Predicting Finite Populations from Spatially Correlated Data

Jay M. Ver Hoef, ADF&G, 1300 College Road, Fairbanks, AK 99701

KEY WORDS spatial statistics, geostatistics, block kriging, covariance, small area estimation, stratified random sampling, variogram

Abstract

Classical sampling methods can be used to estimate the mean of a finite or infinite population. Block kriging also estimates the mean, but of an infinite population in a continuous spatial domain. In this paper, I consider a finite population version of block kriging. The data are assumed to come from a spatial stochastic process. Minimizing mean-squared-prediction errors yields best linear unbiased predictions that are a finite population version of block kriging. Block kriging of finite populations has versions comparable tosimple random sampling and stratified sampling, and includes the general linear model. This method has been tested for several years for moose surveys in Alaska, and an example is given where results are compared to stratified random sampling.

1. INTRODUCTION

Monitoring ecological populations is an important goal for both academic research and management of natural resources. Successful management of moose populations in Alaska depends on obtaining estimates of moose abundance at regular intervals throughout the state. The Alaska Department of Fish and Game developed aerial survey methods to estimate and monitor moose populations (Gasaway et al. 1986). The methods of Gasaway et al. (1986) use stratified random sampling and are based on classical sampling principles that rely on design-based inference, which are very robust. Very few assumptions are required because the distribution for inference comes from the sample design, which is known and under our control. In this paper we will be interested in estimating (predicting) the mean or total number of moose from a fixed geographic area. For design-based methods, sample plots are chosen at random, moose are counted in these plots, and inference is derived from the inclusion probability for sample units (i.e., Horwitz-Thompson estimation). For moose surveys there are a finite number of sample units and so finite population methods are used.

There are some problems with design-based methods. Because few assumptions are required, they may lack power in cases where further assumptions are justified. This appears to be especially true in the case of "small area" estimation, which refers to making an estimate on a smaller geographic area within the overall study area. There may be few or no samples within that small area, so that design-based estimation may not be possible or variances become exceedingly large. An alternative is to assume that the data were generated by a stochastic process and use model-based approaches (see, e.g., Fay and Harriot 1979, Ghosh and Meeden 1986, and Prasad and Rao 1990).

The basic problem considered in this paper is the estimation of some function of the sample units, call it $\tau(\mathbf{z})$, where $\mathbf{z}$ is a vector of the realized values of a spatial stochastic process for all the sample units of a finite population. The function $\tau(\mathbf{z})$ could be the population mean, population total, or the mean or total of a subset of sample units that have few or no observed samples. The goal is to use a predictor based on the set of observed samples $\hat{\tau}(\mathbf{z}_s)$, where $\mathbf{z}_s$ is a vector of observed values for sampled units (see e.g., Bolfarine and Zacks 1992, pg. 6). Geostatistical models and methods are used (for a review, see Cressie, 1993). Geostatistics has been developed for point samples. Because points are infinitesimally small, an infinite population is assumed. The average value over some area can be predicted using methods such as block kriging, which uses aggregation. Thus it appears that this is closely related to small area estimation, but where samples come from point locations rather than a finite set of sample units. In this paper, I consider the case where we have a finite collection of plots and we assume that the data were produced by a spatial stochastic process. It appears this has not been considered in detail. I develop a finite population version of block kriging (FPBK) which has been successfully used for estimating and monitoring moose abundance in Alaska and the Yukon.

1.1 Quick Review of Universal Block Kriging

Kriging is a spatial prediction method that is formulated by minimizing the mean-squared-prediction errors (MSPE), also known as the prediction variance. This treatment follows Cressie (1993, pg. 151). Kriging can be formulated by using variograms or covariance. Here, we show the covariance results. Suppose the data follow some linear model,

$$\mathbf{z} = \mu + \delta, \tag{1}$$

where $\mu = \mathbf{X}\beta$ and $\mathbf{X}$ has dimensions $n \times p$. Assume that the spatial random variable $Z(\mathbf{s})$ is defined at each location $\mathbf{s}$ in some region $\mathcal{D} \subset \mathcal{R}^d$. Define second-order stationarity for the random errors δ as follows: $E[\delta(\mathbf{s})] = 0$ so that $E[Z(\mathbf{s})] = \mu(\mathbf{s}) = \mathbf{x}'(\mathbf{s})\beta$ for all $\mathbf{s} \in \mathcal{D}$, and that the covariance,

$$C(\mathbf{h}) \equiv cov[\delta(\mathbf{s}), \delta(\mathbf{s}+\mathbf{h})], \tag{2}$$

exists and depends only on $\mathbf{h}$. For universal block kriging, define

$$Z(B) \equiv \int_B Z(\mathbf{s})d\mathbf{s}/|B|, \tag{3}$$

and

$$\mu(B) \equiv \int_B \mu(\mathbf{s})d\mathbf{s}/|B|,$$

for some area $B \subset \mathcal{D}$ where $|\mathbf{B}|$ is the area (volume) of B, assuming that the integrals exist for the process $\{Z(\mathbf{s})\}$ (see Cressie, 1993, pg. 106). $Z(B)$ is a random variable for the average value within the block B, and $\mu(B)$ is the expected value within the block. Data are collected at n locations, and assume the data are a realization of the random vector $\mathbf{z} \equiv [Z(\mathbf{s}_1), Z(\mathbf{s}_2), ..., Z(\mathbf{s}_n)]$. Let $\mathbf{a}'\mathbf{z}$ be a linear predictor for the random variable $Z(B)$, subject to the unbiasedness constraint $E(\mathbf{a}'\mathbf{z}) = E[Z(B)]$. Then universal block kriging uses (2) to minimize the MSPE; that is, find a λ such that

$$E[\mathbf{a}'\mathbf{z} - Z(B)]^2 - E[\lambda'\mathbf{z} - Z(B)]^2 \geq 0 \tag{4}$$

for all $\mathbf{a}$ such that $\mathbf{a}'\mathbf{z}_k$ is unbiased. Minimizing $E[\lambda'\mathbf{z} - Z(B)]^2$ in (4) in terms of covariances yields the set of equations,

$$\begin{pmatrix} \Sigma & \mathbf{X} \\ \mathbf{X}' & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \mathbf{m} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_B \\ \mathbf{x}_B \end{pmatrix}, \tag{5}$$

where $\mathbf{c}_B = [c_1(B), c_2(B), ..., c_n(B)]'$ with $c_i(B) \equiv \int_B C(\mathbf{s} - \mathbf{s}_i)d\mathbf{s}/|B|$ for $i = 1, 2, ..., n$, and $\mathbf{x}_B = [x_1(B), x_2(B), ..., x_p(B)]'$ with $x_j(B) \equiv \int_B x_j(\mathbf{s})d\mathbf{s}/|B|$ for $j = 1, 2, ..., p$. The solution of (5) for λ and $\mathbf{m}$ yields the block BLUP $\widehat{Z}(B) = \lambda'\mathbf{z}$, which can be written as

$$\widehat{Z}(B) = \mathbf{c}_B'\Sigma^{-1}(\mathbf{z} - \widehat{\mu}) + \widehat{\mu}_B, \tag{6}$$

where $\widehat{\mu} \equiv \mathbf{X}\widehat{\beta}_{GLS}$ and $\widehat{\mu}_B \equiv \mathbf{x}_B'\widehat{\beta}_{GLS}$ with $\widehat{\beta}_{GLS} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{z}$. The block kriging variance is given by

$$\begin{aligned} E[\lambda'\mathbf{z} - Z(B)]^2 = \sigma^2_{B,B} - \mathbf{c}_B'\Sigma^{-1}\mathbf{c}_B \\ + \mathbf{d}_B'(\mathbf{X}'\Sigma^{-1}\mathbf{X})\mathbf{d}_B, \end{aligned} \tag{7}$$

where $\sigma^2_{B,B}$ is $\int_B \int_B C(\mathbf{s} - \mathbf{u})d\mathbf{s}d\mathbf{u}/|B|^2$ and $\mathbf{d}_B = (\mathbf{x}_B - \mathbf{X}'\Sigma^{-1}\mathbf{c}_B)$.

2. BLOCK PREDICTION FOR FINITE POPULATIONS

One objective of finite population sampling is to estimate the average or total of the values that are actually realized, rather than the mean of some superpopulation from with the data were drawn. The equivalent objective is prediction, not estimation, for spatial processes in model-based approaches. That is, the goal is to *predict* a function of the actual values that occurred, not *estimate* unobservable parameters of a model (see Cressie, 1993, pgs. 13-16 for more details). To formulate this more clearly, suppose that $\mathbf{z}$ is a vector of random variables on a finite spatial lattice. The spatial lattice D is a set that can be indexed, each location denoted by i, $i = 1, 2, ..., N$. The random variable Z_i is located at the ith site in the lattice. Let $\tau(\mathbf{z}) = \mathbf{B}'\mathbf{z}$ be a vector of random variables to be predicted, where $\mathbf{B}$ is the $N \times k$ matrix $[\mathbf{b}_1 | \mathbf{b}_2 | ... | \mathbf{b}_k]$. For example, $\mathbf{b}_j = (1/N)(1, 1, ..., 1)'$ would be the average of the realized values of $\mathbf{z}$ in D. Notice that here $\mathbf{b}'\mathbf{z}$ acts as the finite version of $Z(B)$ in (3). Other possibilities are $\mathbf{b}_j = (1, 1, ..., 1)'$, which is the total of the realized values of $\mathbf{z}$ in D, and $\mathbf{b}_j = (0, 0, ...0, 1, 1, ..., 1, 0, ..., 0, 0)'$ which is the total of the realized values of $\mathbf{z}$ for some subregion (small area) in D. Data are collected from a subset of D, call it the $n \times 1$ vector $\mathbf{z}_s$, and let the unsampled locations be denoted by the $(N - n) \times 1$ vector $\mathbf{z}_u$, and write $\mathbf{z} = (\mathbf{z}_s', \mathbf{z}_u')$. We want some linear combination of the data, call it $\widehat{\tau}(\mathbf{z}_s) = \mathbf{A}'\mathbf{z}_s$, in order to predict $\mathbf{B}'\mathbf{z}$.

Definition 1 *Mean-Squared Prediction Error (MSPE) Matrix*

Let the MSPE matrix for any particular A be,

$$\mathbf{M}_A = E(\mathbf{A}'\mathbf{z}_s - \mathbf{B}'\mathbf{z})(\mathbf{A}'\mathbf{z}_s - \mathbf{B}'\mathbf{z})'. \quad (8)$$

Definition 2 *Best Linear Unbiased Predictor (BLUP)*

The matrix Λ is BLUP if,
1) $E(\Lambda'\mathbf{z}_s) = E(\mathbf{B}'\mathbf{z})$, and
2) $\mathbf{M}_A - \mathbf{M}_\Lambda$ is non-negative definite for every $\mathbf{A} \neq \Lambda$.

For the rest of this paper, assume that $\mathbf{z}$ follows the linear model, $\mathbf{z} = \mathbf{X}\beta + \delta$, or

$$\begin{pmatrix} \mathbf{z}_s \\ \mathbf{z}_u \end{pmatrix} = \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_u \end{pmatrix} \beta + \begin{pmatrix} \delta_s \\ \delta_u \end{pmatrix}, \quad (9)$$

where $\mathbf{X}$ is a matrix of fixed effects, β is a parameter vector, $E(\delta) = \mathbf{0}$, and

$$var(\delta) = \begin{pmatrix} \Sigma_{ss} & \Sigma_{su} \\ \Sigma_{us} & \Sigma_{uu} \end{pmatrix}.$$

To find the BLUP, we need to establish the uniform unbiased conditions for the predictor and then find the Λ that "minimizes" the MSPE matrix.

2.1 Uniform Unbiasedness Conditions

We need to consider all $\mathbf{A}$ such that $E(\mathbf{A}'\mathbf{z}) = E(\mathbf{B}'\mathbf{z})$ for all β in the linear model (9). Taking expectations, we see that $\mathbf{A}'\mathbf{X}_s\beta = \mathbf{B}'\mathbf{X}\beta$ for every β, so that implies $\mathbf{A}'\mathbf{X}_s = \mathbf{B}'\mathbf{X}$, or

$$\mathbf{A}'\mathbf{X}_s = \mathbf{B}_s'\mathbf{X}_s + \mathbf{B}_u'\mathbf{X}_u, \quad (10)$$

where $\mathbf{B}' = [\mathbf{B}_s' | \mathbf{B}_u']$.

2.2 Prediction for Finite Populations

Similar to equations (4), the BLUP is found by finding Λ such that

$$E(\mathbf{A}'\mathbf{z}_s - \mathbf{B}'\mathbf{z})(\mathbf{A}'\mathbf{z}_s - \mathbf{B}'\mathbf{z}) - (\Lambda'\mathbf{z}_s - \mathbf{B}'\mathbf{z})(\Lambda'\mathbf{z}_s - \mathbf{B}'\mathbf{z}) \quad (11)$$

is non-negative definite for all $\mathbf{A}$ such that $\mathbf{A}'\mathbf{z}_s$ is unbiased. By minimizing the MSPE matrix, we obtain the prediction equations,

$$\begin{pmatrix} \Sigma_{ss} & \mathbf{X}_s \\ \mathbf{X}_s' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Lambda \\ \mathbf{M} \end{pmatrix} = \begin{pmatrix} \Sigma_{ss} & \Sigma_{su} \\ \mathbf{X}_s' & \mathbf{X}_u' \end{pmatrix} \begin{pmatrix} \mathbf{B}_s \\ \mathbf{B}_u \end{pmatrix}, \quad (12)$$

which can be compared to equations (5). When (12) are solved for Λ, the FPBK predictor is,

$$\hat{\tau}(\mathbf{B}'\mathbf{z}) = \Lambda'\mathbf{z}_s = \mathbf{B}_s'\mathbf{z}_s + \mathbf{B}_u'\hat{\mathbf{z}}_u, \quad (13)$$

where,

$$\hat{\mathbf{z}}_u = \Sigma_{us}\Sigma_{ss}^{-1}(\mathbf{z}_s - \hat{\mu}_s) + \hat{\mu}, \quad (14)$$

$\hat{\mu}_u = \mathbf{X}_u\hat{\beta}_{GLS}$ and $\hat{\mu}_s = \mathbf{X}_s\hat{\beta}_{GLS}$ with $\hat{\beta}_{GLS} = (\mathbf{X}_s'\Sigma_{ss}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\Sigma_{ss}^{-1}\mathbf{z}_s$. The predictor (14) can be compared to (6). The FPBK predictor (13) is now seen as multiplying the observed sample values times their corresponding coefficients from $\mathbf{B}_s$, and then using universal block kriging to predict all other unsampled units, and these predictions are multiplied by their corresponding coefficients in $\mathbf{B}_u$.

2.3 Prediction Variance (MSPE)

Substituting the solution for Λ for $\mathbf{A}$ in (8), we obtain the MSPE (also called prediction variance) of FPBK,

$$\mathbf{M}_\Lambda = \mathbf{B}'\Sigma\mathbf{B} - \mathbf{C}'\Sigma_{ss}^{-1}\mathbf{C} + \mathbf{D}'\mathbf{V}\mathbf{D}, \quad (15)$$

where

$$\mathbf{C} = \Sigma_{ss}\mathbf{B}_s + \Sigma_{su}\mathbf{B}_u,$$
$$\mathbf{D} = \mathbf{X}'\mathbf{B} - \mathbf{X}_s'\Sigma_{ss}^{-1}\mathbf{C}, \text{ and}$$
$$\mathbf{V} = var(\hat{\beta}_{GLS}) = (\mathbf{X}_s'\Sigma_{ss}^{-1}\mathbf{X}_s)^{-1}.$$

Equation (15) can be compared to (7). In (15), the quantity $\mathbf{B}'\Sigma\mathbf{B}$ is the variance of $\mathbf{B}'\mathbf{z}$, and assuming β is known, the prediction variance of $\mathbf{B}'\mathbf{z}$ is $\mathbf{B}'\Sigma\mathbf{B} - \mathbf{C}'\Sigma_{ss}^{-1}\mathbf{C}$. The additional term $\mathbf{D}'\mathbf{V}\mathbf{D}$ arises because we are estimating β, where $\mathbf{D}$ is, in some sense, the distance between predicted points in the design matrix $\mathbf{B}'\mathbf{X}$ and that of the observed design matrix $\mathbf{C}'\Sigma_{ss}^{-1}\mathbf{X}_s$. Equation (15) can be simplified for computing purposes,

$$\mathbf{M}_\Lambda = \mathbf{B}_u'(\Sigma_{uu} - \Sigma_{us}\Sigma_{ss}^{-1}\Sigma_{su} + \mathbf{W}'\mathbf{V}\mathbf{W})\mathbf{B}_u,$$

where $\mathbf{W} = \mathbf{X}_u' - \mathbf{X}_s'\Sigma_{ss}^{-1}\Sigma_{su}$. If $\mathbf{B}$ has more than one column, the prediction variances of each $\mathbf{b}_j$ are contained as diagonal elements of $\mathbf{M}_\Lambda$ and prediction covariances between $\mathbf{b}_j$ and $\mathbf{b}_{j'}$ are contained as the off-diagonal elements of $\mathbf{M}_\Lambda$.

2.4 Connections to Sampling Theory

Suppose we are interested in predicting the mean over a lattice of N sites. Then $\mathbf{B} = \mathbf{b} = (1/N)(1, 1, ..., 1)'$. Let $\Sigma_{ss} = \sigma^2\mathbf{I}_n$ where $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\Sigma_{uu} = \sigma^2\mathbf{I}_{N-n}$ where

$\mathbf{I}_{N-n}$ is the $(N-n)\times(N-n)$ identity matrix, $\mathbf{\Sigma}_{su}=\mathbf{\Sigma}'_{us}=\mathbf{0}$, and $\mathbf{X}=\mathbf{1}_N$, where $\mathbf{1}_N$ is a vector of N ones. Then from (13) $\mathbf{\Lambda}=\boldsymbol{\lambda}=(1/n)\mathbf{1}_n$ and the predictor is the sample mean

$$\boldsymbol{\lambda}'\mathbf{z}_s=\bar{z}. \tag{16}$$

Likewise, from (15) the

$$\text{MSPE}=(\sigma^2/n)(1-f), \tag{17}$$

where $f=(n/N)$ is the sampling fraction and $1-f$ is the finite population correction factor. Of course, equation (16) is the same estimator of the mean that is used in simple random sampling (e.g., Thompson, 1992, pg. 13) and equation (17) is the variance of the mean estimator used in simple random sampling (e.g., Thompson, pg. 15).

Next, consider stratified sampling. Allow each stratum to be a separate random process, independent from each other, each with its own mean and variance. These are model-based assumptions that are equivalent to stratified random sampling (SRS),

$$\mathbf{X}=\begin{pmatrix}\mathbf{1}_{n_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}_{n_2}\\ \mathbf{1}_{N_1-n_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}_{N_2-n_2}\end{pmatrix}, \tag{18}$$

and

$$\mathbf{\Sigma}_{ss}=\begin{pmatrix}\sigma_1^2\mathbf{I}_{n_1} & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\mathbf{I}_{n2}\end{pmatrix},$$

$$\mathbf{\Sigma}_{uu}=\begin{pmatrix}\sigma_1^2\mathbf{I}_{N_1-n_1} & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\mathbf{I}_{N_2-n2}\end{pmatrix},$$

where $\mathbf{z}=(\mathbf{z}_{s,1},\mathbf{z}_{s,2},\mathbf{z}_{u,1},\mathbf{z}_{u,2})'$. Suppose that now we want to predict the total, $\mathbf{B}=\mathbf{b}=\mathbf{1}_n$. Then, from (13) we obtain the predictor,

$$\boldsymbol{\lambda}'\mathbf{z}_s=N_1\bar{z}_{s,1}+N_2\bar{z}_{s,2}. \tag{19}$$

From (15),

$$\begin{aligned}\text{MSPE}=&(N_1^2\sigma_1^2/n_1)(1-f_1)\\ &+(N_2^2\sigma_2^2/n_2)(1-f_2),\end{aligned} \tag{20}$$

where $f_1=(n_1/N_1)$ and $f_2=(n_2/N_2)$. Of course, equation (19) is the same estimator of the mean that is used in SRS (e.g., Thompson, 1992, pg. 103) and equation (20) is the variance of the mean estimator used in SRS (e.g., Thompson, pg. 103).

Equations (17) and (20) demonstrate that equation (15) is a version of block kriging that provides a reduction in variance when sampling finite populations.

2.5 Modeling Autocorrelation

To make full use of model-based assumptions, we will need to estimate $\mathbf{\Sigma}$ by modeling the spatial autocorrelation in the data. We need to estimate each of the (i,j) entries in $\mathbf{\Sigma}$. One such model for spatial covariance (2) is the exponential model,

$$C(\mathbf{h}|\theta)=\begin{cases}\theta_1+\theta_2 & \mathbf{h}=0,\\ \theta_2\exp(-||\mathbf{h}||/\theta_3) & \mathbf{h}\neq 0,\end{cases} \tag{21}$$

where $\mathbf{h}=\mathbf{s}_j-\mathbf{s}_i$. There are many others (see Cressie, 1993, pg. 61). It is possible to estimate the parameters of $C(\mathbf{h}|\theta)$ using method moments for variograms or covariances (see Cressie, 1993, pg. 69) and then weighted least squares (see Cressie, 1993, pg. 99), or by using restricted maximum likelihood (REML, Patterson and Thompson, 1971, 1974); see Cressie (1993, pg. 92) for spatial REML. In the example below, I will use REML. As a visual diagnostic, I compute the empirical semivariogram,

$$\hat{\gamma}(\mathbf{h})=\frac{1}{2|N(\mathbf{h})|}\sum_{N(\mathbf{h})}[z(\mathbf{s}_i)-z(\mathbf{s}_j)]^2, \tag{22}$$

where $\mathbf{h}=\mathbf{s}_j-\mathbf{s}_i$, $N(\mathbf{h})=\{(\mathbf{s}_i,\mathbf{s}_j):\mathbf{s}_j-\mathbf{s}_i=\mathbf{h}\}$ and $|N(\mathbf{h})|$ is the number of distinct elements in $N(\mathbf{h})$. The fitted model covariance (21) is readily converted to a semivariogram using the relationship, $\gamma(\mathbf{h}|\theta)=C(\mathbf{0}|\theta)-C(\mathbf{h}|\theta)$.

3. EXAMPLES USING MOOSE SURVEY DATA

I give an example from a moose survey conducted in Alaska in the fall of 1999. The survey area was game management unit (GMU) 20A, shown as the darkened area within the state of Alaska in Figure 1.

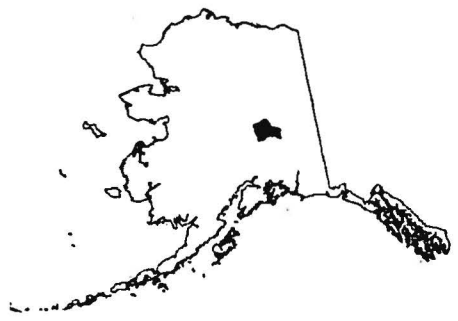

Fig. 1. Map of Alaska, GMU 20A

The survey was flown in November after there was sufficient snow cover to allow moose to be

readily observed. Moose surveys involve five basic elements that include, 1) defining the survey area, 2) stratifying the area, 3) selecting a sample, 4) surveying the sample of units within the area, and 5) analyzing the data. Within a survey area, sample units are laid out in a grid. The north-south boundaries are based on even increments of latitude (2 minutes, starting at 0) and the east-west boundaries are based on increments of longitude (5 minutes, starting at 0). At around 64 degrees latitude, sample unit size is approximately 15 square kilometers. The total area of the survey was 14878 square kilometers. Figure 2 shows an enlarged view of the 20A survey area.

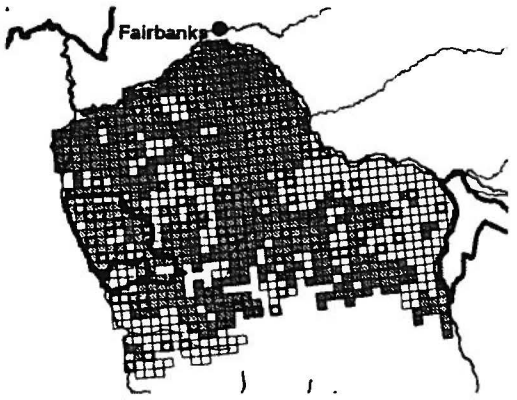

Fig. 2. GMU 20A stratification and samples

The area was stratified into two strata: relatively low moose density (shown as white in Fig. 2) and relatively high moose density (shown by grey in Fig 2), based on the biologist's knowledge of habitats, moose, and their distribution in previous years. The low stratum consisted of 338 sample units and the high stratum consisted of 649 sample units. After the area was stratified, a random sample of 86 samples was drawn with 52 from the high stratum and 34 from the low stratum. The sample units were flown and all moose were counted from the air within each sample. The sampled units are shown with a heavy border in Figure 2. All counts were first changed to density by dividing the counts by the area of each sample, which varied slightly due to the narrowing of longitude as one moves north. An average of 0.973 moose per square kilometer was counted in the high stratum, and an average of 0.398 moose per square kilometer was counted in the low stratum. The covariance between sample units was estimated using an exponential model (21) and REML (see Cressie, pg. 92). The distance between sample units was computed in kilometers from the center of one sample unit to the center of another. For the high stratum, the estimated parameters in (21) were $\theta_1 = 2.614$, $\theta_2 = 0.670$, and $\theta_3 = 23.26$. The empirical semivariogram (22) and the fitted model (21) for the high stratum are given in Fig. 3.

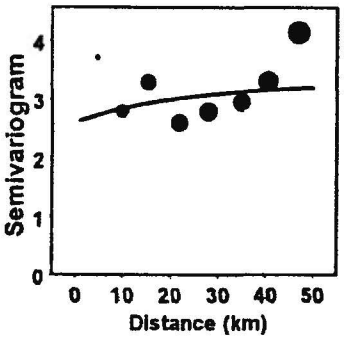

Fig. 3. Semivariograms for high stratum.

In Fig. 3, the size of the circle indicates the number of pairs of locations used for each distance class in the empirical semivariogram, and the line is the fitted model. For the low stratum, the estimated parameters in (21) were $\theta_1 = 0.000$, $\theta_2 = 2.102$, and $\theta_3 = 15.99$. The empirical semivariogram (22) and the fitted model (21) for the low stratum are given in Fig. 4.

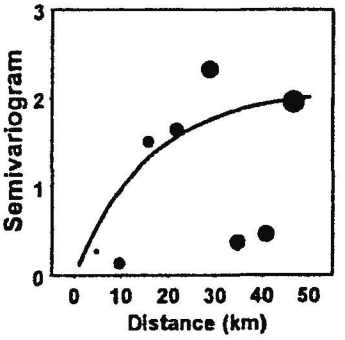

Fig. 4. Semivariograms for low stratum.

All covariances between the two strata are assumed to be 0. The fitted models, given above, are used to fill in the (i,j) entries in Σ. Using the estimated Σ and $\mathbf{B} = \mathbf{b} = (1/N)(1,1,...,1)'$ and $\mathbf{X}$ as in (18), the predictor FPBK (13) of the average moose density per sample was estimated to be 0.7613 moose

per square kilometer, or a total of 11327 moose in the whole study area. The estimated standard error of the average moose denisty using (15) was 0.0043, yielding an estimated standard error of 978 for the total number of moose. For comparision, using SRS (e.g., Thompson, 1992, pg. 103), the estimate of total moose abundance was 11535 with an estimated standard error of 985. It is generally true that if there is autocorrelation in the data, prediction will be more precise when it uses information about the autocorrelation; thus FPBK has slightly smaller variance than SRS. Using the same fitted variogram model, it is possible to make a prediction for any subset of samples; i.e., small area estimation. Fig. 2 also shows the Ferry Trail Managment Area (FTMA) outlined in bold on the left. FPBK yielded an estimate of 1437 moose in the FTMA with a standard error of 153. The SRS estimate using only the samples within the FTMA subset (13 highs and 4 lows) yielded an estimated 1535 moose with an estimated standard error of 227. For small area estimation, the estimated standard error of FPBK was significantly smaller than that of SRS.

4. DISCUSSION AND CONCLUSIONS

Based on twenty different moose surveys over 3 years, the application of geostatistical ideas to finite population sampling gives three main advantages over classical sampling: 1) FPBK is usually more precise than SRS, 2) FPBK allows small area estimation, and 3) FPBK allows nonrandom sampling designs, giving biologists greater flexibility. FPBK also has an advantage over block kriging because FPBK incorporates a finite population correction factor that reduces the prediction variance. In the example above, approximately 9% of the population was sampled, and it often gets as high as 30% for moose surveys in Alaska.

SRS allows a separate variance for each stratum. The analogy for a model-based approach is to have a separate spatial process for each stratum. Geostatistical methods assume a constant variance for a spatial process, and stratification helps meet that assumption. When considering multiple spatial processes induced by stratification, it is possible (even desirable) to model cross-correlation between strata processes. I have investigated this for moose surveys in Alaska and found little or no cross-correlation so that it did not affect the predictions or standard errors. An alternative to stratification is transforming the data to stabilize variance. This is useful when continuous covariates (e.g., elevation) are used in the design matrix of the linear model. These methods require an unbiased backtransformation (similar to transgaussian kriging, Cressie, 1993, pg. 137), and results will be presented later.

5. ACKNOWLEDGMENTS

Financial support for this work was provided by Federal Aid in Wildlife Restoration to the Alaska Department of Fish and Game.

REFERENCES

Bolfarine, H. and Zacks, S. 1992. Prediction Theory for Finite Populations. Springer-Verlag, New York. 207 p.

Cressie, N. 1993. *Statistics for Spatial Data, Revised Edition.* John Wiley and Sons, New York. 900 p.

Fay, R., and Harriot, R. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74: 269-277.

Gasaway, W.C., DuBois, S.D., Reed, D.J., and Harbo, S.J. Estimating moose population parameters from aerial surveys. Biological Papers of the University of Alaska, Number 22, 108 p.

Ghosh, M. and Meeden, G. 1986. Empirical Bayes estimation in finite population sampling. *Journal of the American Statistical Association* 81: 1058-1062.

Patterson, H.D. and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 58: 545-554.

Patterson, H.D. and R. Thompson. 1974. Maximum likelihood estimation of components of variance. Pages 197-207 in *Proceedings of the 8th International Biometric Conference*, Biometric Society, Washington, DC.

Prasad, N.G.N. and Rao, J.N.K. 1990. On the estimation of mean square error of small area predictors. *Journal of the American Statistical Association* 85: 163-171.

Thompson, S. 1992. *Sampling.* John Wiley and Sons, New York. 343 p.