ELSEVIER

Remote Sensing of Environment

Land Cover Estimation in Small Areas Using Ground Survey and Remote Sensing

L. Ambrosio Flores and L. Iglesias Martínez†*

For estimating crop acreage in "small areas" using ground survey and remote sensing, an Empirical Best Linear Unbiased Predictor Estimator is considered. It is a weighted mean of the Survey Regression and the Synthetic Regression estimators. The gain in precision due to the remotely sensed data is estimated for a case study. ©Elsevier Science Inc., 2000

INTRODUCTION

Detailed information about land cover and land use is necessary in order to implement environmentally sensitive policies and practices and to monitor and control such policies. Satellite imagery provides a complete spectral characterization of an area in digital form. This can be used to classify the area by crop types. However, the availability of such spectral data does not eliminate the need for ground data. Since it is difficult to differentiate between land uses (particularly between crops) with a very similar spectral signature, the estimates of land use acreage based only on satellite data are not accurate enough. The design-based Survey Regression estimator (Cochran, 1977) is a well-known method for estimating land use and land cover in large geographical areas (state or region) using remote sensing and ground data (Hanuschak et al., 1982; Allen, 1990; Ambrosio et al., 1993; Deppe, 1998).

However, there is a growing demand for reliable estimates over small areas (counties, irrigated areas). Due to the small sample size in small areas, the design-based Survey Regression estimator is not sufficiently precise for most uses. In this study we follow a model-based approach: we consider a statistical model to "borrow strength" from related small areas in order to obtain precise estimates for a given small area. Based on this model of the relationship between ground and satellite data, a Best Linear Unbiased Predictor (BLUP) estimator is defined, which makes optimal use of the available data, according to statistical criteria.

Since the BLUP estimator has optimal statistical properties, it would be preferred to any other linear estimator for a given sample size. However, it is necessary to verify the model assumptions since the statistical properties of the BLUP estimator are optimal only if the model assumptions are correct. In the specified model, the basic assumption is that the errors (the residuals resulting from the difference between the true scene and the inferred scene by the classification of the image data) are positively correlated within the small areas. This assumption derives from the fact, largely documented in the literature [for some references, see Labovitz and Masouka (1984)], that remotely sensed data are spatially correlated. This spatial correlation is positive and decreases when the distance between pixels increases so that the intrasmall areas correlation (average correlation between pairs of pixels from the same small area) decreases when the small area size increases.

In order to verify the model assumption, a statistic is introduced [Eq. (11)]. If the model assumption is not correct, that is, if the errors are not correlated inside small areas, then, as will be seen, the BLUP estimator turns into a Synthetic Regression estimator. In this case, for a given sample size, the last estimator mentioned is preferred to the BLUP estimator because it is as precise as the BLUP estimator and its calculation is easier than for the BLUP estimator. If the assumption is correct and the sample size in a given small area is high, then, as will be seen, the BLUP estimator turns into a Survey Regression estimator, which would be preferred to the BLUP estimator in this small area for the same two reasons regarding the Synthetic

* Departamento de Economía y Ciencias Sociales Agrarias, Unidad de Estadística, Universidad Politécnica de Madrid, Madrid, Spain

† Departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría—Expresión Gráfica, Universidad Politécnica de Madrid, Madrid, Spain

Address correspondence to L. Ambrosio Flores, Dpto. de Economía y Ciencias Sociales Agrarias, Unidad de Estadística, Escuela Tecnica Superior de Ingenieros Agrónomos, 28040 Madrid, Spain. E-mail: flores@eco.estia.upm.es

Received 25 June 1999; revised 23 February 2000.

REMOTE SENS. ENVIRON. 74:240–248 (2000)

655 Avenue of the Americas, New York, NY 10010

0034-4257/00/$–see front matter
PII S0034-4257(00)00114-0

Regression estimator: The calculations are easier than for the BLUP estimator and precision is similar. A case study is presented showing how to evaluate this assumption in practice and also comparing the relative efficiency of the proposed BLUP estimator with three other estimators (Survey Regression, Synthetic Regression, and Direct Expansion estimator). The gain in precision in the estimates attributable to remotely sensed data is also evaluated.

The data requirements are detailed in the next section. Basically only two kinds of data are necessary: A classified scene from the image data (using "training pixels" in order to identify the image signature that corresponds to each type of ground data) and ground data observed in a sample of "segments." A numerical example has been included as an appendix.

A program written for the IML procedure of the SAS statistical package can be obtained from the authors upon request.

GROUND AND SATELLITE DATA

It is assumed that the ith small area ($i=1, 2, \ldots, m$) is divided into N_i sampling units or "segments." Associated with the jth segment ($j=1, 2, \ldots, N_i$), there are two numbers (y_{ij},x_{ij}): y_{ij} is the true number of hectares (fixed, but unknown) of the land use in the segment and x_{ij} is the number of hectares of classified land use in the segment, observed by remote sensing. In order to estimate the mean per segment, $\overline{Y}_i=\frac{1}{N_i}\sum_{j=1}^{Ni} y_{ij}$, of the y-values in each one of the ith small areas ($i=1, 2, \ldots, m$), a simple random sample of n sampling units or segments is selected from among the $N=\sum_{i=1}^{m} N_i$ total segments. Since N_i is known, an estimate of the total $Y_i=\sum_{j=1}^{Ni} y_{ij}$ is the estimate of the mean multiplied by N_i and the standard error of the total estimator is N_i times the standard error of the mean estimator. The number n_i of sampling units in the ith small area is a random value ranging from 0 to n. For the selected sample both numbers $\{(y_{ij},x_{ij}); j=1,2, \ldots,n_i; i=1,2, \ldots,m\}$ can be observed. Since the satellite data are a complete classification of the landscape, it is possible to establish the x_{ij}-values for each of the N segments of the whole population: $\{x_{ij}; j=1,2, \ldots,N_i; i=1,2, \ldots,m\}$. However, as will be seen, only the totals, $X_i=\sum_{j=1}^{Ni} x_{ij}$ or means $\overline{X}_i=\frac{1}{N_i}\sum_{j=1}^{Ni} x_{ij}$ for $i=1,2, \ldots,m$, are required apart from the sample data $\{(y_{ij},x_{ij}); j=1,2, \ldots,n_i; i=1,2, \ldots,m\}$. Even if $n_i=0$ for a given small area, it will be possible to estimate $\overline{Y}_i$ from X_i.

BEST LINEAR UNBIASED PREDICTOR (BLUP) ESTIMATOR

In order to estimate $\overline{Y}_i$, a model-based estimator is considered (Royall, 1970; Royall and Herson, 1973):

$$\hat{\overline{Y}}_i=f_i\,\overline{y}_i+(1-f_i)\hat{\overline{Y}}_i^{*} \tag{1}$$

where $f_i=n_i/N_i$ is the sampling ratio, $\overline{y}_i$ is the sample average of y_{ij}, and $\hat{\overline{Y}}_i^{*}$ is a predictor of the mean hectares per segment in the remaining segments not included in the sample.

The Model

The predictor $\hat{\overline{Y}}_i^{*}$, in this case considered to be the Best Linear Unbiased Predictor (BLUP), is based on the Linear Mixed Model (Battese et al., 1988):

$$y_{ij}=\beta_1+\beta_2 x_{ij}+v_i+e_{ij} \tag{2}$$

where β_1 and β_2 are unknown parameters with fixed values (fixed-effects), v_i are independent random variables of mean zero and variance σ_v^2 (random effects), and the e_{ij} are independent random variables of mean zero and variance σ_e^2. v_i and e_{ij} are independent so that the variance of $u_{ij}=v_i+e_{ij}$ is $\sigma_u^2=\sigma_v^2+\sigma_e^2$ (σ_v^2 and σ_e^2 are called variance components).

Model [Eq. (2)] can be specified as a fixed-effect model with autocorrelated errors instead of a mixed model:

$$y_{ij}=\beta_1+\beta_2 x_{ij}+u_{ij} \tag{3}$$

where β_1 and β_2 are unknown parameters with fixed values (fixed-effects) and u_{ij} are random variables of mean zero and covariance structure:

$$Cov(u_{ij},u_{i'j'})=\begin{cases}\sigma_v^2+\sigma_e^2; & \forall i=i'; j=j'\\ \sigma_v^2; & \forall i=i'; j\neq j'\\ 0; & \forall i\neq i'\end{cases} \tag{4}$$

Substituting in Eq. (2) $u_{ij}=v_i+e_{ij}$, it can be seen that both models [Eqs. (2) and (3)–(4)] have the same fixed part and the same variance and covariance matrix, $\underline{V}$, of the random part: specific for the whole sample size $n=\sum_{i=1}^{m} n_i$; this matrix is $\underline{V}=\sigma_v^2\underline{J}+\sigma_e^2\underline{I}$, where $\underline{J}=diag\{\underline{J}_1,\underline{J}_2, \ldots,\underline{J}_i, \ldots,\underline{J}_m\}$ is a block diagonal matrix with $\underline{J}_i$ being a square matrix of the order n_i, with all the elements equal to 1, and $\underline{I}$ being the identity matrix of the order n.

Models (2) and (3)–(4) are simple ways of taking into account the fact that the errors $u_{ij}=v_i+e_{ij}=y_{ij}-\beta_1-\beta_2 x_{ij}$ are spatially correlated. This correlation is due to the spatial autocorrelation of the ground data y_{ij}, on the one hand, and of the remotely sensed data x_{ij}, on the other hand. Many quantitative geographical texts refer to the positive correlation of spatial variables, among them y_{ij}, as the first law of geography ["Everything is related to everything, but near things are more related than others" (Csillag and Kabos, 1999)]. The spatial correlation of the remotely sensed data x_{ij} (Labovitz and Masuoka, 1984; Webster et al., 1989) is induced by instruments: The sensors measure light reflectance from the Earth's surface, but this light is scattered so that reflectance from a pixel can be distributed over several contiguous pixels on the image (Haining, 1991;

Forster, 1980). The relationship between y_{ij} and x_{ij} could be exploited more efficiently if this spatial correlation were taken into account, which can be achieved in many different ways. Cokriging is one of them (Dungan, 1998). However, this technique requires a large sample size for empirical semivariogram estimation of both y_{ij} and x_{ij} (Curran, 1988) as well as an empirical cross semivariogram of y_{ij} and x_{ij} in order to assess spatial correlation as a function of the distance between segments [see Tokola et al. (1996) for other methods based on distance].

In this study the effect of spatial correlation is taken into account by specifying in Eq. (4) that the errors u_{ij} are positively correlated inside small areas and uncorrelated between small areas. The correlation coefficient, $\rho=\sigma_v^2/(\sigma_v^2+\sigma_e^2)$ is considered to be the same for every small area.

The BLUP Estimator

The BLUP of $\overline{Y}_i^*$, based on the sample size $n=\sum_{i=1}^{m} n_i$, is (Goldberger, 1962; Robinson, 1991; Cressie, 1991)

$$\hat{\overline{Y}}_i^*=\underline{\overline{X}}_i^*\hat{\underline{\beta}}+\hat{v}_i \tag{5}$$

where

$\underline{\overline{X}}_i^*=[1\overline{X}_i^*]$ where $\overline{X}_i^*$ is the mean of x_{ij} in the remaining (N_i-n_i) segments of the ith small area, not included in the sample,

$\hat{\underline{\beta}}=(\underline{X}^T\underline{\underline{V}^{-1}}\,\underline{X})^{-1}(\underline{X}^T\underline{\underline{V}^{-1}}\,\underline{Y})$ is the estimator of $\underline{\beta}=[\beta_1\beta_2]^T$,

where

$\underline{X}=[1\underline{x}]$ where $\underline{1}$ is a column vector $(n\times 1)$ of ones and $\underline{x}$ the column vector $(n\times 1)$ of the x_{ij} values in the sample,

$\underline{\underline{V}^{-1}}=diag(\underline{\underline{V}_1^{-1}},\underline{\underline{V}_2^{-1}},\ldots\underline{\underline{V}_i^{-1}},\ldots,\underline{\underline{V}_m^{-1}})$ is a block diagonal matrix with

$\underline{\underline{V}_i^{-1}}=\frac{1}{\sigma_e^2}\underline{\underline{I}_{(n_i)}}-\frac{g_i}{(n_i\sigma_e^2)}\underline{\underline{1}_{(n_i)}}\,\underline{\underline{1}_{(n_i)}^T}$, where $\underline{\underline{I}_{(n_i)}}$ is the identity matrix of order n_i and $\underline{\underline{1}_{(n_i)}}$ is a column vector $(n_i\times 1)$ of ones,

$\hat{v}_i=g_i(\overline{y}_i-\underline{\overline{x}_i}\,\hat{\underline{\beta}})$ is the BLUP of v_i (assuming that σ_v^2 and σ_e^2 are known), where $g_i=\sigma_v^2/[\sigma_v^2+(\sigma_e^2/n_i)]$ and $\underline{\overline{x}_i}=[1\overline{x}_i]$, while $\overline{x}_i$ is the sample mean of x_{ij}, in the ith small area,

$\underline{Y}$ is the column vector $(n\times 1)$ of the y_{ij} values in the sample.

Replacing $\hat{\overline{Y}}_i^*$ in Eq. (1) by that of Eq. (5) and ignoring the sampling ratio (i.e., assuming that n_i is small with respect to N_i), the BLUP estimator is found:

$$\hat{\overline{Y}}_i=(1-g_i)\underline{\overline{X}}_i\hat{\underline{\beta}}+g_i[\overline{y}_i+(\underline{\overline{X}}_i-\underline{\overline{x}_i})\hat{\underline{\beta}}] \tag{6}$$

where $\underline{\overline{X}}_i=[1\overline{X}_i]$ and $\overline{X}_i$ is the population mean of x_{ij} for the ith small area.

For the calculation of the estimates $\hat{\overline{Y}}_i^*$ it is not necessary to know $x_{ij'}$ for j' $(\neq j)$ from 1 to N_i-n_i, but only their total X_i, that is, the total land use area classified by remote sensing.

In general, the variance components σ_v^2 and σ_e^2 are unknown. For their estimation, several procedures have been proposed (Khuri and Sahai, 1985). By Henderson method 3, the following are unbiased estimators of σ_v^2 and σ_e^2 (Prasad and Rao, 1990):

$$\hat{\sigma}_e^2=\hat{\underline{e}}^T\hat{\underline{e}}/(n-m-1),$$

$$\hat{\sigma}_v^2=[\hat{\underline{u}}^T\hat{\underline{u}}-(n-2)\hat{\sigma}_e^2]/n^* \tag{7}$$

where [Eq. (8)]

$$n^*=n-trace\left\{(\underline{X}^T\underline{X})^{-1}\sum_{i=1}^{m}n_i^2\underline{\overline{X}}_i^T\underline{\overline{X}}_i\right\}=\sum_{i=1}^{m}n_i[1-n_i\underline{\overline{X}}_i(\underline{X}^T\underline{X})^{-1}\underline{\overline{X}}_i^T, \tag{8}$$

$\hat{\underline{e}}^T\hat{\underline{e}}$ is the residual sum of squares of model (2) fitted by Ordinary Least Squares and taking v_i as fixed, that is, the residual sum of squares of the Dummy Variable model, and $\hat{\underline{u}}^T\hat{\underline{u}}$ is the residual sum of squares of model (2) fitted by Ordinary Least Squares and taking $v_i=0$. The Dummy Variable model is $y_{ij}=\mu_i+\beta_2x_{ij}+e_{ij}$, with $\mu_i=(\beta_1+v_i)$. It is assumed that the intercept μ_i changes from one small area to another: This assumption is specified by associating a variable (Dummy) D_i with each μ_i $(i=1,2,\ldots,m)$, the value of which is 1 for every sample segment from the ith small area and 0 for the remaining segments in the sample $(y_{ij}=\mu_1D_1+\mu_2D_2+\ \ldots\ +\mu_iD_i+\ \ldots\ +\mu_mD_m+\beta_2x_{ij}+e_{ij}\ \forall i;$ $i=1,2,\ldots,m$; the variable D_i takes "n" values, the n_i values corresponding to the segments from the ith small area are equal to 1 and the remaining $n-n_i$ values are equal to 0).

Replacing σ_v^2 and σ_e^2 by $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$, the result is an estimator $\hat{\hat{\overline{Y}}}_i$ of the estimator $\hat{\overline{Y}}_i$, called the Empirical Best Linear Unbiased Predictor (EBLUP) estimator, and replacing σ_v^2 and σ_e^2 by $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ in Eq. (5), the result is an estimator $\hat{\underline{V}}$ of the variance and covariance matrix $\underline{V}$.

THE ESTIMATOR OF THE MEAN SQUARED ERROR OF THE EMPIRICAL BEST LINEAR UNBIASED PREDICTOR (EBLUP) ESTIMATOR

Assuming that the distribution of v_i and e_{ij} is normal, an approximately unbiased estimator of the Mean Squared Error of the EBLUP estimator, MSE $(\hat{\hat{\overline{Y}}}_i)$, is (Prasad and Rao, 1990; Ghosh and Rao, 1994)

$$\hat{MSE}(\hat{\hat{\overline{Y}}}_i)=(1-f_i)^2[h_{1i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)+h_{2i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)+2h_{3i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)] \tag{9}$$

where

$$h_{1i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=g_i\left(\frac{\sigma_e^2}{n_i}\right)+(1-f_i)^2\frac{(N_i-n_i)}{N_i^2}$$

$$h_{2i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=\hat{\sigma}_e^2(\underline{\overline{X}_i^*}-g_i\underline{\overline{x}_i})\underline{A}^{-1}(\underline{\overline{X}_i^*}-g_i\underline{\overline{x}_i})^T$$

$$h_{3i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=\frac{1}{n_i^2}\frac{1}{\left(\hat{\sigma}_v^2+\frac{\hat{\sigma}_e^2}{n_i}\right)^3}[(\hat{\sigma}_e^2)^2\mathrm{Var}(\hat{\sigma}_v^2)+(\hat{\sigma}_v^2)^2\mathrm{Var}(\hat{\sigma}_e^2)$$

$$-2\hat{\sigma}_e^2\hat{\sigma}_v^2 Cov(\hat{\sigma}_e^2,\hat{\sigma}_v^2)]$$

where

$$\underline{A}=\sum_{i=1}^{m}\left[\sum_{j=1}^{n_i}\underline{x}_{ij}^T\underline{x}_{ij}-g_i n_i\overline{\underline{x}}_i^T\overline{\underline{x}}_i\right]$$

$$\underline{x}_{ij}=[1x_{ij}]$$

$$Var(\hat{\sigma}_v^2)=\frac{2}{n_*^2}\left[\frac{1}{n-m-1}(m-1)(n-2)(\hat{\sigma}_e^2)^2+2n^*\hat{\sigma}_e^2\hat{\sigma}_v^2+n^{**}(\hat{\sigma}_v^2)^2\right]$$

$$Var(\hat{\sigma}_e^2)=\frac{2(\hat{\sigma}_e^2)^2}{n-m-1}$$

$$Cov(\hat{\sigma}_e^2,\hat{\sigma}_v^2)=-\frac{1}{n^*}(m-1)Var(\hat{\sigma}_e^2)$$

where

$$n^{**}=\sum_{i=1}^{m}n_i^2(1-n_i\overline{\underline{x}}_i\underline{A}_1^{-1}\overline{\underline{x}}_i^T)+trace\{(\underline{A}_1^{-1}\sum_{i=1}^{m}n_i^2\overline{\underline{x}}_i^T\overline{\underline{x}}_i)^2\}.$$

Because $\underline{A}_1=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\underline{x}_{ij}^T\underline{x}_{ij}$, n^{**} may be simplified to

$$n^{**}=\sum_{i=1}^{m}n_i^2[1-\overline{\underline{x}}_i(\underline{X}^T\underline{X})^{-1}\overline{\underline{x}}_i^T]=n^*-n+\sum_{i=1}^{m}n_i^2$$

RELATIVE EFFICIENCY OF THE EMPIRICAL BEST LINEAR UNBIASED PREDICTOR (EBLUP) ESTIMATOR

Direct Expansion Estimator

The Direct Expansion estimator is the design-based estimator defined by $\overline{y}_i=\sum_{j=1}^{n_i}y_{ij}/n_i$. It does not make use of the remote sensing data, only of the sample information on y_{ij}. The expected variance of this estimator is given by (Hansen et al., 1953, Vol. 2, Chap. 4, Sec. 17]

$$V(\overline{y}_i)=(1-f)S_i^2/(nP_i)+Q_iS_i^2/(nP_i)^2$$

where $P_i=N_i/N$, $Q_i=1-P_i$ and $S_i^2=\sum_{j=1}^{N_i}(y_{ij}-\overline{Y}_i)^2/(N_i-1)$.

Assuming that the variance within small areas S_i^2, is the same in every small area and equal to S_w^2, it can be estimated by $\hat{S}_w^2=\sum_{i=1}^{m}(n_i-1)s_i^2/(n-m)$, where $s_i^2=\sum_{j=1}^{m}(y_{ij}-\overline{y}_i)^2/(n_i-1)$.

An estimator, $\hat{V}(\overline{y}_i)$, of the variance of the Direct Expansion estimator, $V(\overline{y}_i)$, can be defined replacing S_i^2 by $\hat{S}_w^2$. The relative efficiency of the EBLUP estimator with respect to the Direct Expansion estimator would be estimated by $\hat{V}(\overline{y}_i)/\hat{MSE}(\hat{\overline{Y}}_i)$.

The Survey Regression and the Synthetic Regression Estimators

The estimator defined in Eq. (6) belongs to a class of estimators of the form $\hat{\overline{Y}}_i(\delta_i)=\overline{\underline{X}}_i\tilde{\underline{\beta}}+\delta_i(\overline{y}_i-\overline{\underline{x}}_i\tilde{\underline{\beta}})$, where $\tilde{\underline{\beta}}$ is an estimator of $\underline{\beta}$ and δ_i is a nonnegative constant (Harter, 1983). This class is interesting because the most usual estimators are elements of the class. For $\tilde{\underline{\beta}}=\hat{\underline{\beta}}$ and $\delta_i=g_i$ the BLUP estimator is defined in (6). For $\tilde{\underline{\beta}}=\hat{\underline{\beta}}$ and $\delta_i=0$, the Synthetic Regression estimator $\hat{\overline{Y}}_i(0)=\overline{\underline{X}}_i\hat{\underline{\beta}}$ is the result. For $\tilde{\underline{\beta}}=\hat{\underline{\beta}}$ and $\delta_i=1$, $\hat{\overline{Y}}_i(1)=\overline{y}_i+(\overline{\underline{X}}_i-\overline{\underline{x}}_i)\hat{\underline{\beta}}$ is the Survey Regression estimator.

The Mean Squared Error (MSE) of the estimators $\hat{\overline{Y}}_i(0)$ and $\hat{\overline{Y}}_i(1)$ can be expressed as a function of the Mean Squared Error of the BLUP (Harter, 1983), $MSE(\hat{\overline{Y}}_i)$:

$$MSE[\hat{\overline{Y}}_i(0)]=MSE(\hat{\overline{Y}}_i)+g_i^2\left[\sigma_v^2+\frac{\sigma_e^2}{n_i}-\overline{\underline{x}}_i(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\overline{\underline{x}}_i^T\right]$$

$$MSE[\hat{\overline{Y}}_i(1)]=MSE(\hat{\overline{Y}}_i)+(1-g_i^2)\left[\sigma_v^2+\frac{\sigma_e^2}{n_i}-\overline{\underline{x}}_i(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\overline{\underline{x}}_i^T\right] \quad (10)$$

These MSE can be estimated replacing $MSE(\hat{\overline{Y}}_i)$ by $\hat{MSE}(\hat{\overline{Y}}_i)$ and $\underline{V}$ by $\hat{\underline{V}}$. The relative efficiency of the EBLUP estimator with respect to the Synthetic Regression and the Survey Regression estimators will be estimated by $\hat{MSE}(\hat{\overline{Y}}_i(0))/\hat{MSE}(\hat{\overline{Y}}_i)$ and $\hat{MSE}(\hat{\overline{Y}}_i(1))/\hat{MSE}(\hat{\overline{Y}}_i)$, respectively, and with respect to the Direct Expansion estimator by $\hat{V}(\overline{y}_i)/\hat{MSE}(\hat{\overline{Y}}_i)$. The relative efficiency of the Survey Regression estimator with respect to the Direct Expansion estimator will be estimated by $RE_i=\hat{V}(\overline{y}_i)/\hat{MSE}(\hat{\overline{Y}}_i(1))$.

Note that the estimator in (6) is a weighted mean of the extreme estimators $\hat{\overline{Y}}_i(0)$ and $\hat{\overline{Y}}_i(1)$, the weights being g_i and $(1-g_i)$, respectively. When $\sigma_v^2=0$, then $g_i=0$ and the BLUP is reduced to the Synthetic Regression estimator $\hat{\overline{Y}}_i(0)$ and its mean squared error is estimated by Eq. (12) with $\hat{\sigma}_v^2=0$. When $\sigma_v^2>0$, then $g_i\neq 0$ and the BLUP estimator lies between the extreme Synthetic Regression estimator and the Survey Regression estimator, approaching one or the other depending on $\rho=\sigma_v^2/(\sigma_v^2+\sigma_e^2)$ (the correlation coefficient inside small areas) and on the sample size n_i. When σ_v^2 is small with regard to $(\sigma_v^2+\sigma_e^2)$ (i.e., ρ is small), then g_i tends to be low, except when n_i is very high so that on the basis of (10) $\hat{MSE}(\hat{\overline{Y}}_i(0))/\hat{MSE}(\hat{\overline{Y}}_i)$ tends toward 1 and the Synthetic estimator tends to be as efficient as the BLUP estimator. When σ_v^2 is large with regard to $(\sigma_v^2+\sigma_e^2)$ and n_i is high, then g_i is far from zero and the BLUP estimator tends toward the Survey Regression estimator, $\hat{\overline{Y}}_i(1)$ [i.e., on the basis of (10), $\hat{MSE}(\hat{\overline{Y}}_i(1))/\hat{MSE}(\hat{\overline{Y}}_i)$ tends toward 1].

The null hypothesis $\sigma_v^2=0$ versus the alternative hypothesis $\sigma_v^2>0$ can be tested using the statistic:

$$\lambda_{lm}=\frac{n}{2(\overline{n}-1)}\left[\frac{\sum_{i=1}^{m}\left(\sum_{j=1}^{n_i}\hat{u}_{ij}\right)^2}{\sum_{i=1}^{m}\sum_{j=1}^{n_i}\hat{u}_{ij}^2}-1\right]^2 \quad (11)$$

where $\hat{u}_{ij}$ is defined in Eq. (7) and $\overline{n}=n/m$. This statistic is distributed asymptotically as χ^2 with one degree of freedom (Judge et al., 1985).

Table 1. Case Study Results for Irrigated Corn

Small Area	*Sample Size n_i (segm.)*	*Geographical Surface (ha)*	*Estimate $\hat{\bar{Y}}_i$ (ha/segm.)*	$\hat{g}_i$	*Standard Error*				*Relative Efficiency*			
					EBLUP	*Synthetic Regression*	*Syrvey Regression*	*Direct Expansion*	*Surv. Reg./ Dir. Exp.*	*EBLUP/ Surv. Reg.*	*EBLUP/ Dir. Exp.*	*Gain Due to the Remotely Sensed Data*
Agueda	1	1671	0.96	0.08	0.71	0.74	2.37	4.28	3.26	11.14	36.34	10.34
Alba de Tormes	1	311	1.36	0.08	0.70	0.72	2.37	4.29	3.28	11.46	37.56	9.53
Campillo	1	126	0.72	0.08	0.63	0.66	2.35	4.29	3.33	13.91	46.37	9.00
Geria-Villamarciel	1	590	0.95	0.08	0.71	0.73	2.37	4.28	3.26	11.14	36.34	9.97
Marcías-Picaveas	1	2304	1.07	0.08	0.72	0.74	2.37	4.28	3.26	10.84	35.34	10.41
Pollos	1	1029	1.86	0.08	0.71	0.73	2.37	4.28	3.26	11.14	36.34	10.31
Zuzones	1	465	0.85	0.08	0.71	0.73	2.37	4.29	3.28	11.14	36.51	9.83
Aranda	2	3285	0.82	0.14	0.70	0.74	1.68	2.61	2.41	5.76	13.90	6.89
Bajo Carrión	2	6201	0.83	0.14	0.70	0.75	1.69	2.61	2.39	5.83	13.90	6.93
Carrizo	2	3350	1.37	0.14	0.70	0.74	1.69	2.61	2.39	5.83	13.90	6.93
Castilla Sur	2	3763	0.94	0.14	0.70	0.74	1.69	2.61	2.39	5.83	13.90	6.91
Florida	2	1205	1.97	0.14	0.68	0.73	1.68	2.62	2.43	6.10	14.85	6.82
La Maya	2	2190	1.51	0.14	0.69	0.74	1.68	2.62	2.43	5.93	14.42	6.89
La Retención	2	5197	0.83	0.14	0.70	0.75	1.69	2.61	2.39	5.83	13.90	6.93
La Vid	2	345	0.72	0.14	0.63	0.68	1.66	2.62	2.49	6.94	17.30	6.48
Presa de Tierra	2	3852	6.83	0.14	0.74	0.79	1.70	2.61	2.36	5.28	12.44	6.13
San Román y San Justo	2	321	1.43	0.14	0.62	0.67	1.65	2.62	2.52	7.08	17.86	6.51
Toro-Zamora	2	7304	5.19	0.14	0.70	0.75	1.68	2.61	2.41	5.76	13.90	6.87
Velilla	2	1470	3.02	0.14	0.69	0.73	1.68	2.62	2.43	5.93	14.42	6.86
Villalazán	2	1656	1.53	0.14	0.69	0.74	1.68	2.62	2.43	5.93	14.42	6.84
Villoria	2	3909	0.85	0.14	0.70	0.74	1.69	2.61	2.39	5.83	13.90	6.91
Babilafuente	3	3691	0.98	0.20	0.68	0.74	1.38	2.01	2.12	4.12	8.74	5.29
Inés	3	1772	1.58	0.20	0.67	0.73	1.38	2.01	2.12	4.24	9.00	5.22
Tramo Hidoreléctrico	3	672	1.27	0.20	0.63	0.70	1.36	2.01	2.18	4.66	10.18	5.15
Villagonzalo	3	4400	1.38	0.20	0.68	0.74	1.38	2.01	2.12	4.12	8.74	5.32
Almar	4	1763	0.94	0.25	0.64	0.73	1.19	1.69	2.02	3.46	6.97	4.32
Arriola	4	5661	4.58	0.25	0.66	0.74	1.18	1.68	2.03	3.20	6.48	4.40
Castañón	4	2584	7.82	0.25	0.66	0.74	1.15	1.69	2.16	3.04	6.56	4.18
Nava de Campos	4	4885	0.72	0.25	0.66	0.74	1.21	1.69	1.95	3.36	6.56	4.37
Palencia	4	3097	0.90	0.25	0.66	0.74	1.20	1.69	1.98	3.31	6.56	4.35
Riaza	4	5543	0.72	0.25	0.66	0.75	1.21	1.68	1.93	3.36	6.48	4.37
Tera	4	5762	2.03	0.25	0.66	0.74	1.21	1.68	1.93	3.36	6.48	4.41
Tordesillas	4	2196	2.16	0.25	0.64	0.73	1.20	1.69	1.98	3.52	6.97	4.37
Villadangos	4	3062	5.35	0.25	0.67	0.75	1.21	1.69	1.95	3.26	6.36	4.16
Villares	4	6016	4.19	0.25	0.66	0.75	1.21	1.68	1.93	3.36	6.48	4.38
Almazán	5	5537	0.67	0.30	0.64	0.74	1.08	1.48	1.88	2.85	5.35	3.78
Castilla Norte	5	7658	0.85	0.30	0.65	0.74	1.09	1.48	1.84	2.81	5.18	3.79
Villalaco	5	4118	0.88	0.30	0.64	0.74	1.08	1.48	1.88	2.85	5.35	3.78
Arlanzón	6	2091	0.60	0.34	0.60	0.72	0.98	1.33	1.84	2.67	4.91	3.35
Porma Margen Izda	7	6162	5.53	0.37	0.61	0.74	0.92	1.22	1.76	2.27	4.00	3.07
San José	7	5811	5.14	0.37	0.61	0.74	0.93	1.22	1.72	2.32	4.00	3.03
Manganeses	8	5294	3.56	0.40	0.59	0.73	0.86	1.13	1.73	2.12	3.67	2.28
Pisuerga	8	9973	0.57	0.40	0.60	0.73	0.87	1.13	1.69	2.10	3.55	2.88
Esla	9	11454	6.37	0.43	0.60	0.74	0.83	1.06	1.63	1.91	3.12	2.59
Castilla Ramal de Campos	10	10402	0.65	0.46	0.57	0.73	0.78	1.00	1.64	1.87	3.08	2.58
Carrión-Saldaña	13	10361	0.48	0.52	0.53	0.72	0.69	0.87	1.59	1.69	2.69	2.30
Páramo	17	24381	10.25	0.59	0.52	0.69	0.61	0.75	1.51	1.38	2.08	1.91

NONSAMPLED AREAS

In small areas where $n_i=0$, the Synthetic Regression estimator is used as the estimator of $\overline{Y}_i$, and the mean squared error is estimated by

$$\hat{MSE}(\hat{\overline{Y}}_i)=\underline{\overline{X}}_i(\underline{X}^T\hat{\underline{V}}^{-1}\underline{X})^{-1}\underline{\overline{X}}_i^T+\hat{\sigma}_v^2 \qquad (12)$$

RELATIVE EFFICIENCY GAINED FROM REMOTE SENSING DATA

A measure of the gain in precision due to remote sensing data is the following $RE_i^*=\hat{MSE}(\hat{\overline{Y}}_i)_{gd}/\hat{MSE}(\hat{\overline{Y}}_i)$, where $\hat{MSE}(\hat{\overline{Y}}_i)$is as defined in Eq. (9) and $\hat{MSE}(\hat{\overline{Y}}_i)_{gd}$ is the Mean Squared Error of the EBLUP estimator of $\overline{Y}_i$ obtained using the same sample of segments and based on a model ignoring x_{ij}: $y_{ij}=\beta_1+v_i+e_{ij}$ in (2).

The EBLUP estimator of $\overline{Y}_i$ and the estimator $\hat{MSE}(\hat{\overline{Y}}_i)_{gd}$, based on the latest model, are obtained as specific cases of that obtained from model (2), substituting $\underline{X}=\underline{1}$, $\underline{\beta}=\beta_1$, $\underline{\hat{\beta}}=\hat{\beta}_1$, $\overline{\underline{X}}_i=1$, $\overline{x}_i=1$, $\underline{x}_{ij}=1$ and also replacing $(n-m-1)$ by $(n-m)$ and $(n-2)$ by $(n-1)$.

A CASE STUDY

The basin of the Duero river, which flows through Spain and Portugal into the Atlantic Ocean, has been taken as a case study. In order to manage the water resources for irrigation, the basin is divided into small areas called "irrigation zones," following agricultural and administrative criteria: The number of zones in the Spanish part of the basin is $m=53$. Estimates of irrigated crop acreage in each one of these "zones" are required for the main crops (corn, sunflower, and sugar beet) in order to estimate the water irrigation requirements in each "zone." The estimates are to be based on ground and satellite data. As a sampling frame the UTM (Universal Transversa Mercator) grid is used. The segment, or sampling unit, is a squared cell of 500 m×500 m, that is, 25 hectares. Each cell is identified by the UTM coordinates in the southwest corner. A random sample of $n=158$ segments is selected from among the whole population. The number of segments in the sample from a small area or "zone" ranges from zero to 17, the average being 4.

The Estimates

Table 1 shows the estimates for irrigated corn crop acreage, computed by using expression (6) and its standard error [square root of the estimated mean squared error of the estimator computed by using Eq. (9)] as well as the standard error of the remaining estimators. For unsampled small areas, Table 2 shows the Synthetic Regression estimates and their standard error [square root of the estimated mean squared error of the estimator, computed by using Eq. (12)].

Table 2. Estimates for Irrigated Corn in Unsampled Small Areas

Small Area	*Geographical Surface (ha)*	*Estimate (ha/segm.)*	*Standard Error*
Adearregada	601	1.58	0.72
Ejeme Galizancho	738	1.70	0.72
Guma	4157	0.98	0.72
Olmillos	233	0.98	0.72
Villamayor	596	1.18	0.72
Zorita	390	1.12	0.72

Relative Efficiency between Estimators

The EBLUP estimator is the most efficient of the four estimators. The worst estimator is the "Direct Expansion estimator." The relative efficiency of the Synthetic and the Survey regression estimators with regard to the EBLUP estimator depends on g_i [Eq. (5)], which in turn depends on the ratio between the variance components σ_v^2 and σ_e^2 as well as on the sample size n_i. Table 3 shows the estimates of these variance components for the three crops considered. The values of λ_{1m} observed when using Eq. (11) are 6.9713 for corn, 5.6046 for sugar beet, and 18.8368 for sunflower so that σ_v^2 is significantly different from zero [with a significance level of 5% for sugar beet ($\chi^2_{0.95}(1)=3.84$) and 1% for corn and sunflower ([$\chi^2_{0.99}(1)=6.63$)]. Hence, on the basis of Eq. (5), g_i is different from zero (when $n_i\neq 0$). The ratio $\hat{\sigma}_v^2/(\hat{\sigma}_e^2+\hat{\sigma}_v^2)$ is 0.08 for corn and sugar beet, but 0.21 for sunflower. Hence, $\hat{g}_i$ is near zero for corn and sugar beet, and on the basis of Eqs. (6) and (10), the Synthetic estimator is nearer the EBLUP estimator than the Survey Regression estimator, except when n_i is high (in Paramo, where $n_i=17$, the Survey Regression estimator outperforms the Synthetic Regression estimator). For sunflower $\hat{g}_i$ is far from zero and increases when n_i increases so the Survey Regression estimator is nearer the EBLUP estimator than the Synthetic Regression estimator, except when $n_i<4$ (or $g_i<0.5$).

The relative efficiency, RE_i, of the remote sensing data ranged i) from 1.51 and 3.26 for corn, with the average being 2.25, ii) from 1.17 and 2.38 for sunflower, with an average of 1.58, and iii) from 1.91 and 4.13 for sugar beet, with an average of 2.79. These figures are of the order found for large areas (Ambrosio et al., 1993). The differences between the crops are explained by the fact that the spectral signatures of corn and sugar beet crops are more

Table 3. Estimates of the Variance Components and (within Brackets) of Its Standard Error

	Variance Components	
Crop	$\hat{\sigma}_v^2$ *(Standard Error of* $\hat{\sigma}_v^2$*)*	$\hat{\sigma}_e^2$ *(Standard Error of* $\hat{\sigma}_e^2$*)*
Corn	0.4725 (0.4540)	5.5944 (0.6706)
Sugar beet	0.5012 (0.4427)	5.4780 (0.6571)
Sunflower	0.7511 (0.3449)	2.7614 (0.3312)

specific than the spectral signature of sunflower, which is usually confused with ploughed land ready for sowing.

It is suggested to measure the gain due to the remotely sensed data using a given estimator, the EBLUP estimator, as indicated above: $RE_i^* = \hat{MSE}(\hat{\bar{Y}}_i)_{gd} / \hat{MSE}(\hat{\bar{Y}}_i)$. Table 1 shows this measure for corn: it ranges from 1.91 to 10.41, with the average being 5.63. (For sunflower it ranges from 1.20 to 1.56, with the average being 1.37, and for sugar beet from 1.31 to 2.04, with the average being 1.62). Another index proposed to measure this gain is $n_i^* = n_i RE_i^*$, where n_i^* is the segment sample size required when satellite data are not used to achieve the same precision as with n_i segments and satellite data.

For the estimation of corn acreage without remote sensing data, it would, on average, be necessary have a sample size of $n_i^* = 5.63 n_i$ segments in order to achieve the same precision as with n_i segments and remote sensing data ($n_i^* = 1.37 n_i$ for sunflower and $n_i^* = 1.62 n_i$ for sugar beet).

Financial support from the Ministerio de Agricultura Pesca y Alimentación [≈ Ministry of Agriculture, Fishing and Food] (Spain) for Project N° SC93-056 as well as from Tragsatec is gratefully acknowledged. We wish to thank R. Escudero, J. Fernández, and A. Porcuna for the treatment of the satellite images, and the referees for their very helpful comments and suggestions.

REFERENCES

Allen, J. D. (1990), A look at the Remote Sensing Applications Program of the National Agricultural Statistics Service. *J. Official Stat.* 6:393–409.

Ambrosio, L., Alonso, R., and Villa, A. (1993), Estimación de superficies cultivadas por muestreo de áreas y teledetección. Precisión relativa. Estadística Española 35, 132: 91–103.

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988), An error-components model for prediction of county crop areas using surveys and satellite data. *J. Am. Stat. Assoc.* 83:28–36.

Csillag, F., and Kabos, S. (1999), Convergence between GIS and spatial statistics: What? Where? When? Proceedings of the 52nd Session. *Bull. Int. Stat. Inst.* Book 1:249–252.

Cochran, W. G. (1977), *Sampling Techniques*, Wiley, New York.

Cressie, N. A. C. (1991), *Statistics for Spatial Data*, Wiley, New York.

Curran, P. J. (1988), The semivariogram in remote sensing: an introduction. *Remote Sens. Environ.* 24:493–507.

Deppe, F. (1998), Forest area estimation using sample surveys and Landsat MSS and TM data. *Photogramm. Eng. Remote Sens.* 64:285–292.

Dungan, J. (1998), Spatial prediction of vegetation quantities using ground and image data. *Int. J. Remote Sens.* 19:267–285.

Forster, B. C. (1980), Urban residential ground cover using Landsat digital data. *Photogramm. Eng. Remote Sens.* 46:547–558.

Ghosh, M., and Rao, J. N. K. (1994), Small area estimation: an appraisal. *Stat. Sci.* 9:55–93.

Goldberger, A. S.(1962), Best linear unbiased prediction in the generalized linear regression model. *J. Am. Stat. Assoc.* 57:369–375.

Haining, R. (1990), Spatial Data Analysis in the Social and Environmental Sciences, Cambridge University Press, Cambridge.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953), Sample Survey. Methods and Theory. Vol. I: Methods and Applications. Vol. II. Theory, Wiley, New York.

Hanuschak, G. A., Allen, R. D., and Wigton, W. H. (1982), Integration of Landsat data into the crop estimation program of USDA's Statistical Reporting Service. invited paper at the 1982 Symposium, Purdue University, West Lafayette, IN.

Harter, R. M. (1983), Small area estimation using nested-error models and auxiliary data. Ph. D. Iowa State University.

Judge, G. G., Griffiths, W. E., Carter, R., Lütkepohl, H., and Lee, T. (1985), *The Theory and Practice of Econometrics*, Wiley, New York.

Khuri, A. I., and Sahai, H. (1985), Variance components analysis: a selective literature survey. *Int. Stat. Rev.* 53:279–300.

Labovitz, M. L., and Masuoka, E. J. (1984), The influence of autocorrelation in signature extraction—an example from a geobotanical investigation of Cotter Basin, Montana. *Int. J. Remote Sens.* 5:315–332.

Prasad, N. G. N., and Rao, J. N. K. (1990), The estimation of the mean squared error of small-area estimators. *J. Am. Stat. Assoc.* 85:163–171.

Robinson, G. K. (1991), That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* 6:15–51.

Royall, R. M. (1970), On finite population sampling theory under certain linear regression models. *Biometrika* 57:377–387.

Royall, R. M., and Herson, J. (1973), Robust estimation in finite populations I and II. *J. Am. Stat. Assoc.* 68:880–893.

Tokola, T., Pitkänen, J., Partinen, S., and Muinonen, E. (1996), Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. *Int. J. Remote Sens.* 17:2333–2351.

Webster, R., Curran, P. J., and Munden, W. J. (1989). Spatial correlation in reflected radiation from the ground and its implications for sampling and mapping by ground-based radiometry. *Remote Sens Environ.* 29:67–78.

APPENDIX

Numerical Example

Four small areas were considered: 1, 2, 3, and 4. The segment sample size in each small area is 1, 4, 2, and 1, respectively.

The Data

The required data are shown in Tables A.1 and A.2. Table A.1 shows the ground, y_{ij}, and the remotely sensed data, x_{ij}, in each sampling segment. Table A.2 shows the size of each small area (number of segments, N_i) and the mean

Table A.1. Sample Data

Small Area	y_{ij}	x_{ij}
1	1.04	0.10
2	4.56	0.90
2	3.96	0.00
2	7.20	4.78
2	4.19	0.55
3	3.55	7.44
3	1.28	5.70
4	2.05	0.30

Table A.2. Population Data

Small Area	*Number of Segments* (N_i)	*Small Area Mean per Segment Classified by Remote Sensing* ($\bar{X}_i$)
1	12	1.05
2	71	1.91
3	131	4.23
4	14	1.5

per segment of the surface classified by remote sensing as the crop type, $\bar{X}_i$.

Verification of the Model Assumption

Using data from Table A.1, the value of the statistic defined in Eq. (11) is calculated. The vector of ordinary least square (OLS) residuals of model (3) is

$$\underline{\hat{u}}=\underline{Y}-\underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$$

where:

$$\underline{Y}=[1.04 \quad 4.56 \quad 3.96 \quad 7.20 \quad 4.19 \quad 3.55 \quad 1.28 \quad 2.05]^T$$

$$\underline{X}=\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.10 & 0.90 & 0.00 & 4.78 & 0.55 & 7.44 & 5.70 & 0.30 \end{bmatrix}^T$$

So that:

$$\underline{\hat{u}}=[-2.1952 \quad 1.2427 \quad 0.7351 \quad 3.4841 \quad 0.9086 \quad -0.4392 \quad -2.5304 \quad -1.2057]^T$$

$$\underline{\hat{u}}^T\underline{\hat{u}}=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\hat{u}_{ij}^2=27.9176$$

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{n_i}\hat{u}_{ij}\right)^2=55.6744$$

Since the sample size is eight, $n=8$, and the number of small areas four, $m=4$, the average sample size per small area is two, $\bar{n}=8/4=2$, and replacing n and $\bar{n}$ in Eq. (11), the result is $\lambda_{lm}=3.9541$. Since $\chi^2_{0.95}(1)=3.84$, the null hypothesis, $\sigma_v^2=0$, is rejected and the model assumption accepted.

The Estimates

In order to estimate the mean per segment using (6), estimates of $\underline{\beta}$ and g_i are required. Estimates of σ_v^2 and σ_e^2 must be calculated, using (7). The vector of OLS residuals of the Dummy Variable model (2) is

$$\underline{\hat{e}}=\underline{Y}-\underline{\tilde{X}}(\underline{\tilde{X}}^T\underline{\tilde{X}})^{-1}\underline{\tilde{X}}^T\underline{Y}$$

where

$$\underline{\tilde{X}}=\begin{Bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.10 & 0.90 & 0.00 & 4.78 & 0.55 & 7.44 & 5.70 & 0.30 \end{Bmatrix}^T$$

so that

$$\underline{\hat{e}}=[0.0000 \quad 0.0740 \quad 0.1468 \quad -0.1864 \quad -0.0344 \quad 0.4847 \quad -0.4847 \quad 0.0000]^T$$

$$\underline{\hat{e}}^T\underline{\hat{e}}=0.5328$$

$\underline{\bar{x}}_i=[1 \ \bar{x}_i]$, where $\bar{x}_i$ is the sample mean of the x_{ij} from Table A.1: 0.1, 1.5575, 6.57, and 0.3, for the small areas 1, 2, 3, and 4, respectively,

$\underline{\bar{x}}_i(\underline{X}^T\underline{X})^{-1}\underline{\bar{x}}_i^T$ is equal to 0.2142, 0.1382, 0.3915, and 0.1998 for the small areas 1, 2, 3, and 4, respectively.

Hence, $n_*=3.8088$, $\sigma_e^2=0.1776$ and $\sigma_v^2=7.05$.

The estimates of g_i are $\hat{g}_1=0.9754$; $\hat{g}_2=0.9937$; $\hat{g}_3=0.9876$; and $\hat{g}_4=0.9754$ for the small areas 1, 2, 3 and 4, respectively. The estimates of $\underline{V}_i^{-1}$ are

$$\underline{\hat{V}}_1^{-1}=0.1385;$$

$$\underline{\hat{V}}_2^{-1}=\frac{1}{0.1776}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}-\frac{0.9937}{4\times 0.1776}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$=\begin{bmatrix} 4.2318 & -1.3988 & -1.3988 & -1.3988 \\ -1.3988 & 4.2318 & -1.3988 & -1.3988 \\ -1.3988 & -1.3988 & 4.2318 & -1.3988 \\ -1.3988 & -1.3988 & -1.3988 & 4.2318 \end{bmatrix}$$

$$\underline{\hat{V}}_3^{-1}=\frac{1}{0.1776}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}-\frac{0.9876}{2\times 0.1776}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$=\begin{bmatrix} 2.8502 & -2.7804 \\ -2.7804 & 2.8502 \end{bmatrix}$$

and $\underline{\hat{V}}_4^{-1}=0.1385$, for the small areas 1, 2, 3, and 4, respectively.

Using $\underline{\hat{\beta}}=(\underline{X}^T\underline{\hat{V}}^{-1}\underline{X})^{-1}(\underline{X}^T\underline{\hat{V}}^{-1}\underline{Y})$ where $\underline{X}$ and $\underline{T}$ are as above and $\underline{\hat{V}}^{-1}=diag(\underline{\hat{V}}_1^{-1} \ \underline{\hat{V}}_2^{-1} \ \underline{\hat{V}}_3^{-1} \ \underline{\hat{V}}_4^{-1})=$

$$\begin{bmatrix} 0.1385 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 4.2318 & -1.3988 & -1.3988 & -1.3988 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -1.3988 & 4.2318 & -1.3988 & -1.3988 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -1.3988 & -1.3988 & 4.2318 & -1.3988 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -1.3988 & -1.3988 & -1.3988 & 4.2318 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 2.8502 & -2.7804 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -2.7804 & 2.8502 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.1385 \end{bmatrix},$$

the estimate of $\underline{\beta}$ is calculated:

$$\underline{\hat{\beta}}=\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}=\begin{bmatrix} 1.0954 \\ 0.7195 \end{bmatrix}$$

Replacing in (6) g_i by $\hat{g}_i$, $\hat{\hat{\bar{Y}}}_i=(1-\hat{g}_i)(\hat{\beta}_1+\hat{\beta}_2\bar{X}_i)+\hat{g}_i[\bar{y}_i+\hat{\beta}_2(\bar{X}_i-\bar{x}_i)]$, the estimates of the mean per segment in each small area are calculated: $\hat{\hat{\bar{Y}}}_1=2.4462$; $\hat{\hat{\bar{Y}}}_2=5.2137$; $\hat{\hat{\bar{Y}}}_3=0.7736$; and $\hat{\hat{\bar{Y}}}_4=2.8952$ for the small areas 1, 2, 3, and 4, respectively. The estimates of the total are $\hat{\hat{Y}}_1=12\times 2.4475=29.3544$, $\hat{\hat{Y}}_2=71\times 5.2137=370.1727$, $\hat{\hat{Y}}_3=131\times 0.7736=101.3416$ and $\hat{\hat{Y}}_4=14\times 2.8952-40.5328$ for the small areas 1, 2, 3, and 4, respectively.

The Mean Squared Error Estimate

From Eq. (9), the three components of the mean squared error (MSE) of the estimator of the mean per segment are calculated.

The first component of the estimate of the MSE is $h_{11}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=0.2629$, $h_{12}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=0.059$, $h_{13}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=0.0954$ and $h_{14}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=0.2494$ for the small areas 1, 2, 3, and 4, respectively.

For the second component, the value of the vector $\overline{\underline{X}}_i^*=[1\ \overline{X}_i^*]$ where $\overline{X}_i^*=(N_i\overline{X}_i-n_i\overline{x}_i)/(N_i-n_i)$, is required. Using data from Table A.2: $\overline{X}_1^*=2.2273$; $\overline{X}_2^*=1.9310$; $\overline{X}_3^*=4.1937$ and $\overline{X}_4^*=1.5923$, for the small areas 1, 2, 3, and 4, respectively.

In order to calculate the matrix $\underline{A}$, it is necessary to calculate the matrix $\underline{x}_{ij}^T\underline{x}_{ij}$, where $\underline{x}_{ij}=[1\ x_{ij}]$, where x_{ij} is given in Table A.1. For small area 1:

$$\underline{x}_{11}^T\underline{x}_{11}=\begin{bmatrix}1 & 0.10\\ 0.10 & 0.01\end{bmatrix}.$$

For small area 2:

$$\underline{x}_{21}^T\underline{x}_{21}=\begin{bmatrix}1 & 0.90\\ 0.90 & 0.81\end{bmatrix};\ \underline{x}_{22}^T\underline{x}_{22}=\begin{bmatrix}1 & 0.00\\ 0.00 & 0.00\end{bmatrix};$$

$$\underline{x}_{23}^T\underline{x}_{23}=\begin{bmatrix}1 & 4.78\\ 4.78 & 22.8484\end{bmatrix};\ \underline{x}_{24}^T\underline{x}_{24}=\begin{bmatrix}1 & 0.55\\ 0.55 & 0.3025\end{bmatrix}$$

for each of the four sample observations. For small area 3:

$$\underline{x}_{31}^T\underline{x}_{31}=\begin{bmatrix}1 & 7.44\\ 7.44 & 55.3536\end{bmatrix};\ \underline{x}_{32}^T\underline{x}_{32}=\begin{bmatrix}1 & 5.70\\ 5.70 & 32.49\end{bmatrix}.$$

And for small area 4:

$$\underline{x}_{41}^T\underline{x}_{41}=\begin{bmatrix}1 & 0.30\\ 0.30 & 0.09\end{bmatrix}$$

Hence, $\sum_{j=1}^{n_i}\underline{x}_{ij}^T\underline{x}_{ij}$ is

$$\begin{bmatrix}1 & 0.10\\ 0.10 & 0.01\end{bmatrix}$$

for small area 1; for small area 2 it is

$$\begin{bmatrix}4 & 6.23\\ 6.23 & 23.9606\end{bmatrix}$$

for small area 3 it is

$$\begin{bmatrix}2 & 13.14\\ 13.14 & 87.8436\end{bmatrix}$$

and for small area 4 it is

$$\begin{bmatrix}1 & 0.30\\ 0.30 & 0.09\end{bmatrix}$$

For each small area, $\overline{\underline{x}}_i^T\overline{\underline{x}}_i$ is also calculated, where $\overline{\underline{x}}_i=[1\ \overline{x}_i]$. For small area 1, $\overline{\underline{x}}_i^T\overline{\underline{x}}_i$ is

$$\begin{bmatrix}1 & 0.10\\ 0.10 & 0.01\end{bmatrix},$$

for small area 2 it is

$$\begin{bmatrix}1 & 1.5575\\ 1.5575 & 2.4258\end{bmatrix},$$

for small area 3 it is

$$\begin{bmatrix}1 & 6.57\\ 6.57 & 55.1649\end{bmatrix},$$

and for small area 4 it is

$$\begin{bmatrix}1 & 0.30\\ 0.30 & 0.09\end{bmatrix}.$$

Hence, for small area 1, $\sum_{j=1}^{n_i}\underline{x}_{ij}^T\underline{x}_{ij}-g_in_i\underline{\overline{x}}_i^T\underline{\overline{x}}_i$ is

$$\begin{bmatrix}0.0246 & 0.0025\\ 0.0025 & 0.0002\end{bmatrix}$$

and is equal to

$$\begin{bmatrix}0.0252 & 0.0392\\ 0.0392 & 14.3188\end{bmatrix},\ \begin{bmatrix}0.0248 & 0.1629\\ 0.1629 & 2.5843\end{bmatrix},\ \begin{bmatrix}0.0246 & 0.0074\\ 0.0074 & 0.0022\end{bmatrix},$$

for the small areas 2, 3, and 4, respectively.

The matrix $\underline{A}$ is the sum of these four last matrices, each corresponding to a small area:

$$\underline{A}=\begin{bmatrix}0.0992 & 0.2120\\ 0.2120 & 16.9055\end{bmatrix},\ \text{and}\ \underline{A}^{-1}=\begin{bmatrix}10.3582 & -0.1299\\ -0.1299 & 0.0608\end{bmatrix}.$$

For small area 1, the second component of the mean squared error is $h_{21}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=0.0477$ and $h_{22}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=0.0015$, $h_{23}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=0.0584$ and $h_{24}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)=0.0179$, for the small areas 2, 3, and 4, respectively.

For the third component, the values of n_{**} are needed as well as the estimates of the variances, $\text{Var}(\hat{\sigma}_e^2)$ and $Var(\hat{\sigma}_v^2)$, and the covariance, $Cov(\hat{\sigma}_e^2,\hat{\sigma}_v^2)$: $n_{**}=17.8088$, $\hat{V}ar(\hat{\sigma}_e^2)=0.021$, $\hat{V}ar(\hat{\sigma}_v^2)=123.3710$ and $\hat{C}ov(\hat{\sigma}_e^2,\hat{\sigma}_v^2)=-0.0165$. For small area 1, the third component is $h_{31}(\hat{\sigma}_e^2,\hat{\sigma}_v^2)=0.0132$. In the same way, the result is $h_{32}(\hat{\sigma}_e^2,\hat{\sigma}_v^2)=0.0009$; $h_{33}(\hat{\sigma}_e^2,\hat{\sigma}_v^2)=0.0034$ and $h_{34}(\hat{\sigma}_e^2,\hat{\sigma}_v^2)=0.0132$ for the small areas 2, 3, and 4, respectively.

Finally, for small area 1, the mean squared error of the estimator of the mean per segment is: $\hat{MSE}(\hat{\overline{Y}}_1)=0.3370$ and the standard error $\sqrt{\hat{MSE}(\hat{\overline{Y}}_1)}=0.5805$. In the same way there are standard errors, $\sqrt{\hat{MSE}(\hat{\overline{Y}}_2)}=0.2460$, for small area 2, $\sqrt{\hat{MSE}(\hat{\overline{Y}}_3)}=0.4009$ for small area 3 and $\sqrt{\hat{MSE}(\hat{\overline{Y}}_4)}=0.5419$ for small area 4. The standard error of the total estimators are $\sqrt{\hat{MSE}(\hat{\hat{Y}}_1)}=12\times0.5805=6.9660$; $\sqrt{\hat{MSE}(\hat{\hat{Y}}_2)}=71\times0.246=17.7216$; $\sqrt{\hat{MSE}(\hat{\hat{Y}}_3)}=131\times0.4009=5.5179$; $\sqrt{\hat{MSE}(\hat{\hat{Y}}_4)}=14\times0.5419=7.5866$ for the small areas 1, 2, 3, and 4, respectively.