CHAPTER 7

USE OF ADDITIONAL INFORMATION

JUHA LAPPI AND ANNIKA KANGAS

Finnish Forest Research Institute, University of Helsinki, Finland

7.1 CALIBRATION ESTIMATION

If there are not enough sample plots to give sufficiently good inventory results using only forest measurements, we may try to make use of auxiliary variables correlated with forest variables. The most obvious way is to use ratio or regression estimators (Section 2.7). The calibration estimator of Deville and Särndal (1992) is an extension of the regression estimator for obtaining population totals using auxiliary information. Both regression and calibration estimators can be employed if there are auxiliary variables for inventory sample plots known for which the population totals are also known, e.g. variables obtained from remote sensing or from GIS systems. The appeal of calibration estimators for forest inventories comes from the fact that they lead to estimators which are weighted sums of the sample plot variables, where the weight can be interpreted as the area of forest in the population that is similar to the sample plot.

The basic features of the calibration estimator of Deville and Särndal (1992) in terms of estimating means can be described as follows. Consider a finite population U consisting of N units. Let j denote a general unit, thus $U = \{1, \ldots, j, \ldots, N\}$. In a forest inventory the population is a region where units are pixels or potential sample plots. The units in a forest inventory will be referred to here as 'pixels', and it will be assumed that an inventory sample plot gives values to the forest variables for an associated pixel. Each unit j is associated with a variable y_j and a vector of auxiliary variables $\mathbf{x}_j$. The population mean of $\mathbf{x}$, $\overline{\mathbf{X}} = N^{-1}\sum_U \mathbf{x}_j$ is assumed to be known. The y variables in a forest inventory are forest variables and the x variables can be spectral variables from remote sensing or geographical or climatic variables obtained from GIS databases.

A. Kangas and M. Maltamo (eds.), Forest Inventory – Methodology and Applications, 107–117.

Assume that a probability sample S is drawn, and y_j and $\mathbf{x}_j$ are observed for each j in S, the objective being to estimate the mean of y, $\overline{Y} = N^{-1}\sum_U y_j$. Let π_j be the inclusion probability and d_j the basic sampling design weight $d_j = (N\pi_j)^{-1}$, which can be used to compute the unbiased Horvitz-Thompson estimator

$$\hat{\overline{Y}}_d = \sum_s d_j y_j \ . \tag{7.1}$$

A calibration estimator

$$\hat{\overline{Y}} = \sum_s w_j y_j \tag{7.2}$$

is obtained by minimizing the sum of distances, $\sum_s G(w_j, d_j)$, between the prior weights d_j and posterior weights w_j for a positive distance function G, taking account of the calibration equation

$$\sum_s w_j \mathbf{x}_j = \overline{\mathbf{X}} \ . \tag{7.3}$$

If the distance between d_j and w_j is defined as

$$G_1(w_j, d_j) = (w_j - d_j)^2 / d_j \ , \tag{7.4}$$

the calibration estimator will be the same as the regression estimator

$$\hat{\overline{Y}}_r = \sum_s w_j y_j = \hat{\overline{Y}}_d + \left(\overline{\mathbf{X}} - \hat{\overline{\mathbf{X}}}_d\right)' \hat{\mathbf{b}} \ , \tag{7.5}$$

where $\hat{\overline{\mathbf{X}}}_d$ and $\hat{\mathbf{b}}$ (a weighted regression coefficient vector) are

$$\hat{\overline{\mathbf{X}}}_d = \sum_s d_j \mathbf{x}_j \quad \text{and} \tag{7.6}$$

$$\hat{\mathbf{b}} = \left(\sum_s d_j \mathbf{x}_j \mathbf{x}_j'\right)^{-1} \sum_s d_j \mathbf{x}_j y_j \ . \tag{7.7}$$

If the model contains an intercept, the corresponding variable x will be one for all observations, and the calibration equation (7.3) will then guarantee that the weights w_j add up to one. This means that when estimating totals, the

weights Nw_j will add up to the known total number of pixels in the population. Thus Nw_j can be interpreted as the total area, in pixel units, for plots of forest similar to plot j. The standard least squares theory implies that the regression estimator (7.5) can be expressed in the form

$$\hat{\bar{Y}}_r = \sum_s w_j y_j = \overline{\mathbf{X}}' \hat{\mathbf{b}} . \tag{7.8}$$

It is assumed that the intercept is always among the parameters.

Estimator (7.7) is defined if the moment matrix $\sum_s d_j \mathbf{x}_j \mathbf{x}_j'$ is non-singular. Some of the weights w_j in (7.2) implied by Eqs. (7.6)-(7.8) may be negative. Non-negative weights are guaranteed if the distance function is infinite for negative w_j. Deville and Särndal (1992) presented four distance functions producing positive weights.

Minimization of the sum $\sum_s G(w_j, d_j)$ so that (7.3) is satisfied is a non-linear constrained minimization problem. Using Lagrange multipliers, the problem can be reformulated as a non-linear system of equations which can be solved iteratively using Newton's method (for details, see Deville and Särndal 1992). If the initial values of the Lagrange multipliers are set to zero, the first step will produce w_j's of the regression estimator (7.5).

Since the calibration estimator is asymptotically equivalent to the regression estimator, Deville and Särndal (1992) suggest that the variance of the calibration estimator should be computed in the same way as the variance of the regression estimator using regression residuals. There is no design-unbiased estimator of the variance in systematic sampling (Schreuder et al. 1993).

The emphasis on area interpretation for the weights has the same argument behind it as was used by Moeur and Stage (1995) for the most similar neighbour method (MSN), where unknown plot variables are taken from a plot which is as similar as possible with respect to the known plot variables. In both methods each sample plot represents a percentage of the total area, and all the forest variables are logically related to each other. The difference is that in the calibration estimator we obtain an estimate of the area of the sample plot for the whole population whereas in the MSN method each pixel is associated with a sample plot. Since there is no straightforward way of showing that the MSN method produces optimal results in any way at the population level, it may be safer to use the calibration estimator for computing population-level estimates for forest variables. The problem with the calibration estimator is that it does not provide a map. If a map is needed, then the weights provided by the calibration estimator need to be distributed over pixels using separate after-processing.

Lappi (2001) proposed a 'small-area' modification of the calibration estimator which can be used when several subpopulation totals are required simultaneously. He used satellite data as auxiliary information for computing

inventory results for counties. Sample plots in the surrounding inclusion zone are also used for a given subpopulation so that the prior weight decreases as distance increases. The error variance is computed using a spatial variogram model. Block kriging (Cressie 1986) provides an optimal estimator for subpopulation totals under such a model, but kriging can produce negative weights for sample plots, and the weights are different for each *y* variable. Thus it is not possible to give areal interpretations to sample plot weights in kriging.

7.2 SMALL AREA ESTIMATES

Small area estimation is needed when estimates are required for subdivisions or domains of the population. Although the estimates for the whole population may be quite reliable, only a few sample units may fall into a given domain *i*, whereupon the classical design-based estimators may have unacceptably large errors. Accurate estimates for all small areas usually require overall sample sizes that are much too large to be within normal budget constraints (Särndal and Hidiroglou 1989). Thus, in order to improve the estimates of the domains, information from nearby areas can be used.

Small area estimators are typically at least partially model-based (Schreuder et al. 1993) and are referred to as synthetic or global estimators when information for the whole area is used instead of just the information from the domain i of interest (Särndal and Hidiroglou 1989). Estimators based only on information for the domain of interest are referred to as local estimators.

The classical local estimator for a domain *i* is

$$\hat{\bar{y}}_i = \sum_{j \in s_i} \frac{y_j}{n_i}, \tag{7.9}$$

where s_i denotes the sample drawn from domain *i* and n_i is the sample size in *i*. This estimator is unreliable for small sample sizes, however. The simplest possible model that can be used for small area estimation is

$$y_j = \mu + \varepsilon_j \quad for \quad j = 1,...,N . \tag{7.10}$$

Under model (7.10) the global estimator of the mean for domain *i* is thus

$$\hat{\bar{y}}_{iSYN} = \sum_{j \in s} \frac{y_j}{n}, \tag{7.11}$$

where *s* denotes the sample taken from the whole area and *n* is the total sample size. In fact, this is the sample mean for the whole population. The simplest global estimate is thus the overall sample mean for all domains *i*. As $n > n_i$, the estimates obtained with (7.11) will have a lower variance than the local estimates (7.9), but they will be badly biased unless the domain mean is the same as the population

mean, $\overline{Y}_i = \overline{Y}$, in all domains. With this model, the synthetic estimator (7.11) would differ from the domain mean even if all the units in domain *i* were measured, i.e. $n_i = N_i$.

A compromise between these two estimators is to combine the estimators (7.9) and (7.11). Under model (7.10), the best linear unbiased estimator for the domain mean $\overline{Y}_i$ is (Schreuder et al. 1993, p. 318)

$$\hat{\bar{y}}_{iCOM} = \frac{n_i}{N_i}\bar{y}_i + \left(1 - \frac{n_i}{N_i}\right)\bar{y} \tag{7.12}$$

If all the units in domain *i* were measured, the domain mean would have the weight 1 in this case and population mean 0, giving the correct estimate.

If additional information is available, it is possible to use a model (Ericksen 1973,1974, Mandallaz 1991, see sections 3.2, and 2.7)

$$y_j = \mathbf{x}_j\boldsymbol{\beta} + \varepsilon_j \quad for \quad j = 1,\ldots, \tag{7.13}$$

where $\mathbf{x_j}$ is a ($p+1$) vector of independent variables at point (plot) *j*. The coefficients $\boldsymbol{\beta}$ are estimated for the whole population and global estimates for domain *i* are obtained by

$$\hat{\bar{y}}_{iREG} = \overline{\mathbf{X}}_i\hat{\boldsymbol{\beta}}, \tag{7.14}$$

where $\overline{\mathbf{X}}_i$ contains the true average values for the independent variables of domain *i*. The estimator of its variance (assuming infinite population or analytic inference) is

$$Var(\hat{\bar{y}}_{iREG}) = \hat{\sigma}^2\overline{\mathbf{X}}_i(\mathbf{X'X})^{-1}\overline{\mathbf{X}}_i, \tag{7.15}$$

where $\mathbf{X}$ is the $n \times (p+1)$ matrix containing values for the independent variables for each sample point and $\hat{\sigma}^2$ is the estimator for the model residual variance (Eq. 3.7). If only the intercept of model (7.13) is significant, this model reduces to (7.10). The estimator (7.14) is almost the same as the estimator (3.10) presented in section 3.2. The only difference is that in (7.14) the model coefficients are estimated for the whole population whereas $\overline{\mathbf{X}}_i$ is for domain *i*.

Synthetic methods of estimation assume that small areas have characteristics similar to those of the larger areas of which they are part (Gonzales 1973). If this assumption is unjustified, the synthetic estimators will be biased. If the bias component does not tend towards zero as the sample size increases, the estimator is design-biased (Särndal 1984). On the other hand, if an estimator is biased under the assumed model it can be said to be model-biased. A biased

estimator may still be useful if its MSE is smaller than that of an unbiased estimator and if the presence of bias is acceptable.

This bias in synthetic estimators can be reduced by combining an unbiased estimator with a design-biased but low variance estimator, for example, so that the weight of the unbiased estimator increases as the sample size in the small domain increases. Such attempts have included the use of shrinkage or empirical Bayes estimators (Green et al. 1987, see also Hulting and Harville 1991).

It is also possible to correct the estimates obtained with global models by using residuals observed in domain *i* (Särndal 1984, Särndal and Hidiroglou 1989). Mandallaz (1991) proposed a global estimator

$$\hat{\bar{y}}_{iSUR} = \bar{\mathbf{X}}_i\hat{\boldsymbol{\beta}} + (\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}), \tag{7.16}$$

where $\bar{\mathbf{x}}_i$ is the vector of sample means and $\bar{\mathbf{X}}_i$ is the vector of true means in a small area *i*. In (7.16) the synthetic model-based estimator (7.14) is corrected for the bias by means of the residuals observed in the small area *i*.

The estimator of its variance is (Mandallaz 1991)

$$Var(\hat{\bar{y}}_{iSUR}) = \frac{1}{n_i(n_i - 1)}\sum_{j\in s_i}(r_{ji} - \bar{r}_i)^2, \tag{7.17}$$

where r_{ji} is the observed residual in domain *i* and plot *j*.

An alternative model for domain estimation would be

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + c_i + \varepsilon_{ij} \quad for \quad j = 1,... \; and \; i = 1,...,k, \tag{7.18}$$

where $c_i \sim N(0,\sigma_w^2)$ is a random domain effect, $\varepsilon_{ij} \sim N(0,\sigma_e^2)$ is a random plot effect and *c* and ε are mutually independent (Battese et al. 1988). The difference relative to model (7.13) is that the residual error term in (7.18) is divided into two components. The domain effect describes the difference of domain *i* from the population mean, which makes it useful for estimating the domain mean. The global estimator for the domain mean is then (Prasad and Rao 1990)

$$\hat{\bar{y}}_{iMIX} = \bar{\mathbf{X}}_i\hat{\boldsymbol{\beta}} + \hat{c}_i, \tag{7.19}$$

where the domain effect $\hat{c}_i$ can be estimated by

$$\hat{c}_i = \frac{\sigma_w^2}{\sigma_w^2 + \frac{\sigma_e^2}{n_i}}(\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}) = \gamma_i(\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}) \tag{7.20}$$

and γ_i is the (constant) correlation within domain *i*, calculated from the variances in the domain and plot effects and the number of plots.

The estimator of $\hat{c}_i$ (7.20) is biased for a given c_i, but unbiased over the distribution of domains (Lappi 1993). Thus the estimator (7.19) is also model-biased for a given domain but unbiased over the distribution of domains. The larger the within-domain correlation, and the larger the difference $(\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}})$, the larger the predicted $\hat{c}_i$ in (7.19) is. As the variance σ_e^2 approaches infinity, the correlation approaches one and estimator (7.19) approaches estimator (7.16). This means that the global estimator of the mean (7.19) is corrected by means of the observed residuals, as in (7.16), but the amount of this correction depends on the correlation within the domains. The mean square error of (7.19) can be calculated using the theory of linear models, details of which can be found in Prasad and Rao (1990).

If only the intercept of the fixed part of model (7.18) is significant, the estimator (7.19) reduces to a linear combination of the estimate for the total area mean and the observed mean in domain *i*:

$$\hat{\bar{y}}_{iMIX} = \left(1 - \frac{\sigma_w^2}{\sigma_w^2 + \frac{\sigma_e^2}{n}}\right)\hat{\mu} + \left(\frac{\sigma_w^2}{\sigma_w^2 + \frac{\sigma_e^2}{n}}\right)\bar{y}_i \qquad (7.21)$$

This estimator is quite similar to simple James-Stein estimators or the combined estimator (7.13) (Schreuder et al. 1993). Treating the domain effect as random provides a means of combining the domain mean efficiently with the estimator of the population mean.

Geostatistical methods provide interesting possibilities in small area estimation in a forestry context, since in most cases the auxiliary information includes coordinate locations. With these methods it is possible to take the autocorrelations present in the data explicitly into account, instead of just constant within-domain correlation as in the mixed model. In kriging methods, the autocorrelation between the sample plots is usually assumed to depend purely on the distance between the sample plots and to decrease with increasing distance. In a mixed model, however, this correlation is approximated by means of an average correlation over a predefined area. Thus the mixed model approach can be considered a special case of kriging. The kriging method has been presented by Journel and Huijbregts (1978), Burgess and Webster (1980a, 1980b), Ripley (1981) and Cressie (1986), for example, and for small area estimation by Mandallaz (1993) (see also Chapter 10). Examples of small area estimation in forestry are provided by Green et al. (1987), Mandallaz (1991), Kangas (1996) and Lappi (2001), for example.

Example 7.1

The example is based on simulated data. Assume a 1000 hectare area with five distinct regions of interest. The volume of each region is surveyed. There is a satellite image available, and the near-infrared (NIR) channel is used as auxiliary information.

The true data were obtained assuming that the NIR was a normally distributed variable with mean 0.2482 and standard deviation 0.0364. A dataset of 1000 observations for the regions was generated, and the true volumes were obtained from a model

$$V_i = 322.7473 - 714.951\,\mathrm{NIR}_i + \varepsilon_i,$$

where the standard deviation of ε_i was 38.66 m^3/ha. The true mean values for NIR and volume in each area, calculated from these data, are presented in Table 7.1.

Table 7.1 True values for volume and NIR.

District	size, ha	NIR	Volume m^3/ha	STD
1	94	0.22893	155.5	43.82
2	69	0.25104	140.7	40.43
3	123	0.26008	139.2	42.34
4	537	0.28201	120.4	45.40
5	177	0.31497	92.5	44.35
sum/mean	1000	0.27802	122.5	47.84

A sample of 50 plots was taken from the area at random. The values of NIR and volume for each sample plot are presented in Table 7.2.

Table 7.2 The sample.

District	NIR	VOL	District	NIR	VOL
1	0.176512	212.6	4	0.264825	182.3
1	0.212170	154.4	4	0.322347	70.3
1	0.234031	170.8	4	0.313223	130.6
1	0.196743	139.9	4	0.326355	95.3
1	0.261204	159.6	4	0.264030	116.0
1	0.235359	123.9	4	0.308574	93.6
1	0.191436	222.9	4	0.313137	12.8
2	0.244882	133.2	4	0.240222	142.5
2	0.281133	70.8	4	0.281222	75.0
2	0.252457	2.6	4	0.281231	127.1
2	0.268814	119.0	4	0.330613	132.3
2	0.268588	136.6	4	0.313529	114.9
3	0.237107	169.6	4	0.261752	114.9
3	0.253262	115.8	4	0.270598	115.1
3	0.242354	141.9	4	0.327834	57.7
3	0.268941	98.7	5	0.265201	141.3
3	0.301190	163.5	5	0.319447	81.6
3	0.273860	35.5	5	0.321587	120.4
4	0.291545	116.2	5	0.272238	115.9
4	0.259637	98.5	5	0.277245	142.2
4	0.277605	110.7	5	0.309464	82.6
4	0.239459	182.7	5	0.309751	89.2
4	0.339731	56.0	5	0.275660	67.6
4	0.226967	148.9	5	0.366935	28.3
4	0.229883	202.5	5	0.326796	44.8

A linear regression model, having the characteristics presented in Table 7.3, was estimated from the sample data.

Table 7.3 Model statistics.

Regression Statistics	
R^2	0.428969
Adjusted R^2	0.417073
Standard Error	37.28421
Observations	50

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	50125.37	50125.37	36.0585	2.46E-07
Residual	48	66725.41	1390.113		
Total	49	116850.8			

	Coefficients	*Std Error*	*t Stat*	*p-value*
Intercept	330.9492	36.24291	9.13142	<0.000
NIR	−784.192	130.5926	−6.00487	<0.000

Estimates for the small area obtained with various formulae are presented in Table 7.4.

Table 7.4 Estimates for the small area.

District	n	NIR	$\hat{\bar{y}}_i$ (7.9)	$s_e(\hat{\bar{y}}_i)$ (2.12)	$\hat{\bar{y}}_{iSYN}$ (7.11)	$\hat{\bar{y}}_{iREG}$ (7.14)	$\hat{\bar{y}}_{iSUR}$ (7.16)	$s_e(\hat{\bar{y}}_{iSUR})$ (7.17)
1	7	0.22893	169.2	13.79	115.6	151.4	158.5	11.387
2	5	0.25104	92.5	25.35	115.6	134.1	102.0	31.773
3	6	0.26008	120.8	20.36	115.6	127.0	123.0	20.389
4	22	0.28201	113.5	9.45	115.6	109.8	116.3	6.946
5	10	0.31497	91.4	12.21	115.6	84.0	83.1	8.346
total	50	0.27802	115.6	6.91	115.6	112.9	112.9	5.216

REFERENCES

Battese G.E., Harter, R.M. and Fuller, W.A. 1988. An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. Journal of American Statistical Association 83:28-36.

Burgess, T.M. and Webster, R. 1980. Optimal Interpolation and Isarithmic Mapping of Soil Properties. I The Semi-variogram and Punctual Kriging. Journal of Soil Science 31:315-331.

Burgess, T.M. 1980. Optimal Interpolation and Isarithmic Mapping of Soil Properties. II Block Kriging. Journal of Soil Science 31:333-341.

Cressie, N. 1986. Kriging Nonstationary Data. Journal of American Statistical Association 81:625-634.

Deville, J.C. and Särndal, C.E. 1992. Calibration estimators in survey sampling. Journal of the American Statistical Association 87: 376-382.

Ericksen, E.P. 1973. A method for combining sample survey data and symptomatic indicators to obtain estimates for local areas. Demography 10:137-160.

Ericksen, E.P. 1974. A regression method for estimating population changes of local areas. Journal of American Statistical Association 69:867-875.

Gonzales, M.E. 1973. Use and Evaluation of Synthetic Estimates. Proceeding of Social Statistics Section, American Statistical Association. 73:7-15.

Green, E.J., Thomas, C.E. and Strawderman, W.E. 1987. Stein-Rule Estimation of Timber Removals by County. Forest Science 33:1054-1061.

Hulting, F.L. and Harville, D.A. 1991. Some Bayesian and Non-Bayesian Procedures for the Analysis of Comparative Experiments and for Small-Area Estimation: Computational Aspects, Frequentist Properties, and Relationships. Journal of American Statistical Association 86:557-568.

Journel, A.G. and Huijbregts, C.J. 1978. Mining Geostatistics. London, Academic Press.

Kangas, A. 1996. Small area estimates using model based methods. Canadian Journal of Forest Research 26:758-766.

Lappi, J. 1993. Metsäbiometrian menetelmiä. Silva Carelica 24. 182 p.

Lappi, J. 2001. Forest Inventory of Small Areas Combining the Calibration Estimator and a Spatial Model. Canadian Journal of Forest Research 31:1551-1560.

Mandallaz, D. 1991. A Unified Approach to Sampling Theory for Forest Inventory Based on Infinite Population and Superpopulation Models. Chair of Forest Inventory and Planning, Swiss Federal Institute of Technology (ETH), Zurich. 242 p.

Mandallaz, D. 1993. Geostatistical Methods for Double Sampling Schemes: Application to Combined Forest Inventories. Chair of Forest Inventory and Planning, Swiss Federal Institute of Technology (ETH), Zurich. 133 p.

Moeur, M. and Stage, A.R. 1995. Most Similar Neighbor: An Improved Sampling Inference Procedure for Natural Resource Planning. Forest Science 41:337-359.

Prasad, N.G.N. and Rao, J.N.K. 1990. The Estimation of the Mean Square Error of Small Area Estimators. Journal of American Statistical Association 85:163-171.

Ripley, B.D. 1981. Spatial Statistics. John Wiley and Sons. New York. 252 p.

Särndal, C-E. 1984. Design-Consistent versus Model-Dependent Estimation for Small Domains. Journal of American Statistical Association 79:624-631.

Särndal, C-E. and Hidiroglou M.A. 1989. Small Domain Estimation: A Conditional Analysis. Journal of American Statistical Association 84:266-275.

Schreuder, H.T., Gregoire, T.G. and Wood, G.B. 1993. Sampling Methods for Multiresource Forest Inventory. John Wiley and Sons. New York. 446 p.