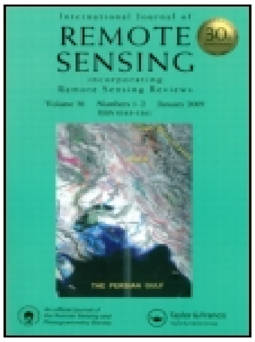

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes

Stephen V. Stehman

International Journal of Remote Sensing, 2014
Vol. 35, No. 13, 4923–4939, http://dx.doi.org/10.1080/01431161.2014.930207

Taylor & Francis
Taylor & Francis Group

TECHNICAL NOTE

Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes

Stephen V. Stehman*

College of Environmental Science & Forestry, State University of New York, Syracuse, NY 13210, USA

(*Received 7 October 2013; accepted 24 January 2014*)

The results of an accuracy assessment are typically organized using an error matrix that displays the proportion of area correctly mapped for each class and the proportion of area misclassified. Stratified random sampling is commonly implemented to obtain the reference data used to estimate the error matrix. When the strata correspond exactly to the map classes, the formulas for estimating accuracy and area are well known. Nevertheless, applications arise in which the strata are different from the map classes, as for example when the stratification is based on the map class labels of one map but the sample is subsequently used to assess the accuracy of other maps. In this paper, the estimators required when the stratum label and map label do not match for all pixels are presented for the proportion of area of each class based on the reference classification and for overall, user's, and producer's accuracies. Standard error formulas are also presented. A numerical example is provided to illustrate the computations.

1. Introduction

Validating land cover and land-cover change products provides critical data quality information to users and producers of these maps. A common practice is to conduct an accuracy assessment of the map based on a spatially explicit comparison of the map to a higher-quality classification called the reference classification. In this article, it will be assumed that a pixel-based accuracy assessment has been conducted, and that the map and reference labels are both based on a hard classification in which a single map class label and a single reference class label are assigned to each pixel. The goal of an accuracy assessment is to estimate a population error matrix (Table 1) and various descriptive measures of accuracy derived from this error matrix.

Because it is impractical to obtain the reference classification for the entire region of interest, a sample is selected and accuracy parameters are estimated from a sample. Stratified random sampling is commonly used in accuracy assessment, with the strata determined by the map label of each pixel (i.e. each map class is a stratum). In such applications, the rows of the error matrix correspond to the strata. However, circumstances may arise in practice in which the strata do not correspond exactly to the map classes. Two common situations in which the stratum label of a pixel may differ from the map label are when a stratified sample is used to assess the accuracy of more than one map and when a map is revised after the stratified sample has been selected. As an example of the first situation, Myeong et al. (2001) evaluated six different classification approaches for mapping urban cover in Syracuse, New York, USA. To conduct the accuracy assessment

*Email: svstehma@syr.edu

Table 1. Population error matrix in terms of pixel counts: N_{ij} = the number of pixels with map class i and reference class j for the population. The notation for a population error matrix expressed in terms of proportion of area replaces N_{ij}, with $P_{ij} = N_{ij}/N$. If the stratum label is equivalent to the map label for all pixels, the rows of the error matrix are also strata and an error matrix of sample counts can be obtained by replacing N_{ij} (the population count) with n_{ij} (the sample count). If the stratum label and map label do not match for some pixels, organizing the error matrix in terms of sample counts is not advised.

Reference class								
		1	2	…	k	…	q	Total
	1	N_{11}	N_{12}	…	N_{1k}	…	N_{1q}	N_{1+}
	2	N_{21}	N_{22}	…	N_{2k}	…	N_{2q}	N_{2+}
Map	:	:	:		:		:	:
Class	k	N_{k1}	N_{k2}	…	N_{kk}	…	N_{kq}	N_{k+}
	:	:	:		:		:	:
	q	N_{q1}	N_{q2}	…	N_{qk}	…	N_{qq}	N_{q+}
	Total	N_{+1}	N_{+2}	…	N_{+k}	…	N_{+q}	N

of these six approaches using a common reference sample, Myeong et al. (2001) selected a single stratified random sample where 50 reference sample points were obtained from each of five strata. The map classes used as the strata in the sample selection were determined from one of the six classification approaches, so for the other five classification approaches, the stratum label would not match the map label for every pixel in the region. The same accuracy assessment sample can still be used to evaluate the accuracy of all six classification approaches in the Myeong et al. (2001) application, but the analysis must take into account that the strata do not align perfectly with the map classes for five of the six approaches.

As an example of the second situation, suppose that a mapping project has a two-year budget and that it is necessary to commence data collection for accuracy assessment during the first year of the project to ensure that the reference data collection is complete in time to assess the final map at the end of the second year. A preliminary map is produced and this map is used to define strata corresponding to the map classes. A stratified random sample is selected and the reference data are obtained. In the meantime, the map producers continue to revise and improve the map and the end-product map differs from the preliminary map used to define the strata. Consequently, some of the sample pixels have a map label that differs from the stratum label and the rows of the error matrix of the end-product map are no longer equivalent to the strata used to select the sample.

As a final example, Wickham et al. (2013) implemented a stratified sampling design to assess the accuracy of a 2001–2006 land-cover change product. The design employed 22 strata defined by different types of change and no change (e.g. strata included urban gain, forest loss, agriculture loss, persistent water, and persistent forest). The assignment of a pixel to a stratum was based on the map label of the pixel for the 2001–2006 classification. The objectives specified by Wickham et al. (2013) also included assessing the accuracy of the 2001 and 2006 maps as single-date land-cover products. The sample pixels from a particular map class for each single-date product could arise from a number of different strata. For example, a sample pixel having the map class of forest in 2001 could have been selected from the persistent (i.e. no change) forest stratum or from the forest loss stratum. The map classes or rows of the error matrix for the 2001 land-cover map would not correspond to the strata used in the sample selection as these strata were determined from the 2001–2006 map.

The objective of this technical note is to describe the sampling theory and methods used for estimating accuracy and area when the map label and stratum label do not match for all pixels. The difficulty that arises when the map classes are not equivalent to the strata is that the familiar and commonly used stratified estimators for accuracy assessment (e.g. Card 1982; Stehman and Foody 2009; Olofsson et al. 2013) applied to estimate an error matrix and associated accuracy parameters are not appropriate in this setting. Fortunately, the stratified estimates of accuracy and area that can be applied to address this problem are relatively easy to compute and can be derived from basic sampling theory (Cochran 1977). The inference framework invoked throughout will be design-based inference (Särndal, Swensson, and Wretman 1992; Stehman 2000).

2. Estimating accuracy and area

A fundamental feature of the approach to estimation when the map classes do not correspond to the strata used to select the sample is that the stratum to which a pixel was assigned when the sample was selected remains a characteristic of that pixel. That is, even though the map class of a pixel may change after the reference sample has been selected, the stratum from which that pixel was selected has not changed. For example, suppose a pixel is selected from the forest stratum (where the forest stratum is defined based on a land-cover map of the region) but for the map that is being evaluated, this pixel is mapped as cropland (i.e. the pixel was relabelled from the forest class in the map version used to construct the strata to cropland in the version of the map to be assessed). This pixel is included in the sample as the result of having been selected from the forest stratum. The information about the process leading to selection of this pixel into the sample must be retained in the analysis, even though the data for this pixel will be included in the cropland map class when the pixel's map and reference class data are incorporated into the error matrix.

2.1. Inclusion probabilities

The sampling theory underlying the general approach to stratified estimation is formalized as follows. The sampling design is the protocol or sequence of steps for selecting the sample of pixels. The sampling design determines the inclusion probability associated with each pixel, where the inclusion probability for pixel, u (denoted as π_u) is defined as the probability that the sample selected will contain pixel u. For stratified random sampling in which a simple random sample of n_h^* sample pixels is selected from the N_h^* pixels of stratum h in the population, the inclusion probability for pixel u in stratum h is $\pi_u = n_h^*/N_h^*$. If the strata and map classes correspond exactly such that for all pixels the map label is the same as the stratum label, the inclusion probability for pixel u in stratum h can also be expressed as $\pi_u = n_h^*/N_h^* = n_{h+}/N_{h+}$. The notation distinguishes stratum population size (N_h^*) and stratum sample size (n_h^*) from the population count (N_{h+}) and sample count (n_{h+}) of the rows (map classes) of the error matrix (Table 1). If the strata are the same as the map classes of the error matrix, then $n_h^* = n_{h+}$ and $N_h^* = N_{h+}$. When the strata are not equivalent to the rows of the error matrix, the inclusion probability is still

$\pi_u = n_h^*/N_h^*$, but it is not possible to determine the inclusion probabilities from the map row totals of the Table 1 error matrix.

When the stratum label does not match the map class label for all pixels, Table 1 is still a valid representation of the population error matrix, but as stated previously, the rows of Table 1 are not equivalent to the strata of the stratified sample (i.e. $n_{h+} \neq n_h^*$ and $N_{h+} \neq N_h^*$). Furthermore, if the strata do not correspond to the map classes, organizing the sample data in terms of sample counts (n_{ij}, $i, j = 1, \ldots, q$) using Table 1 would not be advisable because the sample pixels contributing to n_{ij} in any given cell of the error matrix may have originated from different strata. The stratum origin of each sample pixel must be retained and used when constructing the estimators of accuracy, so an error matrix displaying sample counts will not be informative and instead the cells of the error matrix should be expressed and estimated in terms of proportion of area.

2.2. *Stratified estimator of a mean*

The general estimation theory for stratified random sampling that will be applied to the estimate accuracy and area focuses on a population mean, $\bar{Y}$, as a target parameter to be estimated:

$$\bar{Y} = \sum_{u=1}^{N} y_u/N, \tag{1}$$

where N is number of pixels in the population (i.e. region of interest) and y_u is the observation for pixel u. The definition of y_u depends on the parameter (e.g. overall accuracy, proportion of area) to be estimated. An unbiased estimator of $\bar{Y}$ may be formulated in terms of the inclusion probabilities as

$$\hat{\bar{Y}} = \sum_{u \in s} (y_u/\pi_u)/N, \tag{2}$$

where the summation is over the units included in the sample, s (Särndal, Swensson, and Wretman 1992, Section 2.8). The inclusion probability (π_u) is determined by the sampling design. Because for stratified random sampling the inclusion probability is $\pi_u = n_h^*/N_h^*$, Equation (2) simplifies to

$$\hat{\bar{Y}} = \sum_{h=1}^{H} N_h^* \bar{y}_h/N, \tag{3}$$

where $\bar{y}_h = \sum_{u \in h} y_u/n_h^*$ is the sample mean of the y_u values in stratum h, $u \in h$ indicates that sample pixel u was selected from stratum h, and H is the number of strata (Särndal, Swensson, and Wretman 1992, 102). In Equation (3), each sample unit is accounted for in the stratum to which it was assigned at the sample selection stage, thus highlighting the importance of retaining the stratum identity of each sample pixel.

2.3. *Response design*

Defining the observation y_u is the key to estimating accuracy and area when the strata and map classes are not the same. The response design determines y_u; that is, the response design is defined as the protocol for assigning the different attributes or labels associated with a pixel that are used in the analysis (Stehman and Czaplewski 1998). For example, the reference class label of a pixel could be coded as $y_u = 1$ if pixel u is the reference class of interest, and $y_u = 0$ if pixel u is not the class of interest (e.g. forest could be the class of interest so y_u would be defined as 1 or 0 depending on whether the pixel had a reference label of forest). The stratum label of pixel u determines the inclusion probability (π_u) whereas the response design determines the observation y_u for pixel u. Because the inclusion probability of each pixel is determined by the stratum label that existed at the time of the sample selection, once the sample has been selected the inclusion probability of pixel u is fixed (an exception to this 'fixed' inclusion probability feature is post-stratified estimation, Section 4). But y_u can be defined differently depending on the accuracy or area parameter to be estimated, and the choice of y_u has no impact on π_u because the inclusion probabilities are determined prior to the observation of y_u.

2.4. Conventional stratified estimators of accuracy and area

A parameter is defined as a number describing the population (i.e. a number obtained from a census of the region of interest). The parameters typically used in accuracy assessment are the overall, user's and producer's accuracies. These parameters are readily identified from a population error matrix (see Table 1 for definitions of quantities) as follows.

Overall accuracy:

$$O = \sum_{i=1}^{q} N_{ii}/N = \sum_{i=1}^{q} P_{ii}. \tag{4}$$

User's accuracy of class k:

$$U_k = N_{kk}/N_{k+} = P_{kk}/P_{k+}. \tag{5}$$

Producer's accuracy of class k:

$$P_k = N_{kk}/N_{+k} = P_{kk}/P_{+k}. \tag{6}$$

The proportion of area of a class as determined from the reference classification (e.g. the proportion of area of forest where the forest class is determined from the reference data) is defined by the parameter

$$P_k^{\mathrm{A}} = N_{+k}/N = P_{+k}, \tag{7}$$

(the superscript A is used to highlight that this parameter represents area). Parameters do not depend on the sample, so Equations (4)–(7) are valid regardless of the choice of sampling design.

When the strata correspond to the map classes, the accuracy estimates can be readily obtained from the error matrix of sample counts. The error matrix of sample counts is obtained by replacing N_{ij} in Table 1 with n_{ij}, which is the number of sample pixels with map class i and reference class j, and n_{h+} and N_{h+} can be used instead of n_h^* and N_h^*

because the strata correspond exactly to the map classes. The conventional stratified estimators applicable when the stratum label and map label match for all pixels are

$$\hat{O} = (1/N)\sum_{h=1}^{q}\frac{N_{h+}}{n_{h+}}n_{hh}, \tag{8}$$

$$\hat{U}_k = n_{kk}/n_{k+}, \tag{9}$$

$$\hat{P}_k = \frac{n_{kk}(N_{k+}/n_{k+})}{\sum_{h=1}^{q} n_{hk}(N_{h+}/n_{h+})}, \tag{10}$$

and the stratified estimator (Stehman 2013) of the proportion of area of class k is

$$\hat{P}_k^{\mathrm{A}} = (1/N)\sum_{h=1}^{q}\frac{N_{h+}}{n_{h+}}n_{hk}. \tag{11}$$

When the strata do not correspond to the map classes, the parameters (4)–(7) remain the same, but estimating these parameters from the sample data requires taking into consideration the sampling design. Consequently, the conventional estimators given by Equations (8)–(11) are biased estimators when the rows of the population error matrix do not correspond to the strata used to select the sample, and different estimators must be constructed as described in the following section.

2.5. *Estimating accuracy and area using indicator functions*

When the rows of the population error matrix do not correspond to the strata, the estimators can be more easily derived by expressing the accuracy and area parameters in terms of simple indicator observations denoted as y_u and x_u, respectively, where these observations obtained for pixel u have just two possible values, 0 or 1. For example, if we define

$$y_u = \begin{cases} 1 & \text{if pixel } u \text{ is classified correctly} \\ 0 & \text{if pixel } u \text{ is classified incorrectly} \end{cases}, \tag{12}$$

the parameter for overall accuracy (Equation (4)) can be expressed as

$$O = \sum_{u=1}^{N} y_u/N = \bar{Y}, \tag{13}$$

where $\bar{Y}$ is the population mean based on a census of N pixels in the region, and $\bar{Y}$ is equivalent to the proportion of pixels correctly classified. The proportion of area of class k (where class k is the reference classification) can be defined in an analogous manner using

$$y_u = \begin{cases} 1 & \text{if pixel } u \text{ is reference class } k \\ 0 & \text{if pixel } u \text{ is not reference class } k \end{cases}, \tag{14}$$

and expressing the parameter as

$$P_k^{\mathrm{A}} = \sum_{u=1}^{N} y_u / N = \bar{Y}. \tag{15}$$

The use of 0 and 1 indicator functions (e.g. Equations (12) and (14)) is common in sampling, with typical applications being the estimation of a proportion (Cochran 1977, 50; Scheaffer, Mendenhall, and Ott 2006, 92) and estimation of means of domains or subpopulations (Cochran 1977, 36 and 143; Lohr 2010, 134). The indicator function approach has been previously applied to certain aspects of accuracy estimation (e.g. Edwards, Moisen, and Cutler 1998; Nusser and Klaas 2003; Stehman et al. 2003; Wickham et al. 2013; Zimmerman et al. 2013), but these previous applications do not provide a comprehensive treatment of the indicator function approach for stratified sampling.

Estimating the cell proportions of the error matrix (P_{ij}) follows the same general approach outlined to estimate the proportion of area ($P_k{}^{\mathrm{A}}$) and overall accuracy (O). To estimate the proportion of area for the error matrix entry of row (map class) i and column (reference class) j, we define

$$y_u = \begin{cases} 1 & \text{if pixel } u \text{ is map class } i \text{ and reference class } j \\ 0 & \text{otherwise} \end{cases}. \tag{16}$$

Then the parameter representing the population proportion of pixels in row i and column j of the population error matrix is

$$P_{ij} = N_{ij}/N = \sum_{u=1}^{N} y_u / N = \bar{Y}, \tag{17}$$

where N_{ij} is the total number of pixels in the population having map class i and reference class j. Note that y_u is defined differently (Equations (12), (14), and (16)) depending on the target parameter. But once y_u is defined, the same estimator [3] can be used for the target parameter $\bar{Y}$ where the definition of y_u determines whether the parameter estimated is O = overall accuracy, $P_k{}^{\mathrm{A}}$ = proportion of area of reference class k, or P_{ij} = proportion of area with map class i and reference class j.

Estimating the user's accuracy and producer's accuracy requires an additional indicator function which will be denoted as x_u. For the user's accuracy, define

$$y_u = \begin{cases} 1 & \text{if pixel } u \text{ is classified correctly and has map class } k \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

and

$$x_u = \begin{cases} 1 & \text{if pixel } u \text{ is map class } k \\ 0 & \text{otherwise} \end{cases}. \tag{19}$$

Then the parameter for user's accuracy of class k is expressed in the form of a ratio:

$$R = \frac{\sum_{u=1}^{N} y_u}{\sum_{u=1}^{N} x_u}, \tag{20}$$

where R is the total number of pixels of map class k that are correctly classified (N_{kk}) divided by the total number of pixels mapped as class k (N_{k+}). In this case, $R = N_{kk}/N_{k+}$, the parameter defined as user's accuracy (Equation 5). The parameter representing the proportion of commission errors can also be expressed as a ratio R by defining x_u as in Equation (19) and y_u as

$$y_u = \begin{cases} 1 & \text{if pixel } u \text{ is classified incorrectly and has map class } k \\ 0 & \text{otherwise} \end{cases}. \tag{21}$$

To obtain the parameter for producer's accuracy (Equation (6)) of class k, define

$$y_u = \begin{cases} 1 & \text{if pixel } u \text{ is correctly classified and has reference class } k \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

and

$$x_u = \begin{cases} 1 & \text{if pixel } u \text{ is has reference class } k \\ 0 & \text{otherwise} \end{cases}. \tag{23}$$

If y_u and x_u are defined by Equations (22) and (23), the parameter for producer's accuracy can then be expressed as the ratio R of Equation (20) which is equivalent to $R = N_{kk}/N_{+k}$ (Equation 6). The parameter for the proportion of omission errors of class k can similarly be expressed as a ratio R if x_u is defined by Equation (23) and y_u is defined as

$$y_u = \begin{cases} 1 & \text{if pixel } u \text{ is incorrectly classified and has reference class } k \\ 0 & \text{otherwise} \end{cases}. \tag{24}$$

Given the preceding expressions of the accuracy and area parameters, the task of estimating these parameters for stratified random sampling reduces to two cases, estimating a mean $\bar{Y}$ when the objective is to estimate P_k^{A} = proportion of area of reference class k, O = overall accuracy, and P_{ij} = proportion of area in cell (i, j) of the error matrix; and estimating a ratio R in the case of the user's accuracy, producer's accuracy, commission error, and omission error. Advantages of the approach outlined based on expressing parameters in terms of the indicator observations x_u and y_u are that the estimates are relatively simple to compute and variance estimators are available from sampling theory for stratified estimation.

An unbiased estimator of $\bar{Y}$ for stratified random sampling (Equation 3) is

$$\hat{\bar{Y}} = \sum_{h=1}^{H} N_h^* \bar{y}_h / N,$$

where $\bar{y}_h = \sum_{u \in h} y_u / n_h^*$ is the sample mean of the y_u values in stratum h, $u \in h$ indicates that sample pixel u was selected from stratum h, and H denotes the number of strata. If the map classes are used to define the strata, $H = q$, where q is the number of classes represented by the error matrix. An estimator of the variance of $\hat{\bar{Y}}$ is

$$\hat{V}(\hat{\bar{Y}}) = (1/N^2) \sum_{h=1}^{H} N_h^{*2} (1 - n_h^*/N_h^*) s_{yh}^2 / n_h^*, \tag{25}$$

where the sample variance of the y_u values from stratum h is

$$s_{yh}^2 = \sum_{u \in h} (y_u - \bar{y}_h)^2 / (n_h^* - 1). \tag{26}$$

The standard error of $\hat{\bar{Y}}$ is the square root of the estimated variance. For example, to estimate overall accuracy, define y_u by Equation (12) and use this y_u in Equations (3), (25), and (26). Estimating $P_k{}^{\mathrm{A}}$ and P_{ij} would similarly use these equations with y_u defined by Equations (14) and (16). Numerical examples illustrating the data organization and computational details of the stratified estimators are presented in Section 3.

To estimate the ratio R from a stratified random sample, a combined ratio estimator is used:

$$\hat{R} = \frac{\sum_{h=1}^{H} N_h^* \bar{y}_h}{\sum_{h=1}^{H} N_h^* \bar{x}_h}, \tag{27}$$

where $\bar{y}_h$ and $\bar{x}_h$ are the stratum-specific sample means of y_u and x_u (Cochran 1977, Section 6.11). $\hat{R}$ is not an unbiased estimator of R, but $\hat{R}$ is a design-consistent estimator, where a design-consistent estimator can be interpreted to mean that when the sample size is large, the sampling distribution of the estimator will be tightly concentrated around the parameter (Särndal, Swensson, and Wretman 1992, 166). The variance estimator of $\hat{R}$ is

$$\hat{V}(\hat{R}) = \frac{1}{\hat{X}^2} \sum_{h=1}^{H} N_h^{*2} (1 - n_h^*/N_h^*) (s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2\hat{R} s_{xyh}) / n_h^*, \tag{28}$$

where $\hat{X} = \sum_{h=1}^{H} N_h^* \bar{x}_h$, s_{yh}^2 is as defined by Equation (26), s_{xh}^2 is the sample variance of the x_u values within stratum h (computed analogously to s_{yh}^2), and s_{xyh} is the sample covariance between x_u and y_u for stratum h:

$$s_{xyh} = \sum_{u \in h} (y_u - \bar{y}_h)(x_u - \bar{x}_h) / (n_h^* - 1). \tag{29}$$

The standard error of $\hat{R}$ is the square root of $\hat{V}(\hat{R})$. Equations (27)–(29) can be applied to estimate user's and producer's accuracies, or commission and omission error rates using the specified definitions of y_u and x_u for each parameter.

3. Numerical examples

Numerical examples using the hypothetical data shown in Table 2 illustrate the approach to estimation when the strata do not correspond exactly to the map classes. Suppose the map undergoing the accuracy assessment has four classes designated as A, B, C, and D. The region of interest is partitioned into four strata, but the stratum label does not match the map label for some pixels (e.g. the stratum labels were determined from a map different from the one being assessed). Suppose that the stratum sizes are $N_1^* = 40{,}000$, $N_2^* = 30{,}000$, $N_3^* = 20{,}000$, and $N_4^* = 10{,}000$ and a stratified random sample is selected with equal allocation to each stratum ($n_1^* = n_2^* = n_3^* = n_4^* = 10$). The map and reference classifications for each sample pixel are shown in Table 2. To illustrate application of the stratified estimators, we will examine estimation of the proportion of area of classes A and C (these are area proportions based on the reference classification), overall accuracy, and user's and producer's accuracy of class B. For each estimator, we define y_u and, when needed, x_u, depending on the parameter of interest. The translation of the map and reference class labels to the appropriate y_u and x_u for each parameter is shown in Table 2. The next step is to compute the sample means, variances, and covariances for each stratum using the y_u and x_u observations defined in Table 2. These stratum-specific statistics are presented in Table 3. Given these preliminary calculations, it is now possible to estimate the parameters of interest using Equations (3) and (27), and to estimate the standard errors using Equations (25) and (28).

Proportion of area of class A:

$$\hat{\bar{Y}} = \sum_{h=1}^{4} N_h^* \bar{y}_h / N = [40{,}000(0.6) + 30{,}000(0.3) + 20{,}000(0.10) + 10{,}000(0.00)]/100{,}000$$
$$= [24{,}000 + 9000 + 2000 + 0]/100{,}000 = 35{,}000/100{,}000 = 0.35.$$

Proportion of area of class C:

$$\begin{aligned}\hat{\bar{Y}} &= \sum_{h=1}^{4} N_h^* \bar{y}_h / N \\ &= [40{,}000(0.20) + 30{,}000(0.00) + 20{,}000(0.50) + 10{,}000(0.20)]/100{,}000 \\ &= [8000 + 0 + 10{,}000 + 2000]/100{,}000 = 20{,}000/100{,}000 = 0.20.\end{aligned}$$

Overall accuracy:

$$\begin{aligned}\hat{\bar{Y}} &= \sum_{h=1}^{4} N_h^* \bar{y}_h / N \\ &= [40{,}000(0.60) + 30{,}000(0.80) + 20{,}000(0.40) + 10{,}000(0.70)]/100{,}000 \\ &= [24{,}000 + 24{,}000 + 8000 + 7000]/100{,}000 = 63{,}000/100{,}000 = 0.63.\end{aligned}$$

Table 2. Sample data for example calculation of proportion of area of reference classes A and C, overall accuracy (O), user's and producer's accuracy of class B, and P_{23} = the proportion of area in row 2 (map class) and column 3 (reference class) of the 3 × 3 error matrix estimated from the sample (the numbers of the equations defining y_u and x_u are given in parentheses beside each variable).

			Area (class A)	Area (class C)	O	User's accuracy (class B)		Producer's accuracy (class B)		P_{23}
Stratum	Map class	Reference class	y_u (14)	y_u (14)	y_u (12)	y_u (18)	x_u (19)	y_u (22)	x_u (23)	y_u (16)
1	A	A	1	0	1	0	0	0	0	0
1	A	A	1	0	1	0	0	0	0	0
1	A	A	1	0	1	0	0	0	0	0
1	A	A	1	0	1	0	0	0	0	0
1	A	A	1	0	1	0	0	0	0	0
1	A	C	0	1	0	0	0	0	0	0
1	A	B	0	0	0	0	0	0	1	0
1	B	A	1	0	0	0	1	0	0	0
1	B	B	0	0	1	1	1	1	1	0
1	B	C	0	1	0	0	1	0	0	1
2	A	A	1	0	1	0	0	0	0	0
2	B	B	0	0	1	1	1	1	1	0
2	B	B	0	0	1	1	1	1	1	0
2	B	B	0	0	1	1	1	1	1	0
2	B	B	0	0	1	1	1	1	1	0
2	B	B	0	0	1	1	1	1	1	0
2	B	A	1	0	0	0	1	0	0	0
2	B	A	1	0	0	0	1	0	0	0
2	B	B	0	0	1	1	1	1	1	0
2	B	B	0	0	1	1	1	1	1	0
3	B	C	0	1	0	0	1	0	0	1
3	B	C	0	1	0	0	1	0	0	1
3	C	C	0	1	1	0	0	0	0	0
3	C	C	0	1	1	0	0	0	0	0
3	C	C	0	1	1	0	0	0	0	0
3	C	D	0	0	0	0	0	0	0	0
3	C	D	0	0	0	0	0	0	0	0
3	C	B	0	0	0	0	0	0	1	0
3	B	B	0	0	1	1	1	1	1	0
3	B	A	1	0	0	0	1	0	0	0
4	D	D	0	0	1	0	0	0	0	0
4	D	D	0	0	1	0	0	0	0	0
4	D	D	0	0	1	0	0	0	0	0
4	D	D	0	0	1	0	0	0	0	0
4	D	D	0	0	1	0	0	0	0	0
4	D	D	0	0	1	0	0	0	0	0
4	D	D	0	0	1	0	0	0	0	0
4	D	C	0	1	0	0	0	0	0	0
4	D	C	0	1	0	0	0	0	0	0
4	D	B	0	0	0	0	0	0	1	0

Table 3. Summary statistics required for computing estimates of proportion of area and accuracy using the Table 2 sample data (Var is the sample variance per stratum, s^2_{yh} or s^2_{xh}; covar is the covariance between x_u and y_u per stratum, represented by s_{xyh} in the text).

		Area (class A)	Area (class C)	Overall accuracy	User's accuracy (class B)		Producer's accuracy (class B)		P_{23}
Stratum		y_u	y_u	y_u	y_u	x_u	y_u	x_u	y_u
1	Mean	0.60	0.20	0.60	0.10	0.30	0.10	0.20	0.1
	Var	0.267	0.178	0.267	0.100	0.233	0.100	0.178	–
	Covar	–	–	–	0.0778		0.0889	–	
2	Mean	0.30	0.00	0.80	0.70	0.90	0.70	0.70	0.0
	Var	0.233	0.000	0.178	0.233	0.100	0.233	0.233	–
	Covar	–	–	–	0.0778		0.2333	–	
3	Mean	0.10	0.50	0.40	0.10	0.40	0.10	0.20	0.2
	Var	0.100	0.278	0.267	0.100	0.267	0.100	0.178	–
	Covar	–	–	–	0.0667		0.0889	–	
4	Mean	0.00	0.20	0.70	0.00	0.00	0.00	0.10	0.0
	Var	0.000	0.178	0.233	0.000	0.000	0.000	0.100	–
	Covar	–	–	–	0.0000		0.0000	–	

User's accuracy of class B:

$$\sum_{h=1}^{4} N_h^* \bar{y}_h = [40{,}000(0.10) + 30{,}000(0.70) + 20{,}000(0.10) + 10{,}000(0.00)]$$
$$= [4000 + 21{,}000 + 2000 + 0] = 27{,}000.$$

$$\hat{X} = \sum_{h=1}^{4} N_h^* \bar{x}_h = [40{,}000(0.30) + 30{,}000(0.90) + 20{,}000(0.40) + 10{,}000(0.00)]$$
$$= [12{,}000 + 27{,}000 + 8000 + 0] = 47{,}000.$$

$$\hat{R} = \frac{\sum_{h=1}^{4} N_h^* \bar{y}_h}{\sum_{h=1}^{4} N_h^* \bar{x}_h} = \frac{27{,}000}{47{,}000} = 0.574.$$

Producer's accuracy of class B:

$$\sum_{h=1}^{4} N_h^* \bar{y}_h = [40{,}000(0.10) + 30{,}000(0.70) + 20{,}000(0.10) + 10{,}000(0.00)]$$
$$= [4000 + 21{,}000 + 2000 + 0] = 27{,}000.$$

$$\hat{X} = \sum_{h=1}^{4} N_h^* \bar{x}_h = [40{,}000(0.20) + 30{,}000(0.70) + 20{,}000(0.20) + 10{,}000(0.10)]$$
$$= [8000 + 21{,}000 + 4000 + 1000] = 34{,}000.$$

$$\hat{R} = \frac{\sum_{h=1}^{4} N_h^* \bar{y}_h}{\sum_{h=1}^{4} N_h^* \bar{x}_h} = \frac{27{,}000}{34{,}000} = 0.794.$$

Cell (*i, j*) of the error matrix, P_{ij} (*i* = 2, *j* = 3):

$$\hat{\bar{Y}} = \sum_{h=1}^{4} N_h^* \bar{y}_h / N = [40{,}000(0.1) + 30{,}000(0) + 20{,}000(0.2) + 10{,}000(0)]/100{,}000$$
$$= [4000 + 0 + 4000 + 0]/100{,}000 = 8000/100{,}000 = 0.08.$$

The standard error calculations for the estimates are as follows (finite population correction terms, $1 - n_h^*/N_h^*$, are ignored for these calculations as they have no noticeable effect on the standard error estimates when n_h^*/N_h^* is small as in this example).

Standard error (SE) for proportion of area of class A:

$$\hat{V}(\hat{\bar{Y}}) = (1/N^2) \sum_{h=1}^{H} N_h^{*2} (1 - n_h^*/N_h^*) s_{yh}^2 / n_h^*$$
$$= [1/100{,}000^2][40{,}000^2(0.267/10) + 30{,}000^2(0.233/10)$$
$$+ 20{,}000^2(0.10/10) + 10{,}000^2(0.00/10)]$$
$$= [1/100{,}000^2][42{,}720{,}000 + 20{,}970{,}000 + 4{,}000{,}000 + 0]$$
$$= [1/100{,}000^2][67{,}690{,}000] = 0.006769.$$

$$\mathrm{SE}(\hat{\bar{Y}}) = \sqrt{\hat{V}(\hat{\bar{Y}})} = \sqrt{0.006769} = 0.082.$$

SE for proportion of area of class C:

$$\hat{V}(\hat{\bar{Y}}) = (1/N^2) \sum_{h=1}^{H} N_h^{*2} (1 - n_h^*/N_h^*) s_{yh}^2 / n_h^*$$
$$= [1/100{,}000^2][40{,}000^2(0.178/10) + 30{,}000^2(0/10)$$
$$+ 20{,}000^2(0.278/10) + 10{,}000^2(0.178/10)]$$
$$= [1/100{,}000^2][28{,}480{,}000 + 0 + 11{,}120{,}000 + 1{,}780{,}000]$$
$$= [1/100{,}000^2][41{,}380{,}000] = 0.004138.$$

$$\mathrm{SE}(\hat{\bar{Y}}) = \sqrt{\hat{V}(\hat{\bar{Y}})} = \sqrt{0.004138} = 0.064.$$

SE for overall accuracy:

$$
\begin{aligned}
\hat{V}(\hat{\bar{Y}}) &= (1/N^2)\sum_{h=1}^{H} N_h^{*2}(1 - n_h^*/N_h^*)s_{yh}^2/n_h^* \\
&= [1/100{,}000^2][40{,}000^2(0.267/10) + 30{,}000^2(0.178/10) + 20{,}000^2(0.267/10) \\
&\quad + 10{,}000^2(0.233/10)] \\
&= [1/100{,}000^2][42{,}720{,}000 + 16{,}020{,}000 + 10{,}680{,}000 + 2{,}330{,}000] \\
&= [1/100{,}000^2][42{,}720{,}000 + 16{,}020{,}000 + 10{,}680{,}000 + 2{,}330{,}000] \\
&= [1/100{,}000^2][71{,}750{,}000] = 0.007175.
\end{aligned}
$$

$$
\mathrm{SE}(\hat{\bar{Y}}) = \sqrt{\hat{V}(\hat{\bar{Y}})} = \sqrt{0.007175} = 0.085.
$$

SE for user's accuracy of class B:

$$
\begin{aligned}
\hat{V}(\hat{R}) &= \frac{1}{\hat{X}^2}\sum_{h=1}^{H} N_h^{*2}(1 - n_h^*/N_h^*)(s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2\hat{R}s_{xyh})/n_h^* \\
&= \left[1/47{,}000^2\right]\{40{,}000^2\left[0.10 + 0.574^2(0.233) - 2(0.574)0.0778\right]/10 \\
&\quad + 30{,}000^2\left[0.233 + 0.574^2(0.10) - 2(0.574)0.0778\right]/10 \\
&\quad + 20{,}000^2\left[0.10 + 0.574^2(0.267) - 2(0.574)0.0667\right]/10 \\
&\quad + 10{,}000^2\left[0.00 + 0.574^2(0.00) - 2(0.574)0.00\right]/10\} \\
&= \left[1/47{,}000^2\right]\{13{,}992{,}561 + 15{,}896{,}988 + 4{,}455{,}940 + 0\} \\
&= \left[1/47{,}000^2\right]\{34{,}345{,}489\} = 0.01555.
\end{aligned}
$$

$$
\mathrm{SE}(\hat{R}) = \sqrt{\hat{V}(\hat{R})} = \sqrt{0.01555} = 0.125.
$$

SE for producer's accuracy of class B:

$$
\begin{aligned}
\hat{V}(\hat{R}) &= \frac{1}{\hat{X}^2}\sum_{h=1}^{H} N_h^{*2}(1 - n_h^*/N_h^*)(s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2\hat{R}s_{xyh})/n_h^* \\
&= \left[1/34,000^2\right]\{40{,}000^2\left[0.10 + 0.794^2(0.178) - 2(0.794)0.0889\right]/10 \\
&\quad + 30{,}000^2\left[0.233 + 0.794^2(0.233) - 2(0.794)0.2333\right]/10 \\
&\quad + 20{,}000^2\left[0.10 + 0.794^2(0.178) - 2(0.794)0.0889\right]/10 \\
&\quad + 10{,}000^2\left[0.00 + 0.794^2(0.00) - 2(0.794)0.00\right]/10\} \\
&= \left[1/34{,}000^2\right]\{11{,}367{,}105 + 847{,}007 + 2{,}841{,}776 + 0\} \\
&= \left[1/34{,}000^2\right]\{15{,}055{,}888\} = 0.01302.
\end{aligned}
$$

$$
\mathrm{SE}(\hat{R}) = \sqrt{\hat{V}(\hat{R})} = \sqrt{0.01302} = 0.114.
$$

4. Post-stratified estimation

When the strata used to select the sample do not correspond perfectly with the map class of each pixel, a natural question to ask is whether it is still possible to use the map classes to reduce the SEs of the accuracy and area estimators. The motivation for raising this issue is that for simple random sampling, a stratified estimator using the map classes as 'post-strata' (the technique is commonly referred to as post-stratified estimation) usually results in a smaller SE relative to the conventional simple random sampling estimator (Card 1982; Olofsson et al. 2013; Stehman 2013). If the sampling design is stratified random sampling (the focus of this article), in concept it is still possible to use a post-stratified estimator but the number of situations for which this could feasibly be done in practice may be limited. The complicating factor is that when the sampling design is stratified random, the post-stratified estimator must be applied within each of the original design strata (i.e. the post-strata would represent 'sub-strata' within each of the strata defined for selecting the stratified sample). For example, in the application described in Section 3, the post-stratified estimator could be applied within each of the four strata used in the stratified sampling design where the post-strata would be the map classes A–D. That is, stratum 1 would be partitioned into four post-strata depending on whether the pixel was classified as A, B, C, or D in the map being assessed (recall that the map label of a pixel in the map being evaluated may not match the stratum label), and the other three strata would similarly be partitioned into four post-strata each. A post-stratified estimator would then be applied within each of the original strata of the stratified random design. Unless the overall sample size is very large, some of these post-strata may be represented by few or no sample units. Särndal, Swensson, and Wretman (1992, 407) state that a prudent rule is to require 10 observations per 'cell' (i.e. a map class within one of the original design strata) for a post-stratified estimator to yield good results. Unless the map classes are substantially different from the strata (i.e. the map used to define the strata is very different from that being evaluated), it is likely that many of the post-strata will represent only a small proportion of area and that the overall sample size will need to be very large to yield 10 observations per post-stratum.

5. Discussion

The problem addressed in this article is how to estimate accuracy and area when the map label does not match the stratum label for some of the pixels in the target region. The recommended estimation strategy is based on recognizing that the inclusion probabilities of the sampled pixels are determined by the strata that existed when the sample was selected, and these inclusion probabilities remain fixed even if the map label of the pixel is subsequently changed. Stratified random sampling estimators for a mean ($\bar{Y}$) or ratio (R) can be applied to estimate accuracy and area by defining observations y_u and x_u depending on the parameter of interest. Although the primary focus has been applications in which the strata are based on the map classes, the estimators presented are applicable even when the strata are unrelated to the map classes. For example, if the strata are different geographic regions such as continents (e.g. for a global map) or provinces within a country, the estimators described can be used. This article focuses only on the estimation component of the sampling strategy and does not address issues of sampling design. For example, questions of how to define strata and allocate sample size to strata when the objective is to assess accuracy of several maps of the same region are not addressed.

The article demonstrates that a sample selected based on strata determined from one map can be used to estimate accuracy and area of other maps. Although the accuracy and area estimators will be unbiased (or design-consistent in the case of $\hat{R}$), it is possible that the SEs of these estimators will be higher for other maps than would have been the case had the stratification been created from the particular map being assessed. Nonetheless, the fact remains that a stratified sample has utility for accuracy assessment beyond just the map that was used to define the strata.

Acknowledgements

I thank two anonymous reviewers and the editor for their constructive comments and Dr Huiran Jin for checking the results of the numerical examples.

Funding

S. Stehman was supported by Grant/Cooperative Agreement Number G12AC20221 from the United States Geological Survey. The research is solely the responsibility of the author and does not necessarily represent the official views of the USGS.

References

Card, D. H. 1982. "Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy." *Photogrammetric Engineering and Remote Sensing* 48: 431–439.

Cochran, W. G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons.

Edwards Jr, T. C., G. G. Moisen, and D. R. Cutler. 1998. "Assessing Map Accuracy in a Remotely Sensed, Ecoregion-Scale Cover Map." *Remote Sensing of Environment* 63: 73–83. doi:10.1016/S0034-4257(96)00246-5.

Lohr, S. L. 2010. *Sampling: Design and Analysis*. 2nd ed. Boston, MA: Brooks/Cole.

Myeong, S., D. J. Nowak, P. F. Hopkins, and R. H. Brock. 2001. "Urban Cover Mapping Using Digital, High-Spatial Resolution Aerial Imagery." *Urban Ecosystems* 5: 243–256. doi:10.1023/A:1025687711588.

Nusser, S. M., and E. E. Klaas. 2003. "Survey Methods for Assessing Land Cover Map Accuracy." *Environmental and Ecological Statistics* 10: 309–331. doi:10.1023/A:1025107023980.

Olofsson, P., G. M. Foody, S. V. Stehman, and C. E. Woodcock. 2013. "Making Better Use of Accuracy Data in Land Change Studies: Estimating Accuracy and Area and Quantifying Uncertainty Using Stratified Estimation." *Remote Sensing of Environment* 129: 122–131. doi:10.1016/j.rse.2012.10.031.

Särndal, C.-E., B. Swensson, and J. Wretman. 1992. *Model-Assisted Survey Sampling*. New York, NY: Springer-Verlag.

Scheaffer, R. L., W. Mendenhall, and R. L. Ott. 2006. *Elementary Survey Sampling*. 6th ed. Belmont, CA: Duxbury Press.

Stehman, S. V. 2000. "Practical Implications of Design-Based Sampling Inference for Thematic Map Accuracy Assessment." *Remote Sensing of Environment* 72: 35–45. doi:10.1016/S0034-4257(99)00090-5.

Stehman, S. V. 2013. "Estimating Area from an Accuracy Assessment Error Matrix." *Remote Sensing of Environment* 132: 202–211. doi:10.1016/j.rse.2013.01.016.

Stehman, S. V., and G. M. Foody. 2009. "Accuracy Assessment." In *The SAGE Handbook of Remote Sensing*, edited by T. A. Warner, M. D. Nellis, and G. M. Foody, 297–309. London: Sage.

Stehman, S. V., J. D. Wickham, J. H. Smith, and L. Yang. 2003. "Thematic Accuracy of the 1992 National Land-Cover Data for the Eastern United States: Statistical Methodology and Regional Results." *Remote Sensing of Environment* 86: 500–516. doi:10.1016/S0034-4257(03)00128-7.

Stehman, S. V., and R. L. Czaplewski. 1998. "Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles." *Remote Sensing of Environment* 64: 331–344. doi:10.1016/S0034-4257(98)00010-8.

Wickham, J. D., S. V. Stehman, L. Gass, J. Dewitz, J. A. Fry, and T. G. Wade. 2013. "Accuracy Assessment of NLCD 2006 Land Cover and Impervious Surface." *Remote Sensing of Environment* 130: 294–304. doi:10.1016/j.rse.2012.12.001.

Zimmerman, P. L., I. W. Housman, C. H. Perry, R. A. Chastain, J. B. Webb, and M. V. Finco. 2013. "An Accuracy Assessment of Forest Disturbance Mapping in the Western Great Lakes." *Remote Sensing of Environment* 128: 176–185. doi:10.1016/j.rse.2012.09.017.