)/|B|$ and estimated $\mathbf{x}(B_i)\boldsymbol{\beta} \equiv E(Z(B)/|B|) = \mu$ with $\hat{\mu} = \bar{Y}$. Then, $\boldsymbol{\theta}$ was estimated from Equation (2.23), after also normalizing the covariance function to be $C(B_i, B_j; \boldsymbol{\theta})/(|B_i||B_j|)$, and we obtained $\hat{a} = 5.64$ km and $\tau^2 = 1353453$. As in the

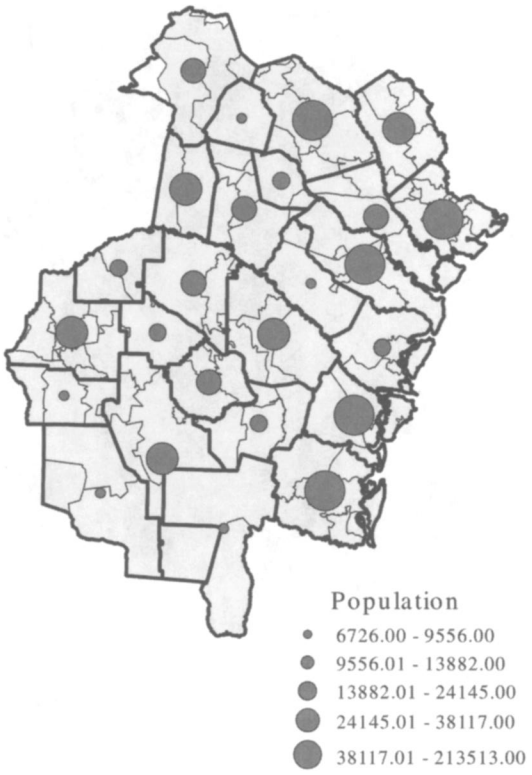

Figure 6. County populations and ZIP code boundaries.

previous example, the spatial integrals were computed using basic numerical integration over a (1 km × 1 km) grid. Note that without covariates, no iteration is required. This gave estimates of $C(B_i, B_j)$ and $C(A, B_i)$ that were used to form the elements of Σ_{BB} and σ_{AB} (Equations (2.6) and (2.8)) needed to compute the weights from Equation (2.5). The optimal predictor of the count associated with each ZIP code unit was then computed using Equation (2.4). When working with large counts with substantial spatial variation, the mass-balance constraint can be severe. To avoid oversmoothing, make more localized predictions, and ensure non-negative predictions, we used an adaptive search neighborhood, fixing each tract prediction on the 5–15 counties closest to the county containing the tract of interest. Proximity was determined using the support averaged distance between counties defined in Equation (3.1). This adaptive search strategy allowed us to keep the ZIP code unit predictions positive, but also allowed us to use as many neighbors as possible in order to keep the prediction mean squared errors as small as possible. A map of the resulting predicted ZIP code populations is shown in Figure 7.

We compared these results to estimates obtained from the commonly used method of proportional allocation, where the population in each county is divided among the ZIP code units according to their areas, that is,

$$\hat{Z}_{\text{pa}}(A_j) = Z(B_i)|A_j|/|B_i|.$$

With proportional allocation, ZIP code units with larger areas receive a greater proportion of the total count, and those with smaller areas receive a smaller proportion of the total count. The mass balance property is preserved by construction and there is no borrowing of information from neighboring counties or tracts as there is with the geostatistical method. We note that proportional allocation is a special case of the geostatistical prediction method developed here obtained by choosing a white noise covariance function for the underlying point-level stochastic process. Predicted counts for each ZIP code unit obtained by proportional allocation are shown in Figure 7.

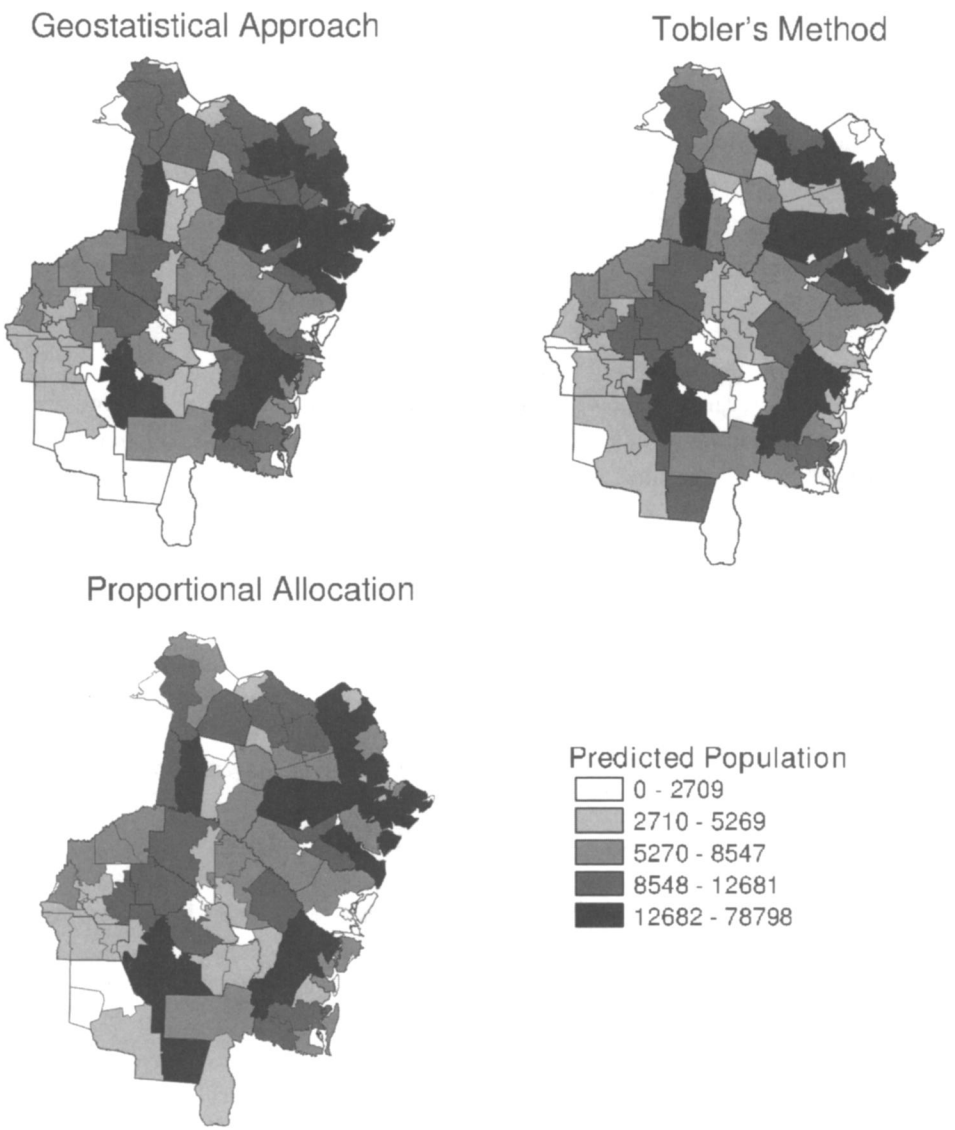

Figure 7. Predicted ZIP code populations.

Since smoothing and mass balance are somewhat at odds with each other, we also compared the predictions from the geostatistical method to another approach that incorporates these two opposing properties. Tobler's pycnophylactic interpolation (Tobler 1979) obtains estimates of $Z(\mathbf{s}) = Z(x, y)$ in Equation (2.1) as values of the spatial function that minimizes

$$\int\int \left[\left(\frac{\partial^2 Z(x, y)}{\partial x^2} \right)^2 + \left(\frac{\partial^2 Z(x, y)}{\partial y^2} \right)^2 \right] dxdy, \tag{3.3}$$

subject to the constraints $Z(\mathbf{s}) \geq 0$ and the constraints imposed by the pycnophylactic property given in Equation (2.18). When applied to point data, the intensity surface that minimizes Equation (3.3) is the surface of minimum curvature. The minimum curvature method produces smooth surfaces from irregularly spaced data that are visually pleasing and constrained to honor (pass through) the data values. In the case of areal data, with the additional constraint of Equation (2.18), the surface is constrained to preserve volume: the intensity process integrates to the observed data for each region.

To use Tobler's approach for areal interpolation, as opposed to density estimation for which it was developed, we first estimated $Z(s)$ by $\hat{Z}_T(\mathbf{s})$, the function that minimizes Equation (3.3) subject to the constraints in Equation (2.18), and then estimated the count in each ZIP code unit using

$$\hat{Z}_T(A_j) = \int_{A_j} \hat{Z}_T(\mathbf{s}) d\mathbf{s}.$$

These estimates are also shown in Figure 7.

Visually, all three maps of predicted ZIP code unit populations are similar. However, the geostatistical method allows neighboring information to influence predictions more than the other two methods. The results from Tobler's method include several zero estimates for ZIP code areas with relatively small populations, suggesting that the boundary conditions may be unduly influencing estimates for these regions. The largest (smallest) values estimated by proportional allocation were larger (smaller) than predictions made by the geostatistical method or Tobler's approach, reflecting the smoothing done by these latter two methods. Surprisingly, these maps roughly capture spatial variation in ZIP code populations obtained by summing the Census block population estimates (Figure 8), even though many important factors affecting the geographic distribution of human populations were not considered in this analysis (e.g., land use). The ability to incorporate covariate information in population predictions is one advantage of the geostatistical approach. A striking illustration of this need is reflected in the results for ZIP code 31314: the estimated population obtained from summing Census block population data is 124, but all three statistical methods predicted a population of approximately 30,000 for this ZIP code. ZIP code 31314 contains Fort Stewart, an active military base.

The geostatistical method has another important advantage over the other approaches considered here in that a measure of prediction uncertainty can be obtained as a result of minimizing the prediction mean squared error. A map of the root mean-squared prediction errors associated with ZIP code unit predictions from the geostatistical method, computed as the square root of Equation (2.10), is shown in Figure 9. Note that predictions with the largest uncertainty (in the southern part of GHCD9) correspond to units with small

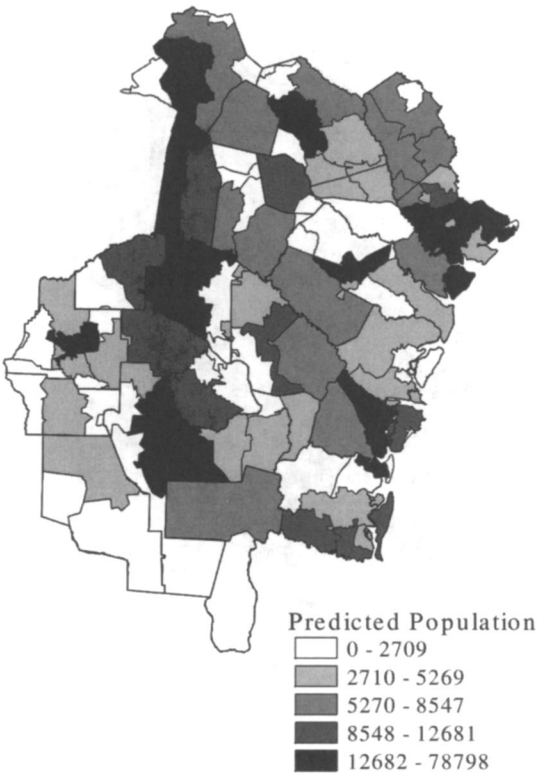

Figure 8. Estimated ZIP code populations using the sum of the block group populations.

predicted populations and predictions with the smallest uncertainty (as in Savannah) tend to be associated with large predicted population counts.

4. SUMMARY AND CONCLUSIONS

This article proposes a geostatistical framework for linking geographically aggregated data from different sources that explicitly uses the spatial support of the data. The new prediction method we propose can be used for upscaling (spatial aggregation), downscaling (spatial disaggregation), or side scaling, a term we use to refer to the prediction of values on one set of spatial units from data on another set of overlapping spatial units. The spatial units may be of point, areal, or volumetric support. When used for disaggregating count data when the target units are nested within the source units, the total for target units within a source unit is equal to the original source unit total (mass balance). When predicting a spatial intensity function of point support from geographically aggregated data, the predicted intensity function satisfies the pycnophylactic property. This geostatistical framework can also be used for prediction of environmental variables that are spatial averages instead of totals. To facilitate the use of this new approach for routine use in commercial GIS software, we also developed a new approach for the simultaneous estimation of mean and covariance functions from aggregated data.

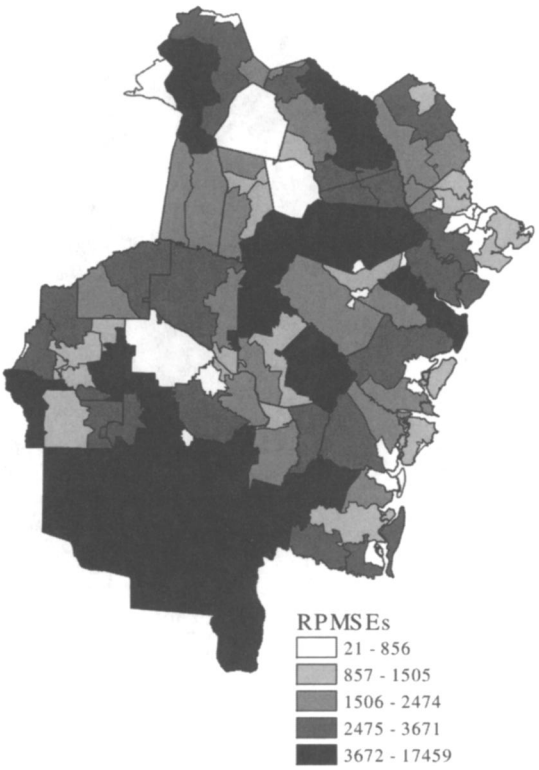

Figure 9. Root prediction mean-squared errors for ZIP code unit predictions from geostatistical prediction.

The primary advantages of this approach over those developed in the geography literature are the ability to adjust predictions for important covariate values and the measure of uncertainty that can be obtained for each prediction. The commonly used method of proportional allocation is a special case of the geostatistical approach. The main advantages of this approach over those presented in the statistical literature include the general framework that can handle a variety of spatial prediction problems, the lack of distributional assumptions that must be made, and the ability to perform the basic computations within a GIS. As with ordinary point kriging, the geostatistical predictor honors the data, that is, the predictor interpolates the observed data.

When working with spatial statistics, the support of the data is a critical component. Geographic information systems can easily combine data with differing supports, but they typically rely on simple arithmetic methods such as proportional allocation and zonal addition for such integration. The geostatistical framework for change of support problems presented in this article provides an attractive alternative, allowing adjustment for covariate values and providing a measure of uncertainty for spatial predictions necessary for aggregation and disaggregation of spatial data. A remaining challenge involves convincing the management of commercial GIS vendors of the utility of statistical methods and developing defaults that allow automatic solutions to change of support problems.

[Received June 2004. Revised July 2006.]

REFERENCES

Bracken, I., and Martin, D. (1989), "The Generation of Spatial Population Distributions from Census Centroid Data," *Environment and Planning A*, 21, 537–543.

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.

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

Cressie, N. (1985), "Fitting Variogram Models by Weighted Least Squares," *Journal of the International Association for Mathematical Geology*, 17, 563–586.

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

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

Flowerdew, R., and Green, M. (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.

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

Gotway, C. A., and Young, L. J. (2002), "Combining Incompatible Spatial Data," *Journal of the American Statistical Association*, 97, 632–648.

——— (2004), "A Spatial View of the Ecological Inference Problem," in *Ecological Inference: New Methodological Strategies*, eds. G. King, M. Tanner, and O. Rosen, New York: Cambridge University Press.

Huang, H.-C., Cressie, N., and Gabrosek, J. (2002), "Fast Resolution-Consistent Spatial Prediction of Global Processses from Satellite Data," *Journal of Computational and Graphical Statistics*, 11, 1–26.

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

Kelsall, J., and Wakefield, J. (2002), "Modeling Spatial Variation in Disease Risk: A Geostatistical Approach," *Journal of the American Statistical Association*, 97, 692–701.

McShane, L., Albert, P., and Palmatier, M. (1997), "A Latent Process Regression Model for Spatially Correlated Count Data," *Biometrics*, 53, 698–706.

Mockus, A. (1998), "Estimating Dependencies From Spatial Averages," *Journal of Computational and Graphical Statistics*, 7, 501–513.

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.

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.

Rogers, J. F., Thompson, S. J., Addy, C. L., McKeown, R. E., Cowen, D. J., and DeCoulfé, P. (2000), "The Association of Very Low Birthweight With Exposures to Environmental Sulfur Dioxide and Total Suspended Particulates," *American Journal of Epidemiology*, 151, 602–613.

Tobler, W. (1979), "Smooth Pycnophylactic Interpolation for Geographical Regions" (with discussion), *Journal of the American Statistical Association,*, 74, 519–536.

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