

DRAFT

Spatial demographics with regression kriging: mapping and interpreting social factors of vulnerability to disaster in the Philippines

C. M. Armstrong^a, Marina Reilly-Collette^b, and Marissa Torres^b

^aERDC-CERL, 2902 Newmark Dr, Champaign, IL 61826, USA; ^bERDC-CRREL, 72 Lyme Rd, Hanover, NH 03755, USA

ARTICLE HISTORY

Compiled June 11, 2019

ABSTRACT

Spatial demographic data is the defining aspect of population geography and a necessary component for decision support. Spatial analysis often demand specialized methodology, and the physical size of spatial demographic data and the many different needs of its users have left it difficult to develop generalized methods and best practices for this data. The following discusses interpretation of spatial demographic data and reviews two methods that underlie most techniques for downscaling this data: area-to-point regression kriging and regression weighted dasymetric mapping. Both methods are ideal for demographic data because both allow modeling with a combination of geographic and demographic factors; an important feature for demographic data with extra-spatial influences. Both methods can also be used for spatial extrapolation. After reviewing methods the article uses regression kriging to map a social vulnerability index in the Philippines. The result is a coherent downscaling of social vulnerability to a one kilometer regular grid that facilitates visualization and combination of multiple spatially-explicit data sets. The paper concludes with a discussion of concerns that arise when fitting a spatial model using census data from one nation and applying that model to make inferences within the borders of another nation.

KEYWORDS

regression kriging; dasymetric mapping; change of support problem; spatial demographics; population geography

Maps of the human population are essential for demographic spatial analysis or to support decision-making with spatial components or consequences. The simplest population maps are based upon demographic data that is aggregated to predetermined boundaries. These are the *choropleth* maps, and for many decades they have been an essential tool of human geography. The drawback of choropleth maps is that the boundaries are often arbitrary and aggregations within the boundaries may not accurately reflect demographic realities. Unequal spatial sizes between bounded areas can also produce a visual bias where viewers perceive larger areas as more important. The problem of arbitrary boundaries is a very generic one, and known as the *modifiable areal unit problem* (MAUP).

The MAUP bears important statistical consequences as well. Openshaw and Taylor (1979) demonstrated that, given the privilege to arbitrarily

choose aggregations of counties in Iowa, they could manipulate the correlation between republican voting and proportion of elderly people in a county to between nearly -1 and 1 . Sometimes data is aggregated at a relatively fine scale of semi-regular enumerative boundaries—e.g. census blocks—and such data may avoid some problems of the MAUP. Another problem, however, is that spatial aggregations are incompatible with any data source not using the same aggregation boundaries. These two problems, the first of arbitrarily aggregated data and the second of incompatible aggregation boundaries, are both manifestations of a specialized variant of the MAUP known as the *change of support problem* (COSP) (Cressie, 1996; Gelfand, Zhu, & Carlin, 2001). Dasymetric mapping and kriging, which the following article will compare, are methods for solving this problem.

Combining data with different spatial supports is a non-trivial but common problem that occurs when at least one of multiple data sets is associated with a set of polygon features. Regression weighted dasymetric mapping (RWDM) solves the COSP by downscaling data from a set of source polygons to target polygons or points, based upon statistical correlations between source and target features. Regression kriging (RK) also uses a regression model to estimate correlations between source and target features, but additionally specifies a spatial autocorrelation model that attempts to explain the residual error of the regression model as a function of spatial autocorrelation. The naïve method for fitting an autocorrelation function to polygons is to collapse them to their centroids, but this simplification ignores the different sizes and shapes of the polygons. Methods to accommodate area-to-area/point (A2A/P) spatial autocorrelation also exist, and will be briefly explained in this article.

The estimated quantities of mapped demographic data are typically expected values or probabilities, and not a physical measurement in space. Nevertheless, spatial population data represent an environmental reality and often can be linked to a real spatial quantity. For example, a map of the risk of childhood leukemia suggests underlying environmental causes that have a physical spatial reality. The presence of carcinogens in the environment contribute to an increased rate of childhood leukemia in an area of space, for example, and within this space induce spatial autocorrelation among incidences of the cancer. An alternative analysis could map concentrations of carcinogens—e.g. concentrations of heavy metals in soil samples. The two variables, risk of childhood leukemia and concentration of heavy metals, both could be down scaled to an arbitrary spatial support, even to a single point. Only heavy metals, however, are observable in a single sample. Risk cannot be directly observed, it is instead estimated from multiple observations and hence is already an aggregate value. This distinction may be the most important for understanding spatial demographic data.

Survey data is georeferenced to administrative or enumeration boundaries for purposes of convenience, privacy, and anonymity. Downscaling procedures for demographic data, both dasymetric and kriging, estimate rates or probability densities over areas of space, and could place spatial limitations on demographic attributes that narrow down the identities of people living in an area. The U.S. Census Bureau has found that georeferenced aggregated tabular data provides means to reconstruct real individuals in the microdata (Abowd, 2018; Abowd & Schmutte, 2019). The conclusion discusses complications that spatial demographic downscaling may pose for privacy of survey data, and additional concerns regarding use of personal data.

The goal of this article is to demonstrate a spatial demographic model in a typical application for the purposes of discussing benefits, difficulties, and caveats of these models in such applications. The discussion assumes familiarity with linear regression and its underlying linear algebra, but requires no knowledge of spatial modeling. The

first section examines the elements of spatial demographic model design, and reviews the history, applications, and interpretations of spatially explicit demographic data. The second section outlines the options available for fitting both RWDM and RK models and the criteria for selecting from among them. The third section demonstrates an RK model in an application of downscaling demographic data for the purpose of visualization. For this application, the model will downscale a social vulnerability index (SVI) from the municipalities where it is aggregated to a grid of roughly one square kilometers. The conclusion summarizes the article, highlights the benefits of down scaled demographic data, and touches upon technical and ethical concerns that users should be aware of when applying spatial models for demographic mapping.

1. Spatial demographic data and models

In any discipline where they might be useful, demographic maps have almost certainly been utilized. Geodemography is a practice oriented science of classifying spatial areas of population, often applied in business and public policy (see (Consolidated Analysis Center Incorporated, 2013-2018) for an example), but also with a parallel academic research arm (Jefferson Institute, 2017; Output Area Classification User Group, 2016; Singleton & Spielman, 2014). Medical geography create maps of demographic data to help identify and clarify spatial relationships between human medical health and the environment (Brown, McLafferty, & Moon, 2009). Population geography is a sub-field of human geography that specializes in generating maps of human population by demographic attributes (Bailey, 2005). Certainly countless other examples exist from the social, political, medical, and economic sciences. Nonetheless, whereas classical statistics are ubiquitous across social sciences, geospatial statistics remain a specialty technique.

One reason classical statistics may enjoy popularity is because it offers a generalized yet simple framework, and importantly is computationally efficient; ordinary least squares (OLS) offers a computational complexity of $O(C^2N)$ versus kriging complexity of $O(N^3)$, required to invert the $n \times n$ kriging covariance matrix. RWDM, which essentially is simple linear regression, can be performed with most contemporary personal computers, whereas RK consumes considerable computing resources and may demand high-performance hardware and software. This stark difference in physical limitations may require that data management and analyses of geospatial data be shifted from personal to cloud computer resources. The lack of a simple and generalized kriging framework, on par with the usability of those available for classical statistics, is also a serious obstacle that receives insufficient attention.

The following section discusses spatial demographic data and provides definitional equations. The first subsection briefly examines types of demographic data and their modeling implications. The second takes a closer look at the COSP. The third and final subsection compares the equations for RWDM and RK, and discusses their differences and similarities. These subsections will discuss properties of the data and models that will be referenced when this article analyzes a RK model for an application of mapping social vulnerability.

1.1. Types of demographic data

For modeling purposes, demographic data come in two types: continuous and discrete. Continuous data exists on the real number line, and thanks to statistical innovations

like the central limit theorem, most continuous data can be modeled using linear models and the normal distribution. Discrete variables may take only certain values; examples include binary data, proportions, and counts. Discrete data are not normally distributed but instead follow the binomial, multinomial, or Poisson distributions. Conventional regression and kriging models assume normally distributed data, and require modifications to work appropriately with discrete data. Regression models fit to variables with two or more discrete classes are known as binomial or multinomial respectively. Indicator kriging is a methodology for fitting kriging models to discretized data, and Goovaerts (2010) recommends variations of indicator kriging for binomial and ordinal spatial demographic data.

RWDM is essentially a regression model fit to georeferenced data. OLS is appropriate for continuous data, and discrete models are available within the framework of generalized linear modeling (GzLM). Using GzLM, one supplies a link function and distribution. The link function linearizes the data, and the distribution describes the distribution of the errors in a model of the data. GzLM fits a model to the data using iteratively reweighted least squares to maximize the likelihood for the given distribution (Nelder & Wedderburn, 1972). OLS is a form of GzLM, with an identity link and normally distributed errors.

Extending kriging models is more difficult, and a generalized framework for both linear and non-linear models doesn't seem to exist for kriging. Kriging may be employed for discrete data, but it may not correctly model the variance properties (Goovaerts, 2010). Goovaerts addressed the issue of kriging small samples of Poisson distributed data by adding a Poisson variance term to the kriging equations that represents the intrinsic unreliability of small counts (Goovaerts, 2006). A like method is applicable to binomial data, replacing the Poisson-distributed variance with binomial variance (Goovaerts, 2004, 2009). Indicator kriging (IK) is another framework for discrete data that was designed to analyze discretizations of continuous data (Journel, 1983). Outliers are pathological for kriging models, but may be removed by discretizing the data as either above or below a threshold. Multiple discrete data sets can be generated for different threshold values, and all data sets cokriged—a multivariate kriging, similar to multivariate regression. Binomial cokriging is a similar method designed for kriging proportions (Lajaunie, 1991; Oliver, 1998), where unobserved continuously varying probability of incidence at some point in space $x(s_0)$ is estimated along with observed aggregated proportions.

Specialized extensions of kriging notwithstanding, the performance of basic kriging models with discrete data is often sufficient, and at times more stable than more complex models (Oliver, 1998). Kriging of binomial data, however, may yield results outside the range of $[0, 1]$. This occurs where samples of data are tightly clustered. Kriging declusters data by distributing weight equally around the clustered points, and the optimization often assigns negative weights to some points of a cluster to balance large positive weights assigned to others. Negative weights can yield results outside observed data ranges, and in general disturb performance of a kriging model (Duetsch, 1996). One can avoid negative weights by using spatial variance-over-distance functions, the so-called variogram, that increase quickly. Using a Gaussian cumulative distribution curve for a variogram function will overweight nearby samples due to the flatness of the curve close the origin: the curve models nearby data points as too similar and overweights nearby points while underweighting outlying data. Overweighting some points of a cluster requires that some other members receive negative weights to balance out the distribution of weights. This is sometimes called the *screen effect*, as negative weights are compensating for overweighting of nearby data that screen out distant

data. Variogram functions less prone to negative weights include the exponential and spherical functions, which both increase quickly at the origin. Gaussian curves are optimal for truly normally distributed data—e.g. distribution of gas concentrations in a body of air—but in general spatial data is not normally distributed.

1.2. *Change of support problem*

The COSP refers to inference of values at points or areas that are different from those where values have been observed (Gelfand et al., 2001). In cases where observations are available only as an aggregation to an area, and where the goal is to estimate new values over unobserved areas, the problem is more generally referred to as the MAUP (Cressie, 1996).

Dasymetric mapping was developed, essentially, to solve the COSP for population data. Generically, dasymetric maps downscale spatial data from a larger source zone to intersecting target zones using ancillary attributes of the target zones to determine population distribution. Regardless of its simplicity, the method consistently performs well at this task (Amos, McDonald, & Watkins, 2017; Barrozo, Pérez-Machado, Small, & Cabral-Miranda, 2016; Eicher & Brewer, 2001; Holt & Lu, 2011). Dasymetric mapping does not directly model spatial autocorrelation, nor can it spatially interpolate data; it depends entirely upon relationships in *feature space*—the ancillary data—rather than in geographical space.

Kriging was developed to model and estimate spatial distribution of natural resources. The notable feature of kriging is that it models spatial autocorrelation of the data, and not the actual data values. In contradistinction to dasymetric mapping, kriging was originally designed to aggregate data—so called block-kriging (Cressie, 1990; Matheron, 1963). Conventionally, kriging analyzes samples of point data in order to estimate a complete spatial distribution for the purpose of summarizing the resources available over an area of space. With its original design scope, kriging required modifications to estimate a spatial distribution from data aggregated in polygons. The solution is A2A/P kriging (Krivoruchko, Gribov, & Krause, 2011; Kyriakidis, 2004).

The basic technique of A2A/P kriging is to discretize the polygons into a series of regular points, and estimate the spatial autocorrelation between any two polygons as the average correlation of their discretizing points (Kyriakidis, 2004); similarly correlation between points and polygons is the average of correlation between the point and discretizing points of the polygon. A2A/P kriging is a type of binomial cokriging, mentioned in the previous section, where the discretized polygon data is cokriged with target discretization points from other polygons or a regular grid where one wishes to interpolate data, i.e. the target support. This whole procedure works because, as the next section will discuss, kriging models only variance as a function of distance between points, and does not need access to underlying data. The kriging model is simply an autocorrelation model that describes how *best*—as in best linear unbiased estimate (BLUE)—to interpolate observed data to unobserved supports.

In regards to downscaling of demographic data, the issue of binomial data and COSP often are the same. Even where raw point data is available, demographic data is often aggregated to a boundary where it can be converted to a proportion. Methods like binomial cokriging reflect this link between the COSP and binomial data.

1.3. Regression kriging and regression weighted dasymetric mapping

RK and RWDM may be compared roughly in terms of the assumptions each makes about space. Kriging assumes that the data varies continuously over space and is spatially autocorrelated, such that a set of closely clustered locations are more similar to each other than a set of widely scattered locations. Dasymetric mapping ignores spatial factors like shape, size, and distance, and instead assumes that spatial relationships can be described entirely by the values of ancillary data sources. The accuracy of either method depends, essentially, upon an ability to identify spatially homogeneous areas (Maantay, Maroko, & Herrmann, 2007; Maantay, Maroko, & Porter-Morgan, 2008). Kriging and RWDM exemplify this difference, in that the former models only spatial autocorrelation, and the latter models only correlations with the ancillary data. RK combines both methods; a topic this section will presently examine.

The kriging interpretation of spatial data is somewhat fraught for demographics, where all data is attached to discrete objects—people—and cannot usefully be viewed as realizations of a smooth and continuous random field (see (Cova, Church, & Goodchild, 2002; Kjenstad, 2006) for discussion of the distinction between spatial fields and objects.) Nonetheless, choropleth maps represent a set of expectations for a variable, each bounded and constant over an area of space, and it's quite natural to imagine how this could be converted to a smoothly varying function over space. Schmid and Maccannell Schmid and Maccannell (1955) and later Tobler Tobler (1979) analyzed the production of contour maps from choropleth maps and control points of population density data. Kriging accomplishes the same task within a framework of analysis of spatial autocorrelation.

The data for both RK and RWDM is similar, consisting of a matrix of m predictors X and a vector of n observations \mathbf{y} thought to be dependent upon X , where typically $m \ll n$. Kriging models covariance between data points of \mathbf{y} , assuming these are realizations of random variables with a multivariate Gaussian distribution, and is generally equivalent to a Gaussian process (GP) regression model (Rasmussen & Williams, 2006). RWDM analyzes covariance between the columns of X and \mathbf{y} using OLS or GzLM. In addition to data values, kriging requires spatial coordinates of the data, usually written $\mathbf{y} = \{y(s_1), \dots, y(s_n)\} = \mathbf{y}(S)$ and $X = \{\mathbf{x}(s_1), \dots, \mathbf{x}(s_n)\} = X(S)$ to emphasize that the data is generated by a function with a spatial coordinate. The values of \mathbf{y} are measurements at the locations of S —e.g. a volume of mineral, inches of rainfall, or height of a plant. The matrix X is a data set of co-located measurements that represent independent variables. Both methods begin with an input of n cases $\{(y(s_1), \mathbf{x}(s_1)), \dots, (y(s_n), \mathbf{x}(s_n))\}$ and derive a vector of weights for estimating values at unobserved location s_0 .

Regression weights, or regression coefficients, weight the impact of variable $x_j(s_0)$ on the estimate of y_0 at location s_0 :

$$\hat{y}(s_0) = \sum_{j=0}^m \beta_j x_j(s_0); \quad x_0(s_0) \equiv 1$$

The constant term x_0 is set to 1. Kriging weights are similar, but rather than weight the effects of variables they instead weight the influence of nearby data points:

$$\hat{y}(s_0) = \sum_{i=1}^n \lambda_i y(s_i)$$

The mathematical notation emphasizes that values of \mathbf{y} and X are indexed to spatial locations by writing them as functions of location: $y(s_0)$ and $\mathbf{x}(s_0)$. We'll see that RWDM doesn't analyze location or spatial arrangement, and is based only on the values generated by those functions. In contradistinction most forms of kriging don't analyze ancillary data, and are based only on analysis of spatial arrangement. RK combines the two methods to analyze both ancillary data and spatial arrangement. We begin by describing the algebra and assumptions of regression that underlay RWDM; RK will use these same regression methods within the kriging framework, and by introducing the regression assumptions first we can see how these assumptions manifest in the kriging framework.

1.3.1. Regression weighted dasymetric mapping

OLS finds regression weights that minimize the sum of squared errors:

$$\sum_1^n (y_n - X\beta)^2 = \sum_1^n \epsilon_n^2$$

The first column of the matrix X are ones for the constant term, the other columns are the data. The weights β that minimize $\sum_1^n \epsilon_n^2$ are a function of the covariance of the data:

$$X^T X \beta = X^T \mathbf{y}$$

Where $X^T X$ and $X^T \mathbf{y}$ are the moment matrices corresponding to the second moments of X and $X^T \mathbf{y}$ with $\dim X^T X = m \times m$ and $\dim X^T \mathbf{y} = m \times 1$. The matrix $X^T X \beta$ is a system of equations, solving by inverting results in the familiar OLS equation:

$$\beta = (X^T X)^{-1} X^T \mathbf{y} \tag{1}$$

The best estimate for an unknown y_0 given input data \mathbf{x}_0 is

$$\hat{y}_0 = \mathbf{x}_0^T \beta$$

with variance equal to

$$\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}$$

$$\sigma^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{n - m}$$

OLS imposes a few important assumptions. The first is that $E[\epsilon|X] = 0$, a result of the fact that OLS minimizes the sum of ϵ^2 . The first assumption is the most important, it is what guarantees that the model is unbiased and free of confounding effects, and violating the first assumption will invalidate the model. The second assumption is that no variables are linearly dependent, and is required to solve the system of equations (i.e. invert) $X^T X$. A linearly dependent variable is also collinear, hence all but one of the collinear variables may be removed to obtain the assumption of linear dependence. The third is that ϵ^2 has the same conditional variance σ^2 for all observations—i.e. homoscedasticity—and that ϵ^2 is uncorrelated between observations, these two criteria together form the assumption of spherical errors. If the assumption of spherical errors does not obtain, the OLS estimates are still valid but the model is less efficient. This third assumption, in fact, is the essential difference between OLS and kriging methods: kriging relaxes the third assumption by modeling spatial autocorrelation between observations.

RWDM does not, in general, preserve the original volumes of the source zone data used to estimate values in the target zones. This property is sometimes referenced as coherence or exactitude, and it is an attractive property of kriging. Because RWDM lacks this property, users should usually rescale the estimates so that, when summing across target zones within a source zone, the sums equal the original source zone data. The equation to accomplish this is:

$$\hat{y}'_{i\cdot} = \hat{y}_{i\cdot} \frac{y_i}{\sum_{j=1}^{n_i} \hat{y}_{ij}}$$

where $\hat{y}'_{i\cdot}$ is a rescaled estimate within source zone i , $\hat{y}_{i\cdot}$ is the raw estimate, y_i is the observed value of source zone i , n_i is the number of target zones that intersect source zone i , and $\sum_{j=1}^{n_i} \hat{y}_{ij}$ sums over all j raw estimates in source zone i .

1.3.2. Regression kriging

RK analyzes data in two steps. The first step is essentially identical the RWDM described in the previous section. Rather than keep the fitted values, however, RK will map the residuals.

$$\hat{\epsilon} = \hat{\mathbf{y}} - \mathbf{y}$$

The second step of RK interpolates the residuals using kriging, with the ultimate effect of allowing the mean of the kriging system to vary based upon the RWDM model output. Simple (SK) and ordinary kriging (OK) are the two most basic forms of kriging, and both assume the data is generated by an intrinsically stationary process with a constant but known (SK) or unknown (OK) mean over a sample area. Residuals by definition have mean 0, hence RK uses SK to interpolate the residuals over an area. All forms of kriging assume that the observed data $\mathbf{y} = \{y(s_i)|i = 1, \dots, n\}$ is a single sample from a multivariate Gaussian process at locations $S = \{s_i|i = 1, \dots, n\}$.

SK is, unsurprisingly, simpler than OK. List 1.3.2 gives the data generating equations modeled by SK, OK, and RK—for comparison purposes. Although the only dif-

ference between SK and OK here is the mean parameter μ , this has important consequences for the objective functions that OK must minimize, and for the properties of the resulting model. OK assumes only intrinsic stationarity, and models a constant local mean μ_{x_0} . SK, on the other hand, assumes a known mean and thus a globally stationary constant μ that can be shifted to zero without loss of generality. Lacking a prior known mean, OK also has no way to compute a covariance function and instead uses a variogram to model covariance. Under most circumstances all major properties of covariance functions carry over to variograms (Gneiting, Sasvári, & Schlather, 2000; Wackernagel, 2003). A final important difference is that OK uses a Lagrangian multiplier to constrain the kriging weights to sum to one. The role of the Lagrangian in OK equations can be difficult to visualize geometrically. In this case, the Lagrangian can be understood as constraining the optimization of the OK equations to a point in the $\dim \boldsymbol{\lambda} = n$ dimensional space where the kriging weights sum to one. The kriging weights are similar to those for a weighted mean, and if they do not sum to one then some values will be over or underweighted yielding a result that is biased from the mean. See Wackernagel (2003) chapter 11 for a formal derivation of the OK system of equations using a Lagrangian multiplier.

- (1) simple: $y_i = e_i$
- (2) ordinary: $y_i = \mu + e_i$
- (3) regression: $y_i = \mathbf{x}_i^T \boldsymbol{\beta} + e_i$

The regression kriging equation in table 1.3.2 is oversimplified for comparison purposes; the complete RK equation is:

$$\hat{y}(s_i) = \mathbf{x}(s_i)^T \boldsymbol{\beta} + \boldsymbol{\lambda}_i (\mathbf{x}(s_i)^T \boldsymbol{\beta} - \mathbf{y}(s_i)) \quad (2)$$

The equations in table 1.3.2 omit the spatial indexes, and collapse the second term in the RK equation to e_i : the interpolated residual. RK is further subdivided based upon how the $\boldsymbol{\beta}$ parameters are obtained. If $\boldsymbol{\beta}$ is derived external to the RK model, and provided as input, this method is conventionally referred to as *universal kriging* (UK) (Hengl, Heuvelink, & Rossiter, 2007; Pebesma, 2006). If on the other hand the $\boldsymbol{\beta}$ parameters are fit along with the coordinates of the data locations, as part of the kriging equations, then this method is usually referred to as *kriging with external drift* (KED). The two methods yield identical results in most cases, however UK, because it accepts externally derived $\boldsymbol{\beta}$ parameters, can be used to extrapolate data to new locations. The ability to fit a RK model in one location and apply it in another raises important technical and ethical issues that will be revisited for a discussion within the concluding section. For derivations of SK and OK see (Wackernagel, 2003) or (Gotway & Stroup, 1997), and for a comparison of RK, UK, and KED see (Hengl et al., 2007) or (Wackernagel, 2003). SK is equivalent to GP regression, as described in (Rasmussen & Williams, 2006).

From the data SK builds a matrix, a row vector, and scalar:

$$V = \begin{bmatrix} c(s_1, s_1) & \cdots & c(s_1, s_n) \\ \vdots & \ddots & \vdots \\ c(s_n, s_1) & \cdots & c(s_n, s_n) \end{bmatrix}$$

$$\mathbf{v}_0 = [c(s_0, s_1), \dots, c(s_0, s_n)]$$

$$v(s_0) = c(s_0, s_0)$$

The function c is the covariance function—or kernel, in GP terminology. Kriging assumes both first and second order stationarity—so that all data points share V and $v(s_0)$. The matrix V is the $n \times n$ covariance matrix for all observations. The vector \mathbf{v}_0 is the $1 \times n$ row vector of variances between the coordinates s_0 and those in S , this vector is subscripted to emphasize that it is recomputed for each s_0 . The scalar $v(s_0)$ is the maximum variance σ_y of y_0 , and because we assume a stationary second moment it also is the maximum variance of all the measurements; therefore it is equal to the values on the diagonal of V . Although the covariances of V are functions only of the distance values in S , the kernel c has a number of free *hyperparameters* (e.g. σ_y) that must be estimated from \mathbf{y} . Indeed, the assumptions of constant mean and variance is equivalent to assuming that the data generation process is sufficiently homogeneous over the area of interest to allow for mean and variance to be estimated from the data; this is what allow the hyperparameters to be estimated from empirical data and build the matrix V . As mentioned previously, OK doesn't assume a known mean and models V using a (semi)variogram rather than a covariance; this distinction is academic for most, however, and most practitioners can treat these two things identically. It's critical to note that, other than hyperparameters, the data in \mathbf{y} isn't used to compute covariance matrix V and friends; these constructs represent only the spatial correlation in S .

The kriging weights are derived by solving the system of equations:

$$V\boldsymbol{\lambda}_0 = \mathbf{v}_0$$

$$\boldsymbol{\lambda}_0 = \mathbf{v}_0 V^{-1}$$

The probability of a particular estimate for y_0 is given by:

$$y_0|\mathbf{y} \sim N(\mathbf{v}_0 V^{-1}\mathbf{y}, v(s_0) - \mathbf{v}_0^T V^{-1}\mathbf{v}_0^T) \quad (3)$$

where the best estimate \hat{y}_0 is the mean parameter of the normal distribution N (the first parameter.) Notice that the kriging weights, $\boldsymbol{\lambda}_0 = \mathbf{v}_0 V^{-1}$, depend upon the location s_0 via \mathbf{v}_0 and weight the impact of \mathbf{y} upon the estimate $\hat{y}(s_0)$. The weights are subscripted to emphasize that they are computed for each s_0 . A consequence of the assumption of a constant mean over the sample area is that $E[\hat{y}_0 - y_0] = 0$, or that the estimate is unbiased. Notice also that, as the covariance is stationary (the second parameter of N), it too doesn't depend on the values of \mathbf{y} . The two most important aspects of these equations is 1) the kriging weights only depend on the locations in S and 2) they model covariance between these locations.

The kriging weights $\boldsymbol{\lambda}_0$ are similar in function to the regression weights $\boldsymbol{\beta}$: $\boldsymbol{\beta}$ weights the effect of the variables in X on y_0 , whereas $\boldsymbol{\lambda}_0$ quantify the influence of the ob-

served values in \mathbf{y} given their distance from the unobserved location s_0 . This difference implies another one of importance: the values of β are the same for any given $\mathbf{x}(s_0)$, whereas kriging weights depend upon the location s_0 ; kriging fits a regression model for each input point, and must develop the vector \mathbf{v}_0 for each point. One might say another difference is that OLS weights imply a varying mean, whereas kriging assumes a constant mean; however this comparison is ill-posed. Kriging allows for varying deviation from the mean, and this property is a better comparison with OLS' varying mean. Moreover, the constant mean of kriging is more akin to the intercept, or constant, term of an OLS model. A similarity between both kriging estimates and those provided by OLS are that all estimates are assumed to have the same variance, as dictated by kriging's assumption of second order stationarity, and OLS' homoscedasticity of variance. This assumption regarding variance for the system of kriging equations should not be confused, however, with the prediction variance of kriging, given by:

$$v(s_0) - \mathbf{v}_0^T V^{-1} \mathbf{v}_0$$

The same as the variance parameter of N in the prediction probability given by equation 3. Prediction variance of kriging expresses the uncertainty of an estimated value as a function of its distance from observed values. Homogeneity of variance, on the other hand, refers to the spherical variance of the covariance matrix V .

The matrix V and friends are also similar to the moment matrices of OLS. The comparison between GP and linear regression is usually drawn between GP and Bayesian linear regression (BLR), as BLR is identical to GP with a linear kernel. BLR also relates to OLS via the equation for the posterior distribution $\text{LIKELIHOOD} \times \text{PRIOR}$. OLS and BLR share identical values for maximum likelihood and as n , the number of cases, increases towards infinity the likelihood "washes out" the priors, and the estimates of OLS and BLR converge. What GP does differently is that, whereas BLR analyzes distributions over parameters, GP analyzes distributions over functions (Williams, 1997). The distinction is clear with a linear kernel $\sigma^2 \phi(X)^T \phi(X)$, where ϕ is a basis function (possibly identity.) GP with a linear kernel is BLR with priors over functions. The matrix form of the linear kernel is quite familiar from the OLS equation $\hat{\beta} = (X^T X)^{-1} X^T y$, from which we may derive the covariance of the weights β around their true values to be $E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] = \sigma^2 (X^T X)^{-1}$, which we can see is identical to the linear kernel (with identity basis). In essence, GP with a linear kernel would be similar to linear regression that computes moment matrices as $\dim X X^T = n \times n$ rather than $\dim X^T X = m \times m$. Kriging rarely uses a linear kernel, however, for which instead conventional kernels include exponential, spherical, Gaussian, and Matérn.

The above brief summary comparison of RK and RWDM is designed to illustrate one important difference. Kriging is a smoothing model that estimates values at unobserved locations based on distance from observed data and the autocorrelation structure among the observed data. RWDM is a linear model that estimates values at unobserved locations based on the covariance of the attributes at the unobserved location with the known data, given the covariance structure of the known data. This is an important distinction for policy and decision makers, where similarity between observed and unobserved locations may depend more upon similarity of their attributes rather than similarity of their spatial location (however, for applications where the estimated data does vary smoothly over the sample area, RK clearly outperforms RWDM (Bourennane, King, & Couturier, 2000)). RWDM, however, may be simpler to apply in practice since it is basically a general(ized) regression model, and doesn't

variable	unit	vulnerability concept
% female	person	gender
% children	person	age
% elderly	person	age
% adults with no secondary degree	person	education
% with no birth registration	person	social dependence
% 20-39 year females with no secondary degree	person	education&gender
% households with female head	household	gender&family
% households with disabled	household	special needs
% heads of household with no secondary degree	household	education
% single heads of household	household	family
average size of household	household	family
% households with no overseas worker	household	social dependence
% households with poor roofing	dwelling	residence
% households with poor walling	dwelling	residence
% households with no tenure	dwelling	social dependence
% houses over 30 years	dwelling	residence
% houses needing repair	dwelling	residence
% houses $< 10m^2$	dwelling	socioeconomic status

Table 1. Components of the Social Vulnerability Index

require the intermediate step of selecting and parameterizing a covariance function.

2. Mapping social vulnerability

Solutions to the COSP help researchers to improve visualization of their data and build models that combine data from multiple sources and spatial resolutions. This final section will demonstrate RK in an application to downscale a social vulnerability index (SVI) from administrative divisions to a regular grid. This section will review the SVI and its construction, and consider independent predictors for the SVI. In addition to the SVI factors, the discussion will also analyze how geographical arrangements can be converted into metrics that are predictors of social vulnerability.

SVI is a nebulous and context dependent concept, and important components of a SVI metric may vary depending upon purpose and location. Many different researchers and organizations have defined an SVI, and this section uses the SVI developed by Ignacio for predicting vulnerability to natural disasters in the Philippines (Ignacio, Cruz, Nardi, & Henry, 2015). Ignacio obtained census data from Philippines Statistics Authority and computed SVI from neighborhood, or barangay, level data. The following procedure uses data from IPUMS-International (Minnesota Population Center, 2018) that is georeferenced to municipalities. Table 1 lists the factors of Ignacio’s SVI, and what each represents. To produce SVI, these factors are combined using principal components analysis (PCA.) We retain only the first component to represent SVI, although some analyses may retain more. This SVI value is normally distributed with mean 0 and standard deviation 1, where positive values indicate greater than average vulnerability and opposite for negative values. The average depends upon the input data—an SVI built with Philippines data represents variance around average vulnerability for a Filipino family.

The unit of analysis of an SVI is the family, however it has limited validity applied to

variable	source	vulnerability relation
urban landcover	(MacDonald, Dettwiler and Associates Ltd., 2013)	accessibility
agricultural landcover	(MacDonald, Dettwiler and Associates Ltd., 2013)	socioeconomic status
nighttime lights	(Elvidge, Baugh, Zhizhin, Hsu, & Ghosh, 2017)	infrastructure
population density	(UT-Battelle, 2019)	socioeconomic status
access to cities	(Weiss et al., 2018)	accessibility

Table 2. Predictors of SVI for RK model

single families, and the values are more usefully aggregated to a geospatial area—so that SVI is a property of a community that encompasses the people, housing, and labour available to a spatial location. The Philippines SVI data is attached to administrative divisions, to which they can be aggregated and geospatially located.

Along with the SVI, we have a set of geospatial predictors at the target support and will use RK to rescale the SVI data from polygons to this support. Some of these data are observations of the physical environment, often collected and compiled for years or decades, and available preprocessed and in a format designed for analysis; these could include climatology, landcover, crop & livestock, human population density, and nighttime lights among others. There also exist complex data on the human-built environment, including buildings, roads, utilities, and infrastructure. These latter data facilitate computation of geodemographic metrics reflecting accessibility to social and economic resources. Pebesma (Pebesma, 2006) notes that such metrics are “cheap” (quotes in original) because this kind of data is generally freely available—e.g. OpenStreetMap (OpenStreetMap contributors, 2018). Cheap they may be, it requires lab time and labour to determine, justify, and implement these metrics. Parsimony demands the metrics be few and simple—which is tragically more difficult than computing masses of random spatial metrics. The SVI RK model uses only one cheap predictor, but it took years of research to devise and justify: Weiss’ accessibility to cities metric (Weiss et al., 2018), listed in table 2 along with the others used in the model. Most metrics used in the SVI RK model are not spatial computations or even zonal statistics, but merely observations at a point in space; this also seems to be the case for most spatial models.

The data in table 2 are the predictors to essentially a RWDM model of social vulnerability. In our case, we will use KED to fit the β values, but as discussed the results are very similar to UK. The SVI is a continuous linear value, so it could be fit with a simple OLS model. The following equation would specify the model.

$$y = \alpha + \beta_1 x_{pop} + \beta_2 x_{pop}^2 + \beta_3 x_{urban} + \beta_4 x_{ag} + \beta_5 x_{travel} + \beta_6 x_{lights}$$

The predictor variables are population x_{pop} , urban landcover x_{urban} , agricultural landcover x_{ag} , travel time to nearest city x_{travel} , and nighttime lights x_{lights} . The equation is quadratic in population density, to allow for the effect of population density upon SVI to vary with population density. The kriging step computes the residuals $\mathbf{x}(s_i)^T \boldsymbol{\beta} - y(s_i)$ of the regression model for a regular grid and fits the kriging model to the entire land mass of the Philippines, using the RK method described in section 1.3.2. The computational procedure for fitting the model allows for options to

control behavior of the fitting procedure, but which we’ve skipped here for brevity. In particular, users of kriging are advised to consider options that force the kriging system to compute covariances using only data within a maximum distance of a target point or only data up to a threshold number of data points. Both these options limit the size of V and \mathbf{v}_0 , which may be a scientifically sensible thing to do, and moreover can allow larger models to be fit on modest hardware.

Figure 1 display the results of downscaling social vulnerability. Associated with the model is interpolation uncertainty quantification, plotted around the National Capital Region, in figure 2b, this represents uncertainty due to the sizes of the input polygons; larger polygons and those further from other polygons have greater interpolation uncertainty—notice that islands and areas associated with Central Luzon—where administrative divisions encompass larger area—have higher uncertainty. The plot margins are the row and column means of the plotted values—these can help highlight isolated and outlying areas that viewers might miss, for example the spots of high vulnerability in Palawan and Muslim Mindanao.

Interpolation uncertainty might help with decision making, but ultimately model results are based upon assumptions not testable within the framework of the model. If we had the information necessary to assess the accuracy, we wouldn’t need to solve the COSP in the first place. The map in figure 1 is a *best linear unbiased prediction* (BLUP) based upon our knowledge of how spatial features may correlate or even cause social vulnerability; by *best* we mean no linear function exists able to achieve better prediction error. We can improve upon it with focused research that tests and establishes relationships between social vulnerability and geographic space, and using data sets with wide geographic coverage to expand the applicability of resulting models to broader, hopefully global, ranges.

If one is comfortable with the assumptions of RK, it can be extrapolated via UK to locations beyond the area where the model was fit. First use KED to fit a RK model and retain the β estimates. Next supply these estimates to a UK model targeting the new area, which will use them to predict SVI before calculating residuals and kriging these values via SK with available georeferenced data. This technique raises ethical and technical concerns that will require more room for discussion than is available now, however the final paragraphs of the section will touch upon these interesting aspects for future research.

The technical concern is that of extrapolation. UK begins with an RWDM or KED model fit to data from a source geographic location, ideally where the data is less aggregated than at a target location where the model will be applied. Looking again at equation 2, we see that the calculated residuals will be the observed \mathbf{y} values, from whatever level of aggregation, subtracted from the predicted value at the target resolution: $\mathbf{x}(s_i)^T \beta - y(s_i)$. Because these residuals are calculated from a model with greater geographic detail, they will share the same levels of geographic variability as the source model, and furthermore will assume the same relationship between the features of the source regression model and the dependent variable. Thus, if some features have a different relationship in the source location, then the predictions of the model may be systematically biased. Despite this additional level of detail in the residuals the kriging weights remain based upon the geographical aggregation of the dependent variable at the target location. This means the details will be smoothed out by the size of the area to which the dependent variable is aggregated, and the uncertainty of the interpolated values will be greater. In the event a sensible covariance function cannot be devised or fit to the data in the target location—e.g. if the data is georeferenced at too great a level of aggregation then the covariance function while have a large

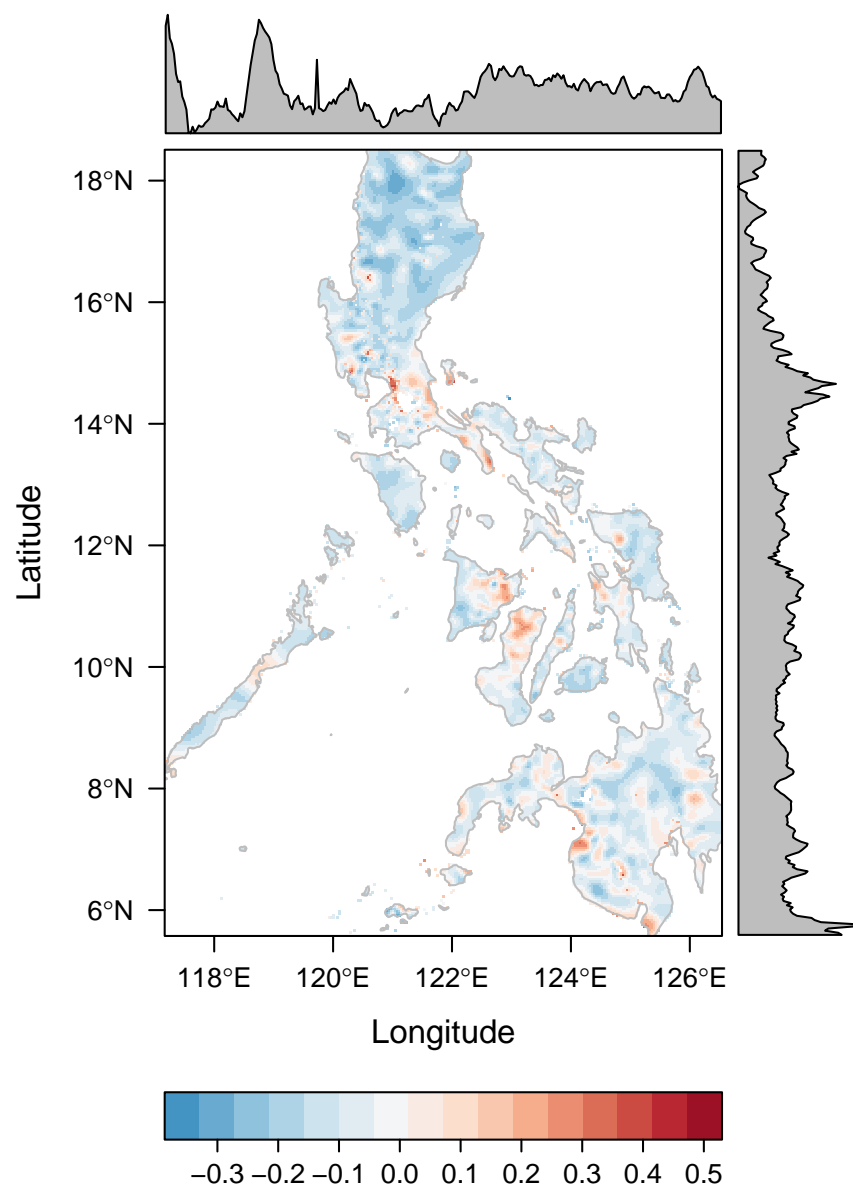
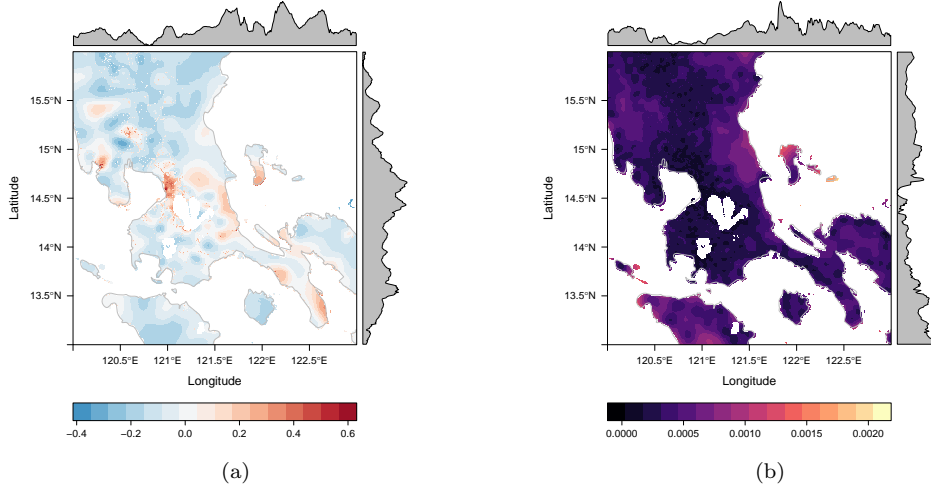


Figure 1. Social vulnerability, Philippines, 1km scale



“nugget”, or a variance in the measured value greater than zero as distance to the measurement approaches zero—then the SK step of UK will still fail or be unreliable.

Technical concerns of extrapolation are familiar, however these models are based on large, often government owned, data sets and may be burdened with subtle ethical if not legal questions. Census data sets like those used to calculate and initially geolocate SVI are provided by foreign governments to non-nationals in goodwill for the purposes of research and education. A model fit to that data, however, no longer needs the original data. It can be applied with known error in the original location, or applied to any location with necessary input data, as long as one is willing to stomach the assumptions. The model is based upon data provided for research, but the model could now be used sans the original data. In this case, the concern is not one of privacy, confidentiality, or anonymity of persons in the data, but instead one of intent for how their data should be used. Census data, like health data, represent personal information. Privacy of citizens is already recognized as a concern for nations releasing their data through IPUMS-International (McCaa, Ruggles, Davern, Swenson, & Palipudi, 2006) and likewise researchers need also to consider issues that could arise when using this personal information to do things that go beyond the communities where it was collected.

3. Conclusion

This article has discussed the COSP as it concerns social sciences. Solutions to the COSP allow social scientist to improve visualization of mapped data, combine data from unequal supports, or make estimates at supports other than where the data was collected. The general solution to COSP is kriging: a flexible and robust spatial interpolation model that can be used for downscaling, upscaling, or sidescaling. The article presented RWDM and RK for models that are suited to the needs of social science. Both models allow ancillary georeferenced data to be used as predictors. The key feature that distinguishes kriging models for population data from other kriging methods is the A2A/P covariance function that calculates covariance between two polygons or a polygon and point as the average covariance of points that discretize

the polygons.

The article presented a spatial model of social vulnerability in the Philippines, and advised of potential technical and ethical concerns. The model, fit in the Philippines, can be applied beyond that country's borders. The Filipino census uses a geographically detailed sampling scheme that provides census data with high geographical resolution. The high resolution provides a lot of information for the model fitting function to find associations between ancillary data and the variables under study; other areas of the world often lack this geographically detailed demographic information, especially in the areas of the world where we often would like to better understand social vulnerability.

The goal of this article was to present results of research into methods for downscaling population data to smaller supports, and to do so in a way that we hope encourage other social scientists to make broader use of these methods. COSP solutions are varied: geographers typically wish to summarize data over larger levels of aggregation—the opposite goal of most social scientists. Nonetheless, many of the solutions are useful for social science needs; the article discussed RWD and RK with A2A/P covariance functions as particularly relevant for social scientists. The methods can shed new light on spatial relationships and assist with spatial reasoning. As social scientists apply these methods, however, they must remain aware that these methods can complicate existing concerns of privacy and usage of personal data collected in census and surveys.

References

- Abowd, J. M. (2018). The u.s. census bureau adopts differential privacy. In *Kdd 2018 proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (p. 2867-2867).
- Abowd, J. M., & Schmutte, I. M. (2019). An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review*, 109(1), 171-202.
- Amos, B., McDonald, M. P., & Watkins, R. (2017). When boundaries collide: Constructing a national voting database of demographic and voting statistics. *Public Opinion Quarterly*, 81, 385-400.
- Bailey, A. (2005). *Making population geography*. Hodder Arnold.
- Barrozo, L. V., Pérez-Machado, R. P., Small, C., & Cabral-Miranda, W. (2016). Changing spatial perception: dasymetric mapping to improve analysis of health outcomes in a megacity. *Journal of Maps*, 12(5), 1242-1247.
- Bourennane, H., King, D., & Couturier, A. (2000). Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. *Geoderma*, 97(3-4), 255-271.
- Brown, T., McLafferty, S., & Moon, G. (Eds.). (2009). *A companion to health and medical geography*. Blackwell Publishing Ltd.
- Consolidated Analysis Center Incorporated, L. (2013-2018). *Acorn*. Retrieved 2018-09-26, from acorn.caci.co.uk
- Cova, T. J., Church, R. L., & Goodchild, M. F. (2002). Extending geographical representation to include fields of spatial objects. *International Journal of Geographical Information Science*, 16(6), 509-532.
- Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, 22(3), 239-252.
- Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, 3, 159-180.
- Duetsch, C. V. (1996). Correcting for negative weights in ordinary kriging. *Computers & Geosciences*, 22(7), 765-773.

- Eicher, C. L., & Brewer, C. A. (2001). Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science*, 28(2), 125-138.
- Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., & Ghosh, T. (2017). VIIRS night-time lights. *International Journal of Remote Sensing*, 38(21), 526-535.
- Gelfand, A. E., Zhu, L., & Carlin, B. P. (2001). On the change of support problem for spatio-temporal data. *Biostatistics*, 2(1), 31-45.
- Gneiting, T., Sasvári, Z., & Schlather, M. (2000). *Analogies and correspondences between variograms and covariance functions* (technical report meeting No. NRCSE-TRS No. 56). National Research Center for Statistics and the Environment.
- Gooaerts, P. (2010). Combining areal and point data in geostatistical interpolation: applications to soil science and medical geography. *Mathematical Geosciences*, 42, 535-554.
- Goovaerts, P. (2004). Simulation-based assessment of a geostatistical approach for estimation and mapping of the risk of cancer. In *Geostatistics banff*.
- 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 kriging. *International Journal of Health Geographics*, 5(52).
- Goovaerts, P. (2009). Combining area-based and individual-level data in the geostatistical mapping of late-stage cancer incidence. *Spatial and Spatio-temporal Epidemiology*, 1(1), 61-71.
- Gotway, C. A., & Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(2), 157-178.
- Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: from equations to case studies. *Computers & Geosciences*, 33(10), 1301-1315.
- Holt, J. B., & Lu, H. (2011). Dasymetric mapping for population and sociodemographic data redistribution. In X. Yang (Ed.), *Urban remote sensing: Monitoring, synthesis and modeling in the urban environment*. John Wiley & Sons.
- Ignacio, J. A. F., Cruz, G. T., Nardi, F., & Henry, S. (2015). Assessing the effectiveness of a social vulnerability index in predicting heterogeneity in the impacts of natural hazards: case study of the tropical storm wash flood in the philippines. *Vienna Yearbook of Population Research*, 13, 91-129.
- Jefferson Institute. (2017). *Patchwork nation*. Retrieved 2018-09-26, from www.patchworknation.org
- Journel, A. G. (1983). Nonparametric estimation of spatial distributions. *Mathematical Geology*, 15(3).
- Kjenstad, K. (2006). On the integration of object-based models and field-based models in GIS. *International Journal of Geographic Information Science*, 20(5), 491-509.
- Krivoruchko, K., Gribov, A., & Krause, E. (2011). Multivariate areal interpolation for continuous and count data. *Spatial Statistics*, 3, 14-19.
- Kyriakidis, P. C. (2004). A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36(3), 259-289.
- Lajaunie, C. (1991). *Local risk estimation for a rare noncontagious disease based on observed frequencies* (Note No. N-36/91/G). Centre de Geostatistique, Fontainebleau, Ecole des Mines de Paris.
- Maantay, J. A., Maroko, A. R., & Herrmann, C. (2007). Mapping population distribution in the urban environment: the cadastral-based expert dasymetric system (ceds). *Cartography and Geographic Information Science*, 34(2), 77-102.
- Maantay, J. A., Maroko, A. R., & Porter-Morgan, H. (2008). Research note—a new method for mapping population and understanding the spatial dynamics of disease in urban areas: asthma in the Bronx, New York. *Urban Geography*, 29(7), 724-738.
- MacDonald, Dettwiler and Associates Ltd. (2013). *Baseview 2013, 30 meter*. Retrieved from <http://www.mdaus.com/products/land-cover-products>
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58, 1246-1266.
- McCaa, R., Ruggles, S., Davern, M., Swenson, T., & Palipudi, K. M. (2006). IPUMS-

- international high precision population census microdata samples: balancing the privacy-quality tradeoff by means of restricted access extracts. In *Proceedings of the 2006 cenex-sdc project international conference on privacy in statistical databases* (p. 375-382). Springer-Verlag.
- Minnesota Population Center. (2018). *Integrated public use microdata series, international: Version 7.1 [dataset]*.
- Nelder, J., & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, 135(3).
- Oliver, M. A. (1998). Binomial cokriging for estimating and mapping the risk of childhood cancer. *IMA Journal of Mathematics Applied in Medicine and Biology*, 15, 279-297.
- Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences* (p. 127-144). London: Pion.
- OpenStreetMap contributors. (2018). *Planet dump retrieved from <https://planet.osm.org>. "https://www.openstreetmap.org"*.
- Output Area Classification User Group. (2016). *Open geodemographics*. Retrieved 2018-09-26, from www.opengeodemographics.com
- Pebesma, E. J. (2006). The role of external variables and gis databases in geostatistical analysis. *Transactions in GIS*, 10(4), 615-632.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press.
- Schmid, C. F., & Maccannell, E. H. (1955). Basic problems, techniques, and theory of isopleth mapping. *Journal of the American Statistical Association*.
- Singleton, A. D., & Spielman, S. E. (2014). The past, present and future of geodemographic research in the united states and united kingdom. *The Professional Geographer*, 66(4), 558-567.
- Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74(367), 519-530.
- UT-Battelle, L. (2019). *Landscan high resolution global population data set*.
- Wackernagel, H. (2003). *Multivariate geostatistics* (3rd ed.). Springer.
- Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., ... Gething, P. W. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553.
- Williams, C. K. I. (1997). *Prediction with gaussian processes: from linear regression to linear prediction and beyond* (Tech. Rep. No. NCRG/97/012). Neural Computing Research Group.