Reconstructing Population Density Surfaces from Areal Data: A Comparison of Tobler's Pycnophylactic Interpolation Method and Area-to-Point Kriging

Eun-Hye Yoo,[1] Phaedon C. Kyriakidis,[2] Waldo Tobler[2]

[1]Department of Geography, University at Buffalo, The State University of New York, Buffalo, NY,
[2]Department of Geography, University of California, Santa Barbara, CA

We compare Tobler's pycnophylactic interpolation method with the geostatistical approach of area-to-point kriging for distributing population data collected by areal unit in 18 census tracts in Ann Arbor for 1970 to reconstruct a population density surface. In both methods, (1) the areal data are reproduced when the predicted population density is upscaled; (2) physical boundary conditions are accounted for, if they exist; and (3) inequality constraints, such as the requirement of non-negative point predictions, are satisfied. The results show that when a certain variogram model, that is, the de Wijsian model corresponding to the free-space Green's function of Laplace's equation, is used in the geostatistical approach under the same boundary condition and constraints with Tobler's approach, the predicted population density surfaces are almost identical (up to numerical errors and discretization discrepancies). The implications of these findings are twofold: (1) multiple attribute surfaces can be constructed from areal data using the geostatistical approach, depending on the particular point variogram model adopted—that variogram model need not be the one associated with Tobler's solution and (2) it is the analyst's responsibility to justify whether the smoothness criterion employed in Tobler's approach is relevant to the particular application at hand. A notable advantage of the geostatistical approach over Tobler's is that it allows reporting the uncertainty or reliability of the interpolated values, with critical implications for uncertainty propagation in spatial analysis operations.

Introduction

Population data are used extensively in decision-making processes in a wide range of social and economic applications, including but not limited to housing, regional

Correspondence: Eun-Hye Yoo, Department of Geography, University at Buffalo, The State University of New York, Buffalo, NY
e-mail: eunhye@buffalo.edu

policy, and health provision (Bracken 1994). Despite the increasing availability of socioeconomic data at the very high spatial resolution required in urban analysis and modeling, population microdata are still either suppressed or incomplete due to confidentiality and the high cost of data collection (Ryan, Maoh, and Kanaroglou 2009). Perhaps the commonly available population data are the summary statistics of population data reported and mapped in irregular geographic areas (e.g., census tracts and enumeration districts) and often need to be transformed into a spatial unit compatible with other data sources (Thurstain-Goodwin and Unwin 2000). A particular case of such a transformation or estimation is the construction of a continuous population density surface, which is often favored in the literature (Tobler 1975, 1979; Goodchild, Anselin, and Deichmann 1993; Langford and Unwin 1994; Martin, Tate, and Langford 2000; Thurstain-Goodwin and Unwin 2000).

Starting from the conventional approach based on choropleth mapping, various areal interpolation methods have been used to construct a density surface from population data. Kernel smoothing, for example, produces a smooth surface that is free of either abrupt discontinuities along boundaries or the strict assumption of within-area homogeneity of source data (Martin 1989, 1996). In this approach, however, the support differences between the source data and the prediction surface are not properly taken into account, because the areal data are collapsed into their corresponding representative points (e.g., polygon centroids). This transformation implicitly assumes that the areal unit or support of the population data is identical to that of the target surface, that is, a point.

In demographic applications of areal interpolation, the absence of empirical data concerning the actual distribution of population at the target spatial resolution is a major obstacle for evaluating any model of population density surface (Martin, Tate, and Langford 2000). Owing to a lack of point-level data, original data reproduction—whether a surface model reproduces the original areal data when the predicted population surface is reaggregated over the spatial units used to collect population data—becomes an essential requirement for accurate interpolation (Tobler 1979; Lam 1983). This property of areal interpolation is also known as the *mass-preserving* or *pycnophylactic* condition in the literature.

In addition, a priori knowledge about the smoothness of the population density surface may be available in the form of a governing equation dictating the smoothness of the interpolated surface; for example, the smoothest possible surface (within the data constraints) might be preferred. Also, information available on boundaries, including natural and human made, may significantly affect the resulting population density surface. Often neglected, but also substantially effective information may be available in the form of inequality constraints such as non-negativity or lower and upper values of the resulting population density surface. With the exception of Tobler's (1979) pycnophylactic interpolation method, however, these additional conditions have not been emphasized in surface modeling; instead, ad hoc adjustments have been applied, if necessary, to the resulting surfaces.

While addressing most of these critical issues, Kyriakidis (2004) cast attribute surface construction as an area-to-point (ATP) interpolation problem within a geostatistical framework. This researcher analytically demonstrated that, once appropriate assumptions are made on the smoothness of the interpolated surface, the geostatistical framework includes as particular cases choropleth mapping, kernel interpolation, and Tobler's pycnophylactic interpolation method. An appealing feature of the geostatistical approach, apart from its generality, is that the quality of the resulting surface is quantified via the prediction error variance associated with each predicted value. What was not provided by Kyriakidis (2004), however, was a practical, real-world demonstration of these theoretical connections.

In this article, we include a small empirical application for illustrative purposes, in which geostatistical ATP interpolation is compared with Tobler's pycnophylactic interpolation method for constructing a smooth population density surface from population data collected within arbitrary areal units. We demonstrate that, upon selection of an appropriate point variogram model in the geostatistical approach the two interpolation approaches yield very similar density surfaces within the limits of numerical errors and discretization discrepancies. The implications of this demonstration are twofold: (1) multiple attribute surfaces can be constructed from areal data using the geostatistical approach, depending on the particular point variogram model adopted—that variogram model need not be the one associated with Tobler's solution and (2) it is the analyst's responsibility to justify whether the smoothness criterion employed in Tobler's approach is relevant to the particular application at hand. As stated, a notable advantage of the geostatistical approach over Tobler's is that it allows for the assessment of the uncertainty or reliability of the interpolated values, with critical implications for uncertainty propagation in spatial analysis operations.

Several geostatistical methods have already been proposed for constructing an attribute surface from areal data. A review article by Gotway and Young (2002) provided a synthesis of geostatistical and deterministic approaches (including Tobler's pycnophylactic method) to surface construction from areal data; however, no formal analytical link was established between the logarithmic variogram used in geostatistics and the Laplacian smoothness criterion of Tobler's method. More recently, Goovaerts (2006) and Gotway and Young (2007) developed non-Gaussian geostatistical methods for surface construction from areal source data reported over irregular polygons; the former involves the application of Poisson's kriging for dealing with count data, while the latter is a generalization of the former that accounts for trends linked to pertinent covariates via a generalized linear model. Last, particular attention should be drawn to the recent work of Goovaerts (2008), who developed an iterative approach for estimating the variogram of the underlying attribute surface from areal source data, based on a postulated functional form of that point variogram model.

We believe that our work in this article complements nicely the approaches mentioned above. By illustrating the equivalence of the geostatistical approach

with Tobler's method in a real-world context, we hope to make a stronger case for the geostatistical approaches, which are not limited to a particular variogram model and allow prediction uncertainty assessment. In essence, we view our contribution to the literature as being a call for Tobler's method to be linked more to the Laplacian smoothness criterion adopted than to the pycnophylactic constraints considered, because there are multiple approaches by which those constraints can be met in surface reconstruction.

Methodology

In this article, population density is conceptualized as a continuous attribute surface $\{z(\mathbf{u}), \mathbf{u} \in \Omega\}$, in short $[z]$, where $z(\mathbf{u})$ denotes the unknown attribute value at a generic location with coordinate vector $\mathbf{u}$ within a domain Ω. The task is to reconstruct the unknown surface $[z]$ from a set of K of areal data $\mathbf{z}_v = [z(v_k), k = 1, \ldots, K]^T$, where v_k denotes the k-th areal unit or support with centroid coordinate vector $\mathbf{u}_k$ and otherwise arbitrary shape, size, and orientation. Notation-wise, we use $|v_k|$ to denote the measure (area in two-dimension [2D]) of the kth support v_k. Any observed areal datum $z(v_k)$ is assumed to stem from a convolution of the unknown point attribute values of surface $[z]$ with a known sampling function $g_k(\mathbf{u})$, as $z(v_k) = \int_{\mathbf{u} \in \Omega} g_k(\mathbf{u}) z(\mathbf{u}) d\mathbf{u}$. Function $g_k(\mathbf{u})$ quantifies the contribution of a generic point attribute value $z(\mathbf{u})$ to the kth areal datum $z(v_k)$ and could simply be an indicator point-in-polygon function, a point-in-a-feature membership function, or a more elaborate distance decay function quantifying proximity to, say, a city center.

Based on the K areal data, we seek to construct an interpolated attribute surface $\{\hat{z}(\mathbf{u}), \mathbf{u} \in \Omega\}$, in short $[\hat{z}]$, under the premise that the smoothness or regularity of the true unknown surface $[z]$ is governed by a deterministic function. Following Tobler (1979), we consider the particular case of Laplace's partial differential equation (PDE) as the governing equation of the solution surface; in principle, however, one could make different assumptions regarding the smoothness of the underlying surface depending on the particular application. For computing purposes, the interpolated attribute surface $[z]$ is typically approximated by a finite set of P attribute values $\{z(\mathbf{u}_p), p = 1, \ldots, P\}$ located on a regular grid. We then seek to predict or interpolate P unknown attribute values $\{\hat{z}(\mathbf{u}_p), p = 1, \ldots, P\}$ from the K areal data, where $\hat{z}(\mathbf{u}_p)$ denotes the predicted attribute value at the pth discretization location. Both the unknown $z(\mathbf{u}_p)$ and the predicted $\hat{z}(\mathbf{u}_p)$ attribute values are assumed to be representative of a fixed cell $c(\mathbf{u}_p)$; for the remainder of this article, we assume a unit cell size $|c(\mathbf{u}_p)| = 1$, $\forall p$.

Tobler (1979) formulated the construction of a population density surface as the discrete solution to Laplace's PDE, subject to several constraints or conditions: (1) a set of K linear equality constraints forcing the interpolated surface to reproduce the K areal data when the values of that surface are aggregated within their respective supports, (2) a set of P inequality constraints dictating that the discretized

solution surface cannot attain negative values, and (3) boundary conditions, either Dirichlet (known attribute values on the boundary) or Neumann (known normal derivatives on the boundary), enforced at a set of discrete boundary locations. Tobler (1979) solved this problem iteratively using a finite-difference scheme, whereby the attribute values at discretization nodes are readjusted under the pycnophylactic constraint so that the value at each node is the average of its neighbors—a consequence of assuming Laplace's PDE to dictate the surface smoothness. This readjustment amounts to minimizing the sum of the directional or partial derivatives of the attribute surface at any discretization grid node.

In what follows, we describe a framework for approximating a continuous attribute surface from areal source data based on the premise that the smoothness of that underlying surface is known. The framework is general; we do not assume that the smoothness of the attribute surface is necessarily the one dictated by Laplace's PDE.

Pycnophylactic interpolation with a priori surface smoothness
In a continuous setting, the unknown solution surface [z] can be expressed as a convolution of a continuous punctual load intensity $[f] = \{f(\mathbf{u}'), \mathbf{u}' \in \Omega\}$ with a kernel or Green's function $G(\mathbf{u}, \mathbf{u}')$ as

$$z(\mathbf{u}) = \int_{-\infty}^{\infty} G(\mathbf{u}, \mathbf{u}') f(\mathbf{u}') d\mathbf{u}', \, \mathbf{u} \in \Omega \qquad (1)$$

where the Green's function $G(\mathbf{u}, \mathbf{u}')$ quantifies how an attribute value placed in isolation at point location $\mathbf{u}'$ is "dissipated" to any other location $\mathbf{u}$ within the domain Ω (Greenberg 1971; Briggs 1974). The unknown attribute value $z(\mathbf{u})$ can therefore be interpreted as an infinite weighted superposition of Green's functions $G(\mathbf{u}, \mathbf{u}')$, with source values $f(\mathbf{u}')$ in the study domain Ω playing the role of weights. We postpone the discussion on the choice of a particular functional form for the Green's function to a subsequent section. For now, one simply needs to assume that the Green's function is known a priori. As one would expect, however, it is this Green's function that ultimately dictates the smoothness of the unknown point attribute surface, for a given or fixed set of areal source data (including their supports).

In the finite case with K point sources $\{f(\mathbf{u}_k), k = 1, \ldots, K\}$, the discrete approximation to the continuous surface in equation (1) is written as

$$z^*(\mathbf{u}_p) = \sum_{k=1}^{K} G(\mathbf{u}_p, \mathbf{u}_k) f(\mathbf{u}_k) + \xi, \, p = 1, \ldots, P \qquad (2)$$

(Kitanidis 1999), where $z^*(\mathbf{u}_p)$ not $\hat{z}(\mathbf{u}_p)$ denotes the predicted attribute value at discretization location $\mathbf{u}_p$ from point, not areal, source data. Term ξ denotes the solution to a homogeneous boundary condition, where the form of the free-space Green's function $G(\mathbf{u}_p, \mathbf{u}'_k)$ is determined by the shape (or dimensionality) of the study domain Ω. This solution yields a discrete approximation $\{z^*(\mathbf{u}_p), p = 1, \ldots, P\}$

to a solution surface $[z^*] = \{z^*(\mathbf{u}), \mathbf{u} \in \Omega\}$ as long as the Green's function satisfies the following condition: $-\nabla^2[G(\mathbf{u}_p, \mathbf{u}_k)] = \delta(\mathbf{u}_p, \mathbf{u}_k)$ for $k = 1, \ldots, K$, where $\delta(\cdot)$ denotes the Dirac delta function (Barton 1989).

When equation (2) is used to predict attribute values at observation points, that is, when $\mathbf{u}_p$ coincides with $\mathbf{u}_k$, interpolation ensures reproduction of the corresponding datum $z(\mathbf{u}_k)$ as

$$\sum_{k'=1}^{K} G(\mathbf{u}_k, \mathbf{u}_{k'})f(\mathbf{u}_{k'}) + \xi = z(\mathbf{u}_k), \ k = 1, \ldots, K \tag{3}$$

under the constraint that the sum of the point source terms $f(\mathbf{u}_{k'})$ be equal to zero, because of the infinite boundary condition (Kitanidis 1999)

$$\sum_{k'=1}^{K} f(\mathbf{u}_{k'}) = 0 \tag{4}$$

The two constraints in equations (3) and (4) (data reproduction and a boundary condition at infinity) constitute a constrained system of equations. The solution of this system of equations yields the $(K+1)$ coefficients $\{f(\mathbf{u}_k), k = 1, \ldots, K, \xi\}$ of the point source interpolator in equation (2). The resulting surface $[z^*]$ constitutes a solution to the PDE corresponding to the particular Green's function $G(\mathbf{u}, \mathbf{u}')$, subject to the prescribed boundary condition at infinity, except at source data locations where singularities occur (Kitanidis 1999). As previously noted, the form of the Green's function is yet to be specified.

When sources represent areal data $\{z(v_k), k = 1, \ldots, K\}$ rather than point values $\{z(\mathbf{u}_k), k = 1, \ldots, K\}$, one needs to replace the point-to-point Green's function $G(\mathbf{u}_p, \mathbf{u}_k)$ in equations (1)–(3) by an integrated ATP Green's function $G(\mathbf{u}_p, v_k) = \int_{\mathbf{u} \in \Omega} g_k(\mathbf{u}) G(\mathbf{u}_p, \mathbf{u}) d\mathbf{u}$; in essence, the Green's function $G(\mathbf{u}_p, v_k)$ associated with an isolated source support v_k is the sampling-function-weighted integral of the Green's functions $G(\mathbf{u}_p, \mathbf{u})$ associated with individual point supports $\mathbf{u} \in v_k$. The solution of the new constrained system of equations yields the coefficients $\{f(v_k), k = 1, \ldots, K, \xi_v\}$ for surface construction based on the K areal data. The discrete approximation of the resulting surface $[\hat{z}]$ is derived from a modified version of equation (2) as

$$\hat{z}(\mathbf{u}_p) = \sum_{k=1}^{K} G(\mathbf{u}_p, v_k)f(v_k) + \xi_v \tag{5}$$

and ensures areal data reproduction; that is, it satisfies the pycnophylactic constraint (Kyriakidis 2004). In addition, the resulting surface satisfies the PDE corresponding to the Green's function adopted only outside the support of each areal datum (Matheron 1971), but provides the closest possible approximation to a solution surface under the K areal data reproduction constraints.

Equivalence with dual kriging

It is rather straightforward to identify equation (2) with the dual form of ordinary kriging with a known, and not necessarily stationary, variogram (or generalized covariance). Generally speaking, kriging is a family of geostatistical spatial interpolation algorithms with both a deterministic and a stochastic interpretation; its dual form is the expression linking kriging and spline interpolation, and the reader should consult Goovaerts (1997), Kitanidis (1997), or Chilès and Delfiner (1999) for further details. In equation (2), the supports of both source and target (unknown) data represent points; hence, the nonstationary Green's function $G(\mathbf{u}_p, \mathbf{u}_k)$ plays the role of a nonstationary generalized covariance between a target point $\mathbf{u}_p$ and a source point $\mathbf{u}_k$. Along the same lines, equation (5) is the dual form of ordinary kriging when point source data are replaced with areal support data, also termed ATP kriging (Kyriakidis 2004). In this case, the nonstationary Green's function $G(\mathbf{u}_p, v_k)$ plays the role of a nonstationary generalized covariance between a target point $\mathbf{u}_p$ and a source support v_k. In equation (2), the point source term $f(\mathbf{u}_k)$ plays the role of a dual point-to-point kriging weight, whereas in equation (5), the areal support source term $f(v_k)$ plays the role of the dual ATP kriging weight.

Up to this point, we have shown that pycnophylactic interpolation can be readily performed once a Green's function or solution kernel has been identified for the particular PDE and the boundary value problem at hand; we have not discussed, however, the particular form of that kernel. This is the equivalent problem of identifying an appropriate point variogram model for performing point-to-point or ATP kriging in a geostatistical framework.

Green's function, Laplace's PDE, and the logarithmic variogram model

The choice of a particular functional form for the Green's function depends on the smoothness assumption postulated for the unknown surface [z]; it is this smoothness that a particular PDE adopted for [z] furnishes. It is well-known that for the diffusion PDE the associated Green's function is the classical Gaussian kernel. The only difference between the Green's function and an ordinary (e.g., Gaussian) kernel is that the former changes depending on where in the domain Ω the source location $\mathbf{u}'$ is situated (e.g., close to boundaries). In other words, boundary conditions are embedded in the Green's function, hence, the nonstationary notation $G(\mathbf{u}, \mathbf{u}')$ instead of $G(\mathbf{u} - \mathbf{u}')$. In the case of a boundary at infinity, a Dirichlet-type boundary condition is equivalent to a Neumann-type boundary condition, and one can then define a free-space Green's function whose derivatives vanish as $\mathbf{u}$ moves infinitely far from a source location $\mathbf{u}'$.

Generally speaking, the Green's function for a boundary value problem cannot be expressed analytically, because either the problem has no solution or the solution is not unique (Barton 1989). Only in particular cases of homogeneous boundary conditions and for geometrically simple domains can one derive analytically the Green's function associated with a PDE and a boundary value problem.

More specifically, that Green's function, if it exists, can be written as $G(\mathbf{u}, \mathbf{u}') = G_0(\mathbf{u}, \mathbf{u}') + G_R(\mathbf{u}, \mathbf{u}')$, where $G_0(\cdot)$ and $G_R(\cdot)$ denote, respectively, the free-space Green's function and the region-dependent Green's function (Kantorovich and Krylov 1958; Greenberg 1971; Barton 1989; Chilès and Delfiner 1999). The free-space Green's function is a particular or fundamental solution to the boundary value problem given in equation (2) when boundary conditions are specified at infinity. In this case, the solution is independent of any boundary condition, hence, independent of location and just a function of distance (isotropic); its mathematical form depends on the dimensionality of the problem (Roach 1970; Greenberg 1971; Barton 1989).

The Green's function for the case of Laplace's PDE in two dimensions is a logarithmic function of distance; see, for example, Greenberg (1971):

$$G_0(\mathbf{u}, \mathbf{u}') = G_0(||\mathbf{u} - \mathbf{u}'||) = -\frac{1}{2\pi} \log(||\mathbf{u} - \mathbf{u}'||) \tag{6}$$

where $||\mathbf{u} - \mathbf{u}'||$ denotes the norm of vector $\mathbf{u} - \mathbf{u}'$. Note that the Green's function for a biharmonic PDE in two dimensions, a case also considered in Tobler (1979), involves the logarithmic distance and the distance squared: $||\mathbf{u} - \mathbf{u}'||^2 \log (||\mathbf{u} - \mathbf{u}'||)$; see, for example, Selvadurai (2001).

Equivalence with the logarithmic variogram model

Equation (6) corresponds to the geostatistical notion of a logarithmic generalized covariance function, a logarithmic variogram model in simpler terms, defined as: $\psi(||\mathbf{u} - \mathbf{u}'||) = \varsigma - \gamma(||\mathbf{u} - \mathbf{u}'||)$, where ς denotes an arbitrary constant and $\gamma(||\mathbf{u} - \mathbf{u}'||)$ denotes an ordinary variogram function (Matheron 1971; Kitanidis 1999). The reader should consult Journel and Huijbregts (1978), Kitanidis (1997), Chilès and Delfiner (1999), and Wackernagel (2003) for more details on generalized covariance functions. It is this equivalence that behooves us to review briefly hereafter the logarithmic variogram model and its properties, little known in the geographical spatial analysis literature.

More precisely, the logarithmic variogram model, also known as the de Wijsian or de Wij model in geostatistics, is defined as

$$\gamma(\mathbf{h}) = 3\alpha \log(||\mathbf{h}||) \text{ for } 0 \leq \alpha < 1 \text{ and } ||\mathbf{h}|| \neq 0 \tag{7}$$

(Krige 1978; Rendu 1981; Chilès and Delfiner 1999), where the parameter α is associated with the pure dispersion state of the phenomenon under study and is independent of the data support (Matheron 1971). In a logarithmic plot of the de Wijsian variogram model, where the variogram values are plotted against the logarithm of lag distances, the slope of the linear model is determined by the *absolute dispersion* α, and this value becomes unity for a process with a spatially random distribution (Krige 1978; Wackernagel, Thiery, and Grzebyk 1999). Plots of the logarithmic variogram model with different α parameters are shown in both linear and log scale in Fig. 1.
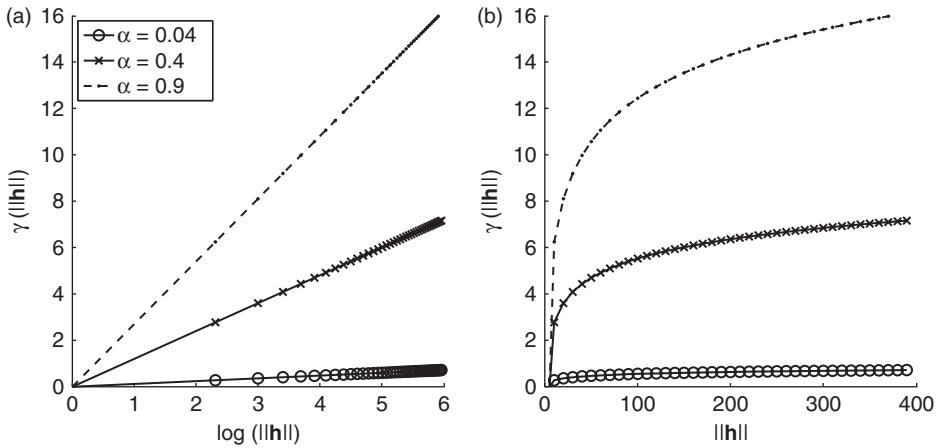
Figure 1. Illustration of logarithmic or de Wijsian variograms with different *absolute dispersion* parameters. The solid lines with circles and x-marks denote, respectively, the de Wijsian model with absolute dispersion parameters 0.04 and 0.4; the dashed-dotted line denotes the variogram model with $\alpha = 0.9$. (a) Plots of de Wijsian variograms versus distance on log scale. (b) Plots of de Wijsian variograms versus distance.

From a practical point of view, it is important to note that the punctual de Wijsian model in equation (7) becomes inapplicable for small distances, because $\log(||\mathbf{h}||)$ approaches to $-\infty$ as the distance $||\mathbf{h}||$ tends to 0. Indeed, the de Wijsian model is not the variogram of a random function but rather a distribution or a random measure at the point level (Matheron 1971). This indicates that the de Wijsian model can represent only properties of samples with support of finite dimension, and only as a regularized variogram (pertaining to nonpoint supports, such as pixels). Regularized variograms are generally approximated via numerical integration, and analytical expressions are only available for selected models in a limited domain (Matheron 1971; Journel and Huijbregts 1978; Rendu 1981). Kitanidis (1999) suggests an alternative form of the regularized de Wijsian variogram model: $\gamma(\mathbf{h}) = 3\alpha \log\left(\frac{\sqrt{||\mathbf{h}||^2 + \kappa^2}}{\kappa} \right)$ where κ is a microstructure parameter interpreted as the radius of curvature of the interpolated function at any observation location. The motivation behind this expression is to avoid singularities at observation locations, and, in practice, one-tenth of the length of the smallest support is recommended for κ.

The logarithmic variogram model has convenient analytical properties that can be used to compute the dispersion variance; this quantity can in turn be used to estimate the only parameter of the model, the absolute dispersion α, from the sample data (Chilès and Delfiner 1999). More precisely, the variance of attribute values defined on a support v within the entire domain Ω, denoted as $\sigma^2(v|\Omega)$ is referred to as the dispersion variance and is associated with the following two

observations: the dispersion around the mean value of a set of data collected within a domain Ω increases with the dimension of Ω, and the dispersion within a fixed domain Ω decreases as the support size $|v|$ increases. When the de Wijsian model is considered, the dispersion variance of the attribute values defined on a support v within the study domain Ω is analytically derived by the ratio between the size of the support $|v|$ and that of the domain $|\Omega|$ as

$$\sigma^2(v|\Omega) = \gamma(\Omega, \Omega) - \gamma(v, v) = 3\alpha \log\left(\frac{|\Omega|}{|v|}\right) \tag{8}$$

where $\gamma(\Omega, \Omega)$ denotes the average point variogram values over the entire region Ω, quantified as $\gamma(\Omega, \Omega) = \frac{1}{|\Omega|^2} \int_{\mathbf{u} \in \Omega} \int_{\mathbf{u}' \in \Omega} \gamma(\mathbf{u} - \mathbf{u}') d\mathbf{u} d\mathbf{u}'$, and $\gamma(v, v) = \frac{1}{|v|^2} \int_{\mathbf{u} \in v} \int_{\mathbf{u}' \in v} \gamma$ $(\mathbf{u} - \mathbf{u}') d\mathbf{u} d\mathbf{u}'$ is the average point-support variogram values within the support v. It should be noted, however, that this condition is quite restrictive because it requires geometrically similar supports within the study region Ω.

The above relationship (the de Wijsian formula) implies strong self-similarity: the variance of the small support v within the large support Ω depends only on the ratio $|\Omega|/|v|$, regardless of the actual scale. This principle of strong similarity also allows connecting the de Wijsian model to fractals (Chilès and Delfiner 1999). In addition, once the experimental variance $\hat{\sigma}_Z^2$ is computed from the sample data and is used as an estimate of the dispersion variance, the absolute dispersion can be derived as

$$3\hat{\alpha} = \frac{\sigma^2(v|\Omega)}{\log(|\Omega|/|v|)} \approx \frac{\hat{\sigma}_Z^2}{\log(|\Omega|/|v|)} \tag{9}$$

(Wackernagel, Thiery, and Grzebyk 1999), where $3\hat{\alpha}$ denotes the estimated value of absolute dispersion.

Accounting for boundary conditions and non-negativity constraints

Regardless of the particular Green's function associated with PDEs, such equations are typically solved subject to boundary conditions; indeed, Tobler's Laplacian smooth pycnophylactic interpolation involves a boundary condition as an input. The effect of boundary conditions on the construction of a smooth surface varies with application, but the presence of a physical boundary is likely to influence the interpolation results. In demographic applications, for example, the estimation of the spatial distribution of population density is expected to be different from one that does not consider boundary conditions; in the former case, the smoothness of the resulting surface is affected at locations not only nearby the boundary but also within the region's interior.

In the geostatistical approach, one can account for boundary conditions by assigning appropriate values, either known attribute values (for a Dirichlet-type condition) or known attribute derivates (for a Neumann-type condition), to a set of locations discretizing the boundary of the study domain. These fictitious additional point data can then be accounted for in the interpolation procedure via cokriging,

after defining appropriately the covariances between boundary locations and areal source supports, based, once again, on the postulated variogram model of the point attribute surface (Yoo and Kyriakidis 2008). The calculation of the corresponding prediction error variances follows similar modifications, and accounts for both the original areal source units and the set of points used to discretize a domain boundary.

Another attribute constraint in demographic applications is the non-negativity of predictions, because negative density values have no physical meaning. In Tobler's interpolation method, the non-negativity of the solution surface is enforced via an iterative procedure that prevents negative predictions simultaneously with mass-preservation and fidelity to boundary conditions. In the geostatistical framework, inequality and equality constraints can be accounted for by ATP kriging. Yoo and Kyriakidis (2006), for example, applied quadratic programming algorithms to ATP kriging with various types of inequality constraints, while accounting for support differences between source data and target predictions and satisfying the pycnophylactic constraint.

Quadratic programming algorithms, however, do not always converge to an optimal solution, particularly when global constraints are imposed or an un-bounded variogram like the de Wijsian model is used. For these reasons (especially the latter), an iterative cokriging approach was used in the current study to ensure non-negative predictions. Initially, a set of prediction locations with negative population density obtained from ATP cokriging is identified. Then, the quadratic programming algorithmsare selectively applied to any areal support containing such negative predictions. This selective application of the quadratic programming algorithms for handling the non-negativity constraint is specific to the problem at hand and does not guarantee convergence, although it allows us to take into account support differences and to satisfy the pycnophylactic condition. The biggest disadvantage of this geostatistical approach lies in the construction of confidence intervals for the final predictions: although the actual predictions are guaranteed to be non-negative, this is not true for the entire conditional distribution modeling the uncertainty about the unknown point attribute value given the areal data (Yoo and Kyriakidis 2006).

Case study

The case study uses population data collected over irregular polygons to demonstrate empirically that the population density surface obtained from the geostatistical approach using a logarithmic generalized covariance function is very similar to that obtained from Tobler's Laplacian smooth pycnophylactic interpolation. Population data were collected in 1970 for $K = 18$ census tracts in Ann Arbor, Michigan, and were used by Tobler (1979). Here, we use population densities (rather than counts) at the census tract level as areal data, and they are shown in Fig. 2a with a choropleth map. Univariate statistics of the population density data at
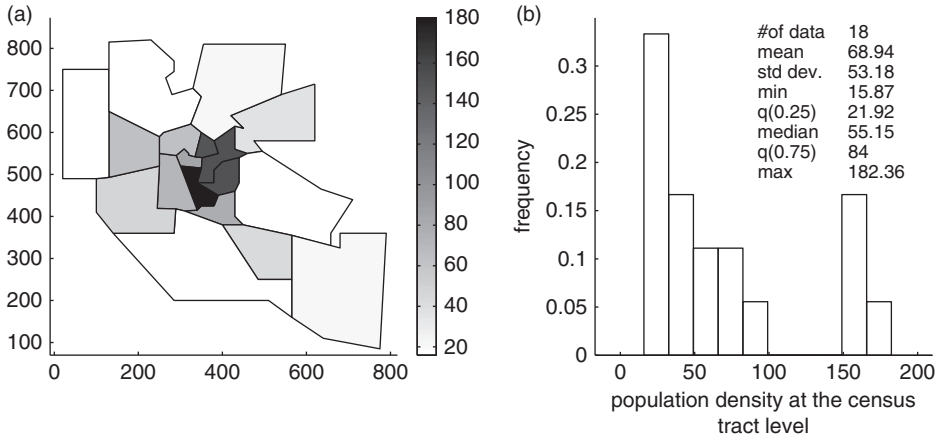
Figure 2. (a) Choropleth map representation of population density at the census tracts in Ann Arbor (1970). (b) Histogram and summary statistics of the mean population density values at census tracts.

the census tract level are presented in Fig. 2b via a histogram; the average population density at the census tract level is 68.94, and the variance is 53.18^2; all density values are between 15.87 and 182.36.

Population density surfaces are constructed using the geostatistical approach and Tobler's approach for the case of an unbounded domain where the boundary condition is implicitly placed at infinity (see section ''Unconstrained population density surface'') and for the case of a Dirichlet-type boundary condition where zero population density is known outside the study region together with a non-negativity constraint (see section ''Constrained population density surface''). In all cases, we discretize the solution surface using a (73×73) regular grid with a (11×11) cell size; this grid specification is fairly close to that of the original work by Tobler (1979). Generally speaking, the quality of interpolation relies on the discretization scheme, but the current choice of discretization is inconsequential because the main goal of this study is to examine the relative differences between the resulting density surfaces. All results obtained via Tobler's interpolation method involved 400 iterations of the averaging processes.

As mentioned in the previous section, the variogram model corresponding to the free-space solution of Laplace's equation in 2D is the de Wijsian model, which is parameter free except for the absolute dispersion, determined solely by the variance of logarithmic attribute values and the geometric configuration of the source supports. In our case, however, census tracts are irregular and of unequal size, hence, there is no unique logarithmic ratio value $\log (|\Omega|/|v|)$ to use in equation (9); instead such logarithmic ratios range from 1.87 to 5.27 with a mean of 3.21. Similar to practical demonstrations on the use of the de Wijsian variogram for different support sizes (Krige 1978), we adopt the mean ratio value as a
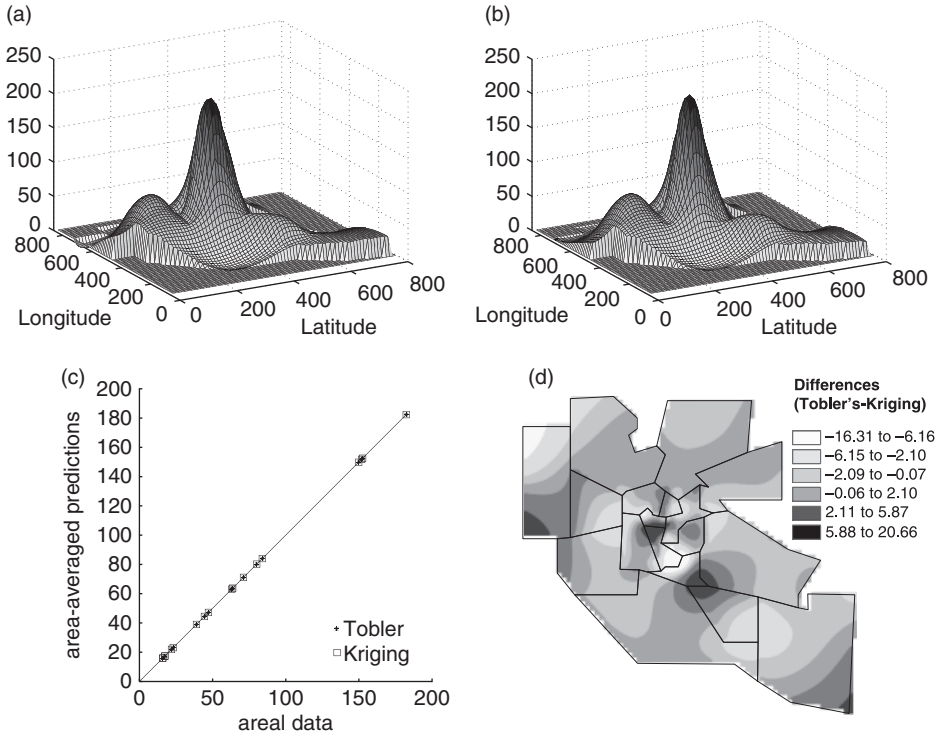
Figure 3. (a) Tobler's solution surface under a Neumann-type boundary condition. (b) Area-to-point (ATP) kriging prediction surface with boundary conditions at infinity. (c) Scatter-plot of original areal data (population density at the census tract level of Fig. 2a) versus area-averaged point predictions derived by Tobler's method (asterisks) and by kriging (squares). (d) Map of differences between point predictions obtained by Tobler's interpolation method (a) and ATP kriging (b).

representative geometric configuration. Based on a variance of logarithmic population density 0.72 computed from the $K = 18$ areal data, the absolute dispersion is approximated from equation (9) as $3\alpha \approx 0.72/3.21 = 0.22$. Note that ATP ordinary kriging predictions using the de Wijsian model are independent of the sill (or scale) parameter, but the corresponding prediction variances are not.

Unconstrained population density surface

In what follows, we aim to reconstruct a smooth surface of population density under the sole constraint that the original population densities computed over irregularly shaped geographical regions are reproduced when the values of that surface are reaggregated within such regions. Using a logarithmic variogram model, stemming from the Laplacian smoothness criterion of Tobler's pycnophylactic interpolation, ATP ordinary kriging yields a smoothly varying population density surface Fig. 3b. A very similar density surface (Fig. 3a) is obtained

via Tobler's method without enforcing the non-negativity constraint, and using a Neumann-type boundary condition where zero derivatives of population density are prescribed along a set of discrete point locations on the boundary. As shown in Fig. 3c, both density surfaces are pycnophylactic, that is, they reproduce the areal data when the predicted values are reaggregated within each support. Figs. 3a and 3b indicate that the two interpolated population density surfaces reproduce the major characteristics of the areal data; high population density at the center of the study domain in accordance with Fig. 2a, but abrupt changes between tract boundaries disappear. From a visual inspection, the two maps in Figs. 3a and 3b appear almost identical to each other. Relatively small differences exist, however, as is shown in Fig. 3d. Tobler's method, for example, predicts lower population density in the north-west part of the study region but higher population densities at the center (see Fig. 3d); this may be due to the Neumann-type boundary condition imposed on Tobler's pycnophylactic interpolation surface.

One of the advantages of the geostatistical approach over Tobler's approach is that the former allows the assessment of the uncertainty associated with each point value of the predicted surface. Figs. 4a and 4b show the maps of ATP ordinary kriging standard deviations associated with the ATP predictions of Fig. 3b for different values of absolute dispersion α. As mentioned in the section on ''Green's function, Laplace's PDE, and the logarithmic variogram model,'' α is the only parameter of the de Wijsian model which can be used to model the pure dispersion state of the phenomenon under study. Fig. 4b shows kriging standard deviations with unity absolute dispersion, that is, $\alpha \approx 1$, as an extreme case of spatially random distribution, and Fig. 4a is the map of kriging standard deviations with α corresponding to an absolute dispersion calculated from the geometric configuration of the study domain with the estimated variance of population density at the census tract level; this yields $3\alpha = 0.23$. As expected from theory, this parameter
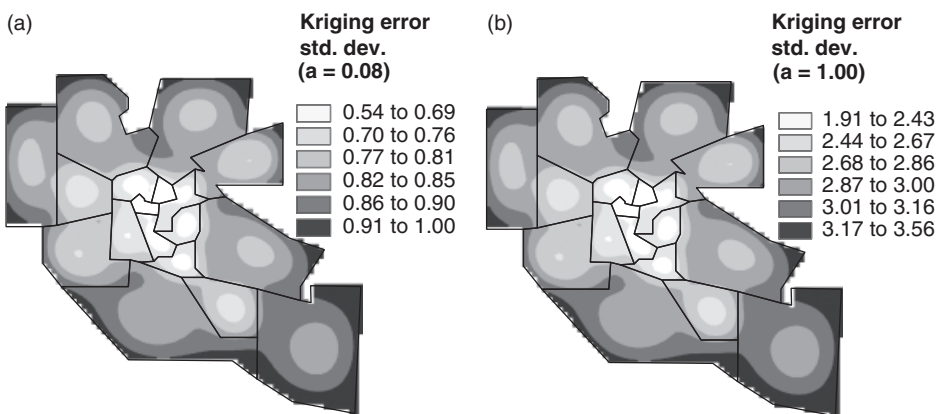

Figure 4. Error standard deviations associated with the area-to-point kriging predictions of Fig. 3b with absolute dispersion $\alpha \approx 0.08$ (a) and $\alpha \approx 1$ (b).

does not affect the kriging predictions shown in Fig. 3b, but alters significantly the uncertainty associated with these predictions, as indicated by the range of kriging standard deviation values; note the differences in the legend of maps in Figs. 4a and 4b. The interpretation of the kriging standard deviation as a single uncertainty measure requires some caution, however, because it is only a function of the spatial configuration of source and target supports and the point covariance model, not of the actual data values themselves.

Constrained population density surface

To construct a population density surface while accounting for a Dirichlet-type boundary condition, we use dual ATP ordinary cokriging using the 18 areal data with a total of 283 additional points along the edge of the study area. The resulting prediction surface is shown in Fig. 5b, along with the corresponding Tobler's solution surface for comparison in Fig. 5a. In both cases, the predicted population
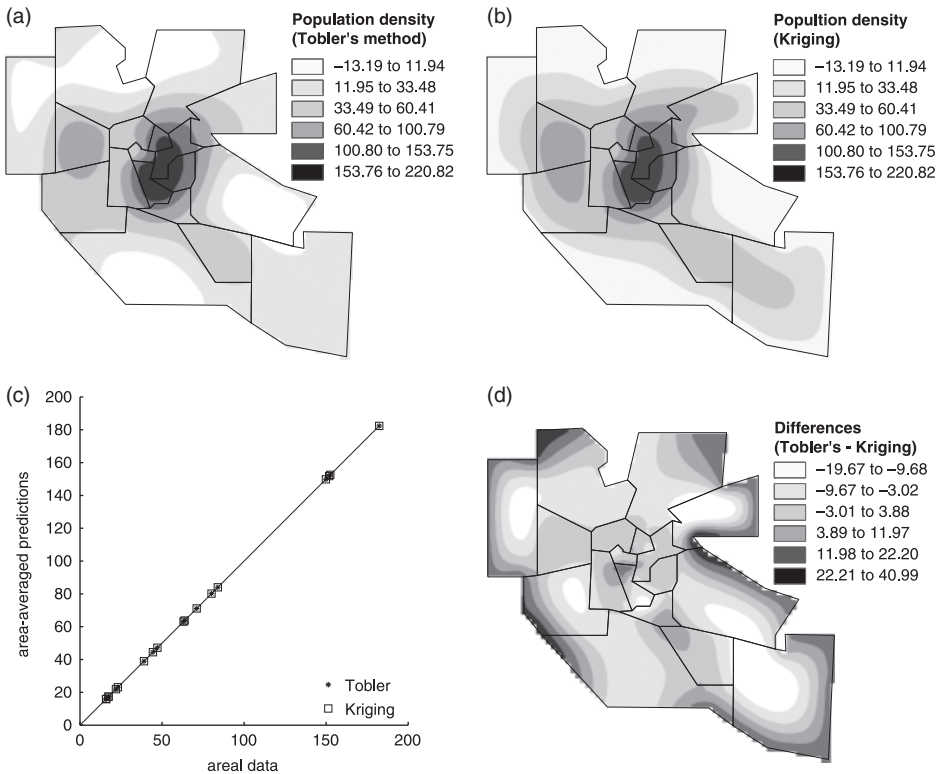

Figure 5. Population density surface derived from Tobler's interpolation method (a) and from area-to-point (ATP) cokriging (b), both under a Dirichlet-type boundary condition. (c) Scatter plot of original areal data (of Fig. 2a) versus area-averaged point predictions derived by Tobler's method (asterisks) and by kriging (squares). (d) Map of differences between point predictions obtained by Tobler's interpolation method (a) and ATP cokriging (b).
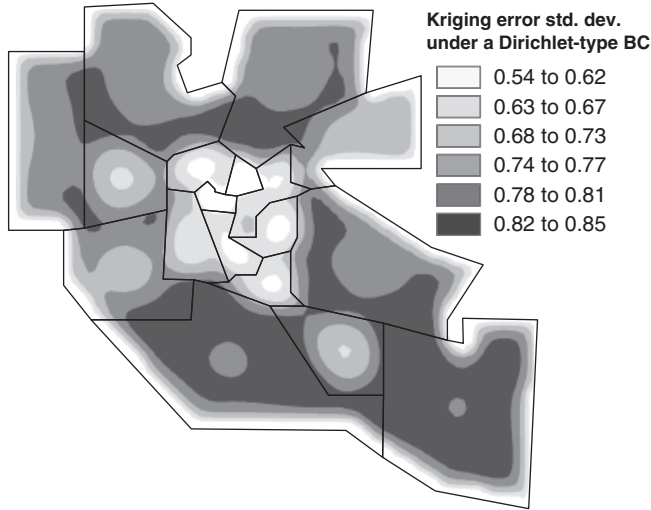
Figure 6. Error standard deviations associated with the area-to-point cokriging predictions subject to a Dirichlet-type boundary condition, shown in Fig. 5b.

density values along the domain boundary are very similar to those prescribed by the boundary condition, here zeros. As shown in Fig. 5c, both density surfaces are pycnophylactic. The two predicted surfaces are fairly similar to each other, yet relatively small discrepancies do exist between them as shown in the difference map of Fig. 5d. Such discrepancies may arise from the differences between the numerical approximation algorithm used in Tobler's interpolation and the numerical integration of the de Wijsian variogram model, but it could be also due to the different locations of discrete points representing the boundary. The differences range from -19.67 to 40.99, which are relatively small considering that the total range of density is -13.19 to 220.82. The major differences are found near the border due to the different handling of the boundary condition (Fig. 5d).

The impact of the boundary condition on the construction of a smooth surface may be better illustrated via the map of kriging error standard deviations shown in Fig. 6. Relatively smaller standard deviations are observed near the boundary of the study area compared with those in the map of kriging standard deviations with a boundary condition at infinity and $\alpha = 0.08$ (Fig. 4a). Also, relatively smaller standard deviations are found at the center of study area, where census tracts have smaller supports.

We also perform surface reconstruction subject to a Dirichlet-type boundary condition, plus an inequality constraint, that of non-negative predicted density values. Recall that in ATP cokriging prediction without a non-negativity constraint (Figs. 3b and 5b), the smooth density surface contains some negative values, although their magnitude is relatively small. To bypass this problem, a two-step procedure is adopted. Initially, a set of prediction locations with negative population
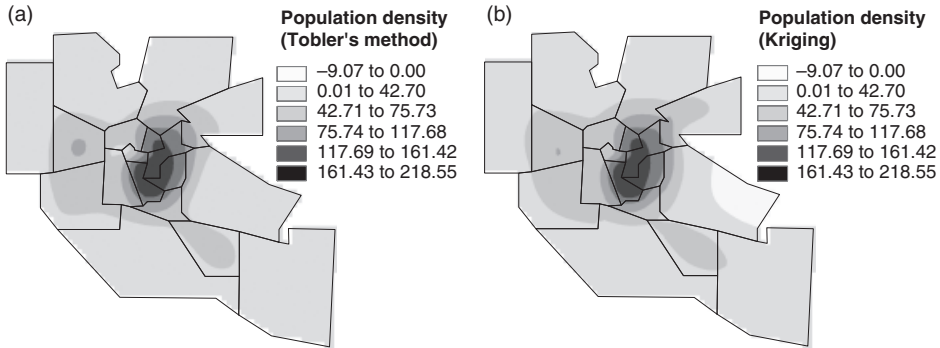
Figure 7. Population density surface predicted from Tobler's interpolation method (a) and from areal-to-point cokriging (b), both subject to a Dirichlet-type boundary condition and non-negativity constraint.

density values is identified, and then the quadratic programming algorithms are selectively applied to any areal support containing such negative predictions. Unfortunately, quadratic programming algorithms converge at only three among four census tracts; the updated final results are shown in Fig. 7b. This selective application of quadratic programming algorithms for handling the non-negativity constraint is specific to the problem at hand and does not guarantee convergence; it allows us to take into account support differences and to satisfy the pycnophylactic condition. Fig. 7a shows the corresponding density surface constructed using Tobler's approach for comparison. The two predicted surfaces are fairly similar to each other, yet relatively small discrepancies do exist between the two. Again, such differences are most likely due to the different discretization schemes used by the two methods to account for the particular (here Dirichlet) boundary condition.

Last, we perform ATP cokriging, subject to a Dirichlet-type boundary condition, using alternative, judiciously selected point variogram models to assess the impact of the choice of that model on the kriging-derived solution surfaces. These models (shown in Fig. 8) were selected based on their similarity, particularly in terms of their shape near the origin, to the de Wijsian model, and they include (1) an exponential model with effective range 100 distance units and sill 0.85; (2) a spherical model with effective range 100 distance units and sill 0.85; and (3) a Gaussian model with effective range 300, partial sill 1.05, and a very small nugget effect; this model's shape near the origin gradually deviates from that of the de Wijsian model.

The alternative kriging-derived surfaces (not shown) were compared in terms of their correlation coefficient, mean absolute error (MAE), and root mean squared error (RMSE) to Tobler's solution surface derived under the same boundary condition; the results are summarized in Table 1. The correlation coefficient between the alternative kriging prediction surfaces and Tobler's solution surface decreases as the shape of the point variogram model used in kriging deviates more
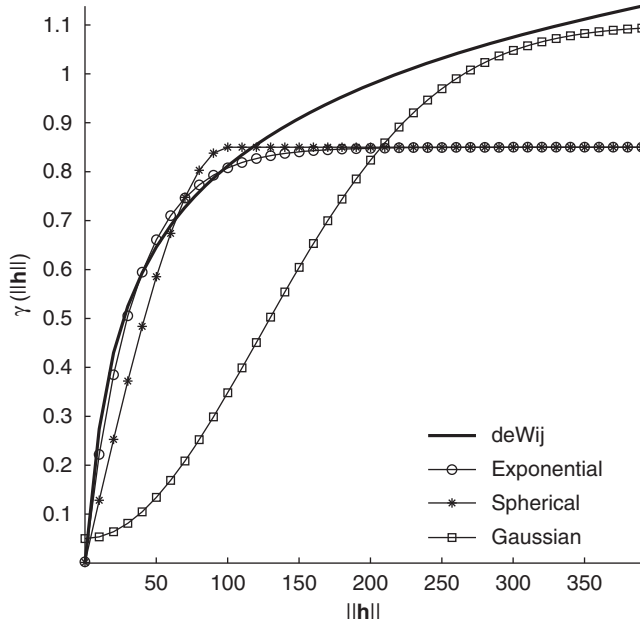
Figure 8. Alternative point variogram models used for area-to-point cokriging: de Wijsian model (solid line), exponential model with effective range 100 and sill 0.85 (solid line with circles), spherical model (solid line with stars) with same parameters as the exponential, and Gaussian model with effective range 300 and sill 1.05 plus a very small nugget effect (solid line with squares).

from the shape of the de Wijsian model at the origin. On the contrary, the corresponding summary measures of dissimilarity (MAE, RMSE) increase.

These results imply that the ATP ordinary kriging with the de Wijsian model provides the closest approximation to Tobler's solution surface among other point variogram models. As stated in the introduction, there is no a priori reason why one should use a de Wijsian variogram model to begin with. Here, we are illustrating, in practice, that the geostatistical approach can yield very similar density surfaces with Tobler's approach, but one is certainly not restricted to the de Wijsian model when working within the geostatistical framework.

Discussion and conclusions

As an alternative representation of census-based population data, that is, via choropleth mapping, we adopt a continuous population density surface model in which the resulting latent surface is free of arbitrary partitioning effects. Prior information about the smoothness on the unknown density surface is critical for reconstructing that surface from known areal data. As expected, the smoothness assumption imposed on the latent point field has a significant impact on the predicted surface visually, also reflects the spatial process model corresponding to

Table 1 Summary Statistics Comparing Tobler's Solution Surface with Alternative Kriging-Derived Surfaces Constructed Using Different Point Variogram Models

Statistic/Model	de Wijsian	Exponential	Spherical	Gaussian
Correlation coefficient	0.97	0.96	0.95	0.84
Mean absolute error	7.51	7.75	9.74	18.24
Root mean squared error	9.60	9.97	12.21	23.22

that surface. In particular, when little is known about the attribute surface except the aggregate source data, a continuous and smooth surface is considered as the choice of estimation (Tobler 1979; Kitanidis 1999); one can argue that even this is debatable (Chilès and Delfiner 1999).

In this article, we have demonstrated with a case study that the population density surface constructed via Tobler's pycnophylactic interpolation method is very similar (apart from numerical errors and discretization differences) with the geostatistical method of ATP kriging using a logarithmic point variogram model. We have also shown that the geostatistical method can be extended to incorporate prespecified boundary conditions as well as inequality constraints, if they exist. Combining such auxiliary information with areal data can improve the quality of the reconstructed surface, although some modifications of the proposed ATP kriging are necessary.

The main reason behind this demonstration of equivalence between the two interpolation methods is to draw attention to the smoothness criterion employed in Tobler's method. In other words, we believe it is this Laplacian smoothness criterion that should be the defining feature of Tobler's method, not its pycnophylactic property. We justify this statement by highlighting that geostatistical methods are by construction pycnophylactic and allows for a wide spectrum of smoothness criteria encapsulated in alternative point variogram models.

As a closing statement, we would like to make a call for users of spatial interpolation methods in geographical applications involving areal data to address the question of what is the particular variogram model (smoothness criterion) that one needs to employ for a particular attribute within a given spatial context or environment. One could attempt a deconvolution approach, such as the one proposed by Goovaerts (2008). That approach, however, relies on the notion of an average distance between supports, which in the case of drastically different unit shapes is an approximation. In addition, the sole dispersion parameter (α) of the logarithmic variogram model can be rigorously estimated only if abundant point-level data (e.g., microdata) exist. In either case, that parameter affects only the uncertainty in the interpolated surface and might be better estimated using theoretical (process-based) arguments. In conclusion, the particular smoothness criterion dictating the spatial distribution of the unknown attribute surface should be selected on the basis of a spatial process model. One such model is the Laplace PDE embedded in Tobler's pycnophylactic interpolation method; other spatial

process models exist, such as fractals or diffusion models. Once such a smoothness criterion (or equivalently punctual variogram function corresponding to a spatial process model) is selected, geostatistical methods should be the default choice for surface construction from areal source data, because they provide a means for assessing uncertainty in the predicted surface and can be used in a stochastic simulation mode for uncertainty propagation purposes.

Acknowledgements

References

Barton, G. (1989). *Elements of Green's Functions and Propagation.* New York: Oxford University Press.

Bracken, I. (1994). ''A Surface Model Approach to the Representation of Population-related Social Indicators.'' In *Spatial Analysis and GIS*, 247–59, edited by S. Fotheringham and P. Rogerson. London: Taylor and Francis.

Briggs, I. C. (1974). ''Machine Contouring Using Minimum Curvature.'' *Geophysics* 34, 39–48.

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

Goodchild, M. F., L. Anselin, and U. Deichmann. (1993). ''A Framework for the Areal Interpolation of Socioeconomic Data.'' *Environment and Planning A* 25, 383–97.

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

Goovaerts, P. (2006). ''Geostatistical Analysis of Disease Data: Accounting for Spatial Support and Population Density in the Isopleth Mapping of Cancer Mortality Risk Using Area-to-Point Poisson Kriging.'' *International Journal of Health Geographics* 5(52).

Goovaerts, P. (2008). ''Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units.'' *Mathematical Geosciences* 40(1), 101–28.

Gotway, C., and L. Young. (2002). ''Combining Incompatible Spatial Data.'' *Journal of the American Statistical Association* 97(458), 632–48.

Gotway, C. A., and L. J. Young. (2007). ''A Geostatistical Approach to Linking Geographically Aggregated Data from Different Sources.'' *Journal of Computational & Graphical Statistics* 16(1), 115–35.

Greenberg, M. D. (1971). *Application of Green's Functions in Science and Engineering.* Englewood Cliffs, NJ: Prentice Hall.

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

Kantorovich, L. V., and V. I. Krylov. (1958). *Approximate Methods of Higher Analysis.* New York: Interscience Publishers.

Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications in Hydrogeology.* New York: Cambridge University Press.

Kitanidis, P. K. (1999). ''Generalized Covariance Functions Associated with the Laplace Equation and their Use in Interpolation and Inverse Problems.'' *Water Resources Research* 35, 1361–67.

Krige, D. J. (1978). *Lognormal-de Wijsian Geostatistics for Ore Evaluation, Series in Geostatistics No.1.* Johannesburg, South Africa: Southern African Institute of Mining and Metallurgy.

Kyriakidis, P. C. (2004). ''A Geostatistical Framework for the Area-to-Point Spatial Interpolation.'' *Geographical Analysis* 36(3), 41–50.

Lam, N. S.-N. (1983). ''Spatial Interpolation Methods: A Review.'' *The American Cartographer* 10, 129–49.

Langford, M., and D. J. Unwin. (1994). ''Generating and Mapping Population Density Surfaces within a Geographical Information System.'' *The Cartographic Journal* 31, 21–26.

Martin, D. (1989). ''Mapping Population Data from Zone Centroid Locations.'' *Transactions of the Institute of British Geographers NS* 14(8), 90–97.

Martin, D. (1996). ''An Assessment of Surface and Zonal Models of Population.'' *International Journal of Geographical Information Systems* 10, 973–89.

Martin, D., N. J. Tate, and M. Langford. (2000). ''Refining Population Surface Models: Experiments with Northern Ireland Census Data.'' *Transactions in GIS* 4, 343–60.

Matheron, G. (1971). The Theory of Regionalized Variables and its Applications, Centre de Géostatistique de l'Ecole des Mines de Paris, France. Fasc. 5, Ecole Nat. Sup. des Mines, Paris.

Rendu, J. M. (1981). *An Introduction to Geostatistical Methods of Mineral Evaluation, Series in Geostatistics No. 2.* Johannesburg, South Africa: South African Institute of Mining and Metallurgy.

Roach, G. (1970). *Green's Functions.* London: Cambridge University Press.

Ryan, J., H. Maoh, and P. Kanaroglou. (2009). ''Population Synthesis: Comparing the Major Techniques Using a Small, Complete Population of Firms.'' *Geographical Analysis* 41, 181–203.

Selvadurai, A. P. S. (2001). *Partial Differential Equations in Mechanics I: Fundamentals, Laplace's Equation, Diffusion Equation, Wave Equation.* Berlin, Germany: Springer Verlag.

Thurstain-Goodwin, M., and D. Unwin. (2000). ''Defining and Delineating the Central Areas of Towns for Statistical Monitoring Using Continuous Surface Representations.'' *Transactions in GIS* 4, 305–17.

Tobler, W. R. (1975). ''Linear Operators Applied to Areal Data.'' In *Display and Analysis of Areal Data*, 14–38, edited by J. Davis and M. McCullagh. London: Wiley.

Tobler, W. R. (1979). ''Smooth Pychnophylactic Interpolation for Geographical Regions.'' *Journal of the American Statistical Association* 74(367), 519–30.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications.* Berlin, Germany: Springer.

Wackernagel, H., L. Thiery, and M. Grzebyk. (1999). ''The Larsen Model from a de Wijsian Perspective.'' In *GeoENV II: Geostatistics for Environmental Applications*, 125–35, edited by J. Gomez-Hernandez, A. Soares, and R. Froidevaux. Dordrecht, The Netherlands: Kluwer.

Yoo, E.-H., and P. C. Kyriakidis. (2006). ''Area-to-Point Kriging with Inequality-Type Data.'' *Journal of Geographical Systems* 8, 357–90.

Yoo, E.-H., and P. C. Kyriakidis. (2008). ''Area-to-Point Predictions under Boundary Conditions.'' *Geographical Analysis* 40, 355–79.