Constrained map-based inventory estimation

PAUL C. VAN DEUSEN[1]* AND FRANCIS A. ROESCH[2]

[1] National Council for Air and Stream Improvement, Inc. 600 Suffolk St, Lowell, MA 01854, USA
[2] USDA Forest Service, Southern Research Station, 200 WT Weaver Boulevard, Asheville, NC 28804-3454, USA
*Corresponding author. E-mail: PVandeusen@ncasi.org

Summary

A region can conceptually be tessellated into polygons at different scales or resolutions. Likewise, samples can be taken from the region to determine the value of a polygon variable for each scale. Sampled polygons can be used to estimate values for other polygons at the same scale. However, estimates should be compatible across the different scales. Estimates are often required for zones within a region, where a region might be a state and counties could be the zones. A method is developed for estimating high-resolution (pixel) values that are constrained to be compatible with results for lower resolution values. The high-resolution values are constrained to sum to totals for zones within a region, where the totals are being simultaneously estimated from measurements taken at a different scale. If the zone estimates are unbiased, then the pixel-based estimates for the zone will be less biased. Sums of pixels in arbitrary polygons are thereby constrained to approach unbiased estimates. Approximate variance estimators are developed for the summed pixel estimates. Two example applications are provided. The first example is based on simulated data and verifies that the proposed variance estimators give reasonable results. The second example estimates the volume in a circle around a possible mill site in North Carolina. This example uses publicly available US Forest Service inventory data and simulated inventory data that the mill would provide.

Introduction

Thematic maps are relatively easy to make and provide useful visual impressions about the distribution of various categories of land use or natural resources. Estimates derived from these maps can be adjusted to have known variance and bias properties (Card, 1982; Story and Congalton, 1986; Green *et al.*, 1993; Van Deusen, 1994, 1996).

The approach of making thematic maps and adjusting derived summary statistics has been thoroughly researched. Less well studied is the idea of adjusting pixel values using measurements taken at a different scale so that map-based estimates of means and totals are unbiased. This would allow the user to draw a polygon on the map and obtain reliable estimates by summing the pixels contained in the polygon. Methods are developed here to adjust the pixel estimates to conform to larger zones whose totals are based on sample data taken at a different scale. The methods developed here are for quantitative pixel values rather than categorical values. These methods would therefore be useful for maps that depict tree volume but not for tree species.

This generalizes work in Roesch and Van Deusen (1995) where pixel-based estimates were corrected according to externally derived summary statistics. In that study, large 1-km pixel values and USDA Forest Service inventory and analysis (FIA; US Department of Agriculture, Forest Service, 2005) data

© Institute of Chartered Foresters, 2007. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

Forestry, Vol. 80, No. 4, 2007. doi:10.1093/forestry/cpm021
Advance Access publication date 19 July 2007

are used to improve estimates based on 30-m pixels. We extend this idea to multiple scales or levels. The methods could also be applied to inventories where not all levels involve remotely sensed data.

The next section outlines an example that is designed to motivate the need for the methods being developed here. This is followed by a section on notation and a method for stochastic adjustment of pixel values to correspond to estimated zone totals obtained at a different scale. This ensures that the sum of all pixels in a zone will be compatible with other estimates of zone totals. A section on variance estimation outlines procedures for estimating the variance of sums of pixels in arbitrary polygons. The method is tested with an application based on simulated data. This demonstrates that the proposed variance estimators and constraint methods perform as expected. A second example addresses the practical issue of combining company inventory data with FIA data to obtain estimates in a procurement circle around a mill. This is discussed in more detail in the next section.

Motivating example

An important justification for these methods is the need to estimate the volume of wood within a certain distance of a paper mill, i.e. a mill circle. FIA data are available for the states that contain the mill circle, but the plot locations are not available due to privacy issues. Suppose that another set of inventory or photo plots is available to the mill with exact locations so the plot values can be modelled from remote sensing spectral band values. Alternatively, a simple model could be developed that relates the plot values to x and y coordinates.

One method to estimate the total volume in a circle or polygon would be to classify 30-m pixels as forest or non-forest for the circle. Then, the forested pixels whose centres are contained in the circle can be counted and the forested area in the circle is readily computed. Now, multiply the circle's forest area estimate by the mean volume estimate from plots in the circle to get the total volume estimate. This could be accomplished with National Land Cover Data (NLCD) (Vogelmann *et al.*, 2001) to compute forest area and FIA plots to compute mean volume or biomass in the circle.

The mill has its own plots, which could also be used to estimate mean volume in the circle. However, these plots might not have been measured recently or may be based on photo interpretation so the mill feels uncomfortable using them as the only basis for the estimate. The procedures developed here provide a method to combine the FIA and the mill plots to obtain estimates that may be more reliable than using either data source alone.

A mixed estimator is developed (Theil, 1971) that will constrain the equation fit to the mill data to give means that are similar to the FIA means in specified zones. In the second example application, these zones are the four quadrants of the mill circle. The constraint reflects the notion that FIA data provide an unbiased and accepted estimate for a state. Therefore, volume estimates for the mill circle should be expected to conform to FIA estimates. Our method is applied to FIA and NLCD data for central North Carolina to demonstrate its efficacy in the example application section.

Notation and equations

The development here focuses on the common situation where there are two scales, but the results generalize easily to multiple scales. The basic data come in the form of four vectors for each level or scale: $L^a = (X^a, Y^a, M^a, Z^a)$ and $L^b = (X^b, Y^b, M^b, Z^b)$, where the a and b superscripts denote the level or scale. X^b denotes an $N_b \times p_b$ matrix of concomitant variables that should be available for all N_b units at the b scale. Y^b denotes the vector that contains the variable of interest for the study and M^b is a measurement indicator vector that contains 0s when the y-measurement is not available and 1s otherwise. The number of observations is n_b, which is the number of 1s in M^b. There are K_b zones labelled $1,\ldots,K_b$ and the vector Z^b contains labels to indicate the zone membership of each unit at level b. The definitions of the four vectors at level a correspond to the definitions for level b. Another level could be added by defining L^c analogously to L^a and L^b. Throughout, we denote a scalar with a subscript, e.g. N_b, and a matrix or vector with a superscript, e.g. X^b. Vectors or matrices that contain the complete data for a level are in upper case, e.g. Y^b, whereas lower case is used for the observed data, e.g. y^b.

Observation equations

The measurement indicator allows us to create vectors that correspond to observed values $L^a_{\text{obs}} = (x^a, y^a)$, where each observed vector has n_a rows. The observed values are used to estimate unknown parameters, but the full X-matrices are used to enforce constraints. Assume that the following equations describe the observed data:

$$y^a = x^a \beta^a + e^a \tag{1}$$

and

$$y^b = x^b \beta^b + e^b, \tag{2}$$

where β^a and β^b are vectors of unknown coefficients. The error vectors, e^a and e^b, have mean zero and are assumed to have a joint variance–covariance matrix of the form

$$\Sigma = \begin{bmatrix} \Sigma^a & \Sigma^{ab} \\ \Sigma^{ab'} & \Sigma^b \end{bmatrix}. \tag{3}$$

In most cases, $\Sigma^a = \sigma^2_a I^a$ with Σ^b having a similar form, i.e. the elements within e^a and e^b are independent. Likewise, e^a and e^b will often be independent and thus, $\Sigma^{ab} = 0$.

Constraints

The constraints that cause summations across different levels to be compatible are imposed using mixed estimation methods (Theil, 1971). The constraints are put in the form of equations that are solved simultaneously with the observation equations, equations (1) and (2). One vector constraint equation is required for a two-level problem

$$r = J^a X^a \beta^a c^a - J^b X^b \beta^b c^b + v, \tag{4}$$

where X represents the full X-matrix (not just observed rows). The $Q_b \times 1$ error vector, v, has one element for each constraint and the variance–covariance matrix $E(vv') = \sigma^2_v I^{Q_b}$. The error vector, v, is uncorrelated with the observation equation errors. The r-vector allows for the possibility that the zone sums are not equal at the different levels, but usually $r = 0$. The J-matrices have Q_b rows and enough columns to be conformable with the matching X-matrix. The qth row of J has 1s to pick off the rows of X that are involved in the constraint, and has 0s elsewhere. Therefore, $J^a X^a \beta^a$ gives the sums of all a-level elements involved in the constraint. Conversion factor vectors, c^a and c^b, are needed to put the constraint sums at the different levels into compatible units. It makes sense to have that all prediction equations give results in the same units, e.g. volume per hectare. However, level a might consist of 30-m pixels with 1-km pixels at level b. Summing per hectare volume predictions for the different size pixels will not result in correct zone totals without the conversion factors. A simple conversion when the constraint includes all observations is to let $c^a = 1/n_a$ and $c^b = 1/n_b$, so the means at each level are constrained.

For some problems, the zones in level b do not completely tessellate the region or include areas that should be ignored. This would leave a number of level-a elements without a zone b membership. This implies that the membership vector, Z^a, can include missing values. The summations occurring in equation (4) must exclude these elements at level a by putting a 0 in the corresponding position in J^a. The second example demonstrates a typical situation where this occurs. If only forested pixels are of interest, then the non-forest pixels must be excluded from the computations even though they are physically within the zones.

The number of constraints is limited by the number of parameters being estimated. In general, $Q_b \leq p_a + p_b$. There cannot be more constraints than the number of unknown parameters in β^a and β^b.

Estimation

Development of estimators is facilitated by creating more compact matrix notation for the observation and constraint equations given above. The first step is to write the combined equations as

$$\begin{bmatrix} y^a \\ y^b \\ r \end{bmatrix} = \begin{bmatrix} x^a & \mathbf{0} \\ \mathbf{0} & x^b \\ & R \end{bmatrix} \begin{bmatrix} \beta^a \\ \beta^b \end{bmatrix} + \begin{bmatrix} e^a \\ e^b \\ v \end{bmatrix}, \tag{5}$$

where $\mathbf{0}$ is a matrix of 0s and

$$R = \begin{bmatrix} J^a X^a c^a & -J^b X^b c^b \end{bmatrix}. \tag{6}$$

Carry the compaction into matrix notation one step further to get

$$\begin{bmatrix} Y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} e \\ v \end{bmatrix}. \qquad (7)$$

Now, the estimator for β can be written in standard mixed estimator form (Theil, 1971) as

$$\hat{\beta} = \left[X'\Sigma^{-1}X + \frac{1}{\sigma_v^2} R'R \right]^{-1} \left[X'\Sigma^{-1}Y + \frac{1}{\sigma_v^2} R'r \right]. \qquad (8)$$

The approximate covariance matrix of $\hat{\beta}$ is

$$\mathrm{Var}(\hat{\beta}) \approx \left[X'\Sigma^{-1}X + \frac{1}{\sigma_v^2} R'R \right]^{-1}. \qquad (9)$$

Variance component estimation

For most applications of these methods, there will only be a few unknown variance components to estimate. The observation equation components are σ_a^2 and σ_b^2. They can be estimated iteratively using

$$\hat{\sigma}_a^2 = \frac{(y^a - x^a\hat{\beta}_u^a)'(y^a - x^a\hat{\beta}_u^a)}{n_a - p_a} \qquad (10)$$

and

$$\hat{\sigma}_b^2 = \frac{(y^b - x^b\hat{\beta}_u^b)'(y^b - x^b\hat{\beta}_u^b)}{n_b - p_b}, \qquad (11)$$

where $\hat{\beta}_u$ denotes an unconstrained estimate.

The first iteration requires a starting value for σ_a^2 and σ_b^2, e.g. set them both to 1.0. Then, estimate $\hat{\beta}_u$ with equation (8), where $\sigma_v^2 = \infty$, i.e. set σ_v^2 to a very large number, say 10 000. Then estimate σ_a^2 and σ_b^2 with equations (10) and (11). Cycle through this process until convergence. The unconstrained estimate of β ensures that the estimates of σ_a^2 and σ_b^2 are unbiased. Then get the constrained estimate of β in a second step using the unconstrained estimates of σ_a^2 and σ_b^2.

The constraint equation variance component, σ_v^2, is somewhat more difficult to estimate. One approach would be to adjust σ_v^2 until the constraint is met to within a user-specified tolerance. For example, a typical constraint would be to ensure the scaled mean over all level-*a* units is nearly the same as the mean of all level-*b* units in the same zone. The value of σ_v^2 can be adjusted until these means are close enough to satisfy the user, i.e. decrease σ_v^2 if the difference in the means is too large. This puts more weight on the constraint.

A more objective approach is to find the value, $\tilde{\sigma}_v^2$, that minimizes the log-likelihood function

$$L \propto \log(|\Sigma|) + (Y - X\beta)'\Sigma^{-1}(Y - X\beta) + \frac{1}{\sigma_v^2}(r - R\beta)'(r - R\beta) + Q_b \log(\sigma_v^2). \qquad (12)$$

The minimization of L could be done with a simple grid search. This can be included in the iterative process of estimating β, σ_a^2 and σ_b^2 as an additional step. An initial value for σ_v^2 is required to start the iterations.

The likelihood approach will often be appropriate for finding an objective value of $\tilde{\sigma}_v^2$. However, the overall purpose of the mixed estimator is to control the bias in estimates across different zones. The method of reducing σ_v^2 until the zone totals are sufficiently similar across levels can also be employed with the recognition that it yields a subjective result.

Variance of zone sums

A method is needed to estimate the variance of the sum of high-resolution elements contained within an arbitrary polygon drawn on the map. These elements might be 30-m pixels from satellite data or hexagons that tessellate the region. Begin by considering the estimated sum of level-*a* elements over Q_b zones

$$\hat{S}^a = J^a X^a \hat{\beta}^a, \qquad (13)$$

where J^a would now contain Q_b rows, one for each zone of interest. The qth row of J^a would have 1s to pick off the level-*a* elements in zone q and 0s elsewhere. The variance of $\hat{S}^a$ can be estimated by

$$\widehat{\mathrm{Var}}(\hat{S}^a) = J^a X^a \mathrm{Var}(\hat{\beta}^a) X^{a'} J^{a'}, \qquad (14)$$

where $\mathrm{Var}(\hat{\beta}^a)$ is the $p_a \times p_a$ covariance matrix from the upper left partition of $\mathrm{Var}(\hat{\beta})$ in equation (9).

The variance estimate in equation (14) gives the variance of the expected zone sums. It would be more realistic to treat the sums in equation (13) as predictions since nearly all the level-a values are being estimated. The variance of predicted sums (Theil, 1971) is

$$\widehat{V}ar(\hat{S}^a) = J^a X^a \left[\mathrm{Var}(\hat{\beta}^a) + I_{p_a} \right] X^{a'} J^{a'}. \quad (15)$$

Applications

The methods for multilevel constrained estimation apply to a range of problems. The first example applies the methods to an entirely simulated dataset that is meant to emulate a two-level problem, where level a is based on satellite data and level b uses FIA data. There is currently much interest in combining satellite imagery with FIA data for improved estimation (McRoberts *et al.*, 2002). The primary objective of the first example is to demonstrate that the method can recover the known parameters that were used to generate the data.

The second example uses NLCD data and FIA data to demonstrate a simple application of these methods to a more realistic dataset and problem. The objective is to estimate the total dry biomass in a circle (81 km radius) in central North Carolina.

Example 1

This exercise begins by generating N_a = 10 000 members of the level-a population using the following functions:

$$X^a = x_{min} + 5(Z^a - 1) + e^{X^a},$$
$$Y^a = \beta_a X^a + e^{Y^a} + bias_a,$$

where $x_{\min}$ = 100, $e^{X^a} \sim N(0,0.04x_{min})$, β_a = 3.0 and $e^{Y^a} \sim N(0,0.2x_{min})$. The bias$_a$ term is added to ensure that the level-a observations are biased relative to the level-b observations. This bias causes zone totals across the two levels to be unequal without a constraint. Both levels have observations in the same 30 zones labelled 1,...,30 in the zone label vectors, Z^a and Z^b. Each zone gets an equal number of the observations at each level. A sample of size n_a = 20 is drawn from the 10 000-member level-a population.

There are N_b = 300 level-b observations (n_b = 10) generated as

$$X^b = \left[x_{min} + 5(Z^b - 1) + e^{X^b} \right] / 2,$$
$$Y^b = \beta_b X^b + e^{Y^b},$$

where β_b = $2\beta_a$, $e^{X^b} \sim N(0,0.1x_{min})$ and $e^{Y^b} \sim N(0,0.2x_{min})$. The vector X_b has the same mean as X_a except that it is divided by 2. Y_b is generated from the same equation as Y_a, except $\beta_b = 2\beta_a$ to compensate for X_b being divided by 2. Now, the objective of the simulation is to test the ability of the method to constrain the zone sums across levels, to rediscover β_a and β_b and to deal with the bias in level a. All level-b observations are used in the analysis.

A constraint is employed to make the level-a mean estimate be equal to the level-b mean. This constraint enforces the following relationship:

$$v = \beta_{1b}\bar{X}^b - \beta_{0a} - \beta_{1a}\bar{X}^a, \quad (16)$$

where β_{1b} is the level-b slope parameter, $\bar{X}^b$ is the mean level-b x-value, β_{0a} is the level-a intercept parameter, β_{1a} is the level-a slope parameter and $\bar{X}^a$ is the mean level-a x-value. The constraint variance is controlled by the value of σ_v, which is initially set to 10 000 so the constraint has almost no effect on the results.

Results for Example 1

First, we set bias$_a$ = 0 and run the simulation to verify that the correct values for β_a and β_b are recovered. The model for the level-a observations has an intercept and a slope, so the intercept can account for the bias term. The level-b observations use a model with a slope and no intercept. The simulation results (Table 1) for 1000 iterations indicate that the slope parameters were accurately estimated and that the variance estimate from equation (9) is similar to the simulation variance for the estimated parameters. The mean of both Y_a and Y_b is ~517, which makes the level-a intercept relatively close to zero, as it should be with no added bias.

Repeating the simulation with bias$_a$ = 20 gives insight (Table 2) into the value of constraints to control bias. This run is first made without constraints, i.e. σ_v = 10 000. The results show that the level-a intercept reflects the bias term and the variance of parameter estimates from equation (9)

Table 3: Results of two-level simulation application with $\text{bias}_a = 20$ and $\sigma_v = 1$

Parameter	$\hat{\beta}$	$\text{Var}(\hat{\beta})$	
		Equation (9)	Simulation
β_{0a}	5.97	311.98	324.46
β_{1a}	2.98	1.04×10^{-02}	1.07×10^{-02}
β_{1b}	6.01	1.56×10^{-04}	1.75×10^{-04}

The average is shown for $\hat{\beta}$ along with the estimate of $\text{Var}(\hat{\beta})$ from equation (9) and the simulation.

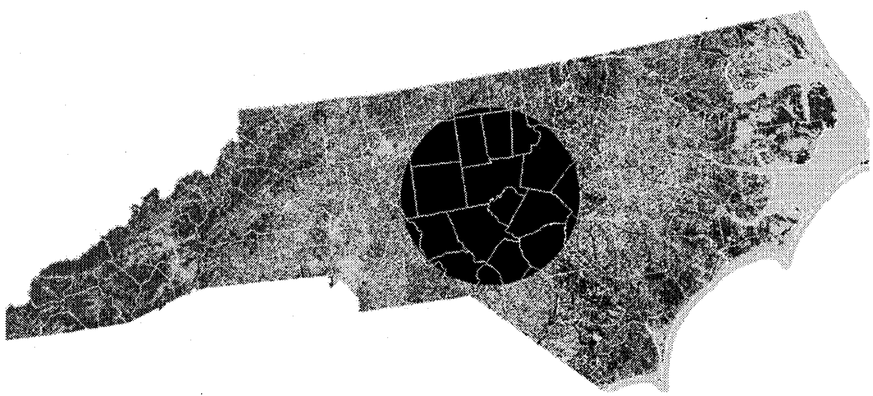

Figure 1. Central North Carolina mill circle (81 km radius). The circle is black, non-forest is light grey and forest is dark grey.

error vector and the unknown coefficients are a_j, $j = 0, 1, 2$. The low-resolution dataset is from the FIA plots in the bounding box. The FIA dry biomass values are simply modelled as a mean, i.e.

$$Y^b = \mu + e^{Y^b}. \tag{18}$$

There are four model coefficients being estimated, which allows for four constraints in the R-matrix. We decided to constrain the pixel means in each quadrant to be similar to the overall FIA mean biomass value. The quadrants are the north-east, north-west, south-east and south-west sections of the bounding box. This will ensure that the estimate within the mill circle will be compatible with the FIA data. The mill circle estimate will be derived in two steps. First, count the forested pixels whose centres fall inside the circle. Then, compute the forest area in the circle using the fact that each forested pixel represents 900 m^2 of forest. Second, compute the mean biomass estimate of all pixels in the circle. Multiply the estimates from steps 1 and 2 to get an estimate of total biomass in the circle.

Analysis details for Example 2

This analysis was done using GRASS (2005) for manipulating the NLCD data and the R-project software (R Development Core Team, 2007) for implementing the mixed estimator. All 29 million NLCD pixels are incorporated into the 4×4 constraint matrix (equation 6), which may be the most computationally demanding part of this exercise. Each row of the R-matrix corresponds to one of the quadrants of the mill circle bounding box. The non-forest pixels are excluded when the R-matrix is formed since they are not used in the computations and should have no biomass by definition.

The analysis is not computationally demanding, after the R-matrix (equation 6) is computed.

It involves the 2000 simulated mill plots, the 918 FIA plots and the 4 × 4 *R*-matrix. The data preparation can be demanding, but the application of the mixed estimation steps will be quite similar across problems.

The maximum likelihood method (equation 12) is employed to find the constraint variance parameter. It is appropriate for this example because it provides an objective result. It would not be clear how to set this parameter's value otherwise. There should be some similarity between the quadrant estimates and the FIA plot estimate for the bounding box, but they should not necessarily be identical. The maximum likelihood estimate provides a justifiable result.

Results for Example 2

The unconstrained results from fitting equation (17) indicate (Table 4) that there is a trend in the data in both the *X* and *Y* directions. The intercept is not significant probably due to the artificial nature of the pseudo-biomass variable.

The mixed estimation process was applied using the maximum likelihood estimate for the constraint variance parameter. This resulted in slightly modified coefficient estimates (Table 5).

The total (pseudo) biomass in the mill circle can now be calculated by computing the average biomass value for all forested pixels in the mill circle with equation (17) and the mixed estimator coefficients (Table 5). These coefficients result in estimates of total dry biomass in tons per acre (1 ton acre^{-1} = 2.2417 tonnes ha^{-1}). Depending on the units selected for the final result, the 30-m pixels each represent either 0.22239395 acres or 0.09 ha.

Conclusions

A method was developed for simultaneously estimating population parameters from data taken at different scales. An example application used simulated data where the high-resolution data could represent satellite-derived pixels and the low-resolution data might be from a forest inventory. This example indicated that these methods provide reliable variance estimates and help to ensure compatibility of estimates across scales. A second example demonstrated how these methods could be used to estimate the volume of wood in a mill circle by combining photo plots with FIA data. The methods developed here are not difficult to implement, but require the ability to perform some customized statistical analysis.

The motivation for this work was to develop a method for combining data measured at different scales or resolutions. However, this method could generally be used to combine data from different sources. For example, two inventories of the same area could be combined. For this purpose, the mixed estimator is similar in

Table 4: Unconstrained results from fitting the pseudo-biomass data to equation (17)

Coefficient	Estimate	Standard Error	*t*-value	Pr(>|*t*|)
Intercept	$-1.374 \times 10^{+01}$	$1.843 \times 10^{+01}$	−0.746	0.456
X	-9.054×10^{-05}	9.326×10^{-06}	−9.708	$<2 \times 10^{-16}$
Z	1.219×10^{-04}	8.819×10^{-06}	13.819	$<2e \times 10^{-16}$

Multiple *R*-squared: 0.1165, Adjusted *R*-squared: 0.1156. *F*-statistic: 131.6 on 2 and 1997 df, *P*-value: $<2.2 \times 10^{-16}$.

Table 5: Mixed estimator results from fitting the pseudo-biomass data to equation (17)

Coefficient	Estimate	Standard Error	*t*-value	Pr(>|*t*|)
Intercept	$-1.330 \times 10^{+01}$	1.83401	−0.725	0.460
X	-8.941×10^{-05}	9.274×10^{-06}	−9.641	$<2 \times 10^{-16}$
Z	1.204×10^{-04}	8.771×10^{-06}	13.727	$<2 \times 10^{-16}$

spirit to Bayesian, James–Stein or composite estimators (Green and Strawderman, 1990). These other approaches are typically used when the two estimates are to be combined with weights based on variances. The mixed estimator can also combine two estimates, but it provides additional flexibility in the form of constraints. The constraints were used here to help ensure that the two inventories would give similar estimates within predefined zones.

Conflict of Interest Statement

None declared.

References

Card, D.H. 1982 Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogramm. Eng. Remote Sens.* **48**, 431–439.

GRASS Development Team 2005 *Geographic Resources Analysis Support System (GRASS) 2. Grass Programmer's Manual.* ITC-irst, Trento, Italy. http://grass.itc.it/devel/index.php.

Green, E.J. and Strawderman, W.E. 1990 Combining inventory estimates with possibly biased auxiliary information. *For. Sci.* **36**, 688–699.

Green, E.J., Strawderman, W.E. and Airola, T.M. 1993 Assessing classification probabilities for thematic maps. *Photogramm. Eng. Remote Sens.* **59**, 635–639.

McRoberts, R.E., Nelson, M.D., Wendt, D.G. 2002 Stratified estimation of forest area using satellite imagery, inventory data, and the k-nearest neighbors technique. *Remote Sens. Environ.* **82**, 457–468.

R Development Core Team 2007 *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Roesch, F.A. and Van Deusen, P.C. 1995 A comparison of various estimators for updating forest area coverage using AVHRR and forest inventory data. *Photogramm. Eng. Remote Sens.* **61**, 307–311.

Story, M. and Congalton, R.G. 1986 Accuracy assessment: a user's perspective. *Photogramm. Eng. Remote Sens.* **59**, 635–639.

Theil, H. 1971 *Principles of Econometrics.* Wiley, New York.

US Department of Agriculture, Forest Service 2005 *Forest Inventory and Analysis National Core Field Guide, Volume 1: Field Data Collection Procedures for Phase 2 Plots, Version 3.0.* US Department of Agriculture, Forest Service, Washington Office. Internal report. On file with: US Department of Agriculture, Forest Service, Forest Inventory and Analysis, Rosslyn Plaza, 1620 North Kent Street, Arlington, VA 22209.

Van Deusen, P.C. 1994 Correcting bias in change estimates from thematic maps. *Remote Sens. Environ.* **50**, 57–73.

Van Deusen, P.C. 1996 Unbiased estimates of class proportions from thematic maps. *Photogramm. Eng. Remote Sens.* **62**, 409–412.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C. R., Wylie, B.K., and Van Driel, J.N. 2001 Completion of the 1990s National Land Cover Data Set for the conterminous United States. *Photogramm. Eng. Remote Sens.* **67**, 650–662.

Received 2 February 2007