Classification Accuracy for Stratification with Remotely Sensed Data

Raymond L. Czaplewski and Paul L. Patterson

ABSTRACT. Tools are developed that help specify the classification accuracy required from remotely sensed data. These tools are applied during the planning stage of a sample survey that will use poststratification, prestratification with proportional allocation, or double sampling for stratification. Accuracy standards are developed in terms of an "error matrix," which is familiar to remote sensing specialists. In addition, guidance is provided to determine when new remotely sensed classifications are needed to maintain acceptable levels of statistical precision with stratification. For. Sci. 49(3):402–408.

Key Words: Forest inventory and monitoring, forest statistics.

Forest surveys often use a simple random or systematic sample of field plots to characterize forest conditions over extensive regions. Stratification with remotely sensed data can improve precision of statistical estimates for areal extent of different land uses and types of forest cover. For example, the USDA Forest Service's Forest Inventory and Analysis (FIA) Program measures 376,000 1 ha field plots, 125,000 of which are forested, to characterize the 300 million ha of forest and woodland ecosystems in the USA (Czaplewski 1999). In addition, FIA has used double-sampling for stratification with 9.4 million photo-interpreted plots in the first phase. Stratification with "wall-to-wall" Landsat satellite data can replace photo-interpretation for stratification.

The objective of this article is to suggest accuracy standards for remotely sensed classifications that will be used in prestratification with proportional allocation or post-stratification (Cochran 1977, p. 91, 134). Generalizations are developed that can be applied during the early stages of survey planning. The gain in statistical efficiency with stratification is approximated for a wide range of classification accuracies, using metrics that are familiar to the remote sensing community (e.g., user's and producer's accuracies). All that is required is a target for statistical precision and assumptions regarding the prevalence and rates of change for each category of land use or forest cover in the study area.

We begin by developing recommendations for two subpopulations and their corresponding two strata, such as forest and nonforest. These generalizations are then extended to three or more subpopulations and their corresponding strata. Recent changes in land cover or land use reduce accuracy of older remotely sensed data, and this degrades the gains in statistical efficiency with stratification. We develop methods that help determine when new remotely sensed data are needed to restore efficient stratification.

Statistical Estimators

In this section, we introduce the "error matrix" for remotely sensed classifications (e.g., Congalton 1991), and relate this to a statistical estimator that uses stratification. Next, the "design effect" is offered as a scalar measure of the gain in statistical efficiency achieved through a stratified sampling design. Objectives for gains in statistical efficiency will be expressed in terms of this design effect, and the design effect will be incorporated into the error matrix. This allows

Raymond L. Czaplewski, Project Leader, USDA Forest Service, Rocky Mountain Research Station, 2150 Centre Ave, Bldg. A, Suite 350, Fort Collins, CO 80526-1891—Phone: (970) 295-5973; Fax: (970) 295-5959; E-mail: rczaplewski@fs.fed.us. Paul L. Patterson, Project Scientist, USDA Forest Service, Rocky Mountain Research Station, 2150 Centre Ave, Bldg. A, Suite 350, Fort Collins, CO 80526-1891—Phone: (970) 295-5966; E-mail: plpatterson@fs.fed.us.

Acknowledgments: This research was funded by the USDA Forest Service, Forest Inventory and Analysis Program (http://fia.fs.fed.us/) and the Rocky Mountain Research Station (http://www.fs.fed.us/rm/ftcol/rwu4804.htm).

Manuscript received December 14, 2000, accepted March 5, 2002.

Joint probabilities

$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$	$P(A_1) = \sum_{j=1}^{m} P(A_1 \cap B_j)$
$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$	$P(A_2) = \sum_{j=1}^{m} P(A_2 \cap B_j)$
$P(B_1) = \sum_{i=1}^{m} P(A_i \cap B_1)$	$P(B_2) = \sum_{i=1}^{m} P(A_i \cap B_2)$	$\sum_{i=1}^{m} \sum_{j=1}^{m} P(A_i \cap B_j) = 1$

Conditional probabilities (accuracy) within strata

$P(A_1 \mid B_1) = \dfrac{P(A_1 \cap B_1)}{P(B_1)}$	$P(A_1 \mid B_2) = \dfrac{P(A_1 \cap B_2)}{P(B_2)}$
$P(A_2 \mid B_1) = \dfrac{P(A_2 \cap B_1)}{P(B_1)}$	$P(A_2 \mid B_2) = \dfrac{P(A_2 \cap B_2)}{P(B_2)}$
$\sum_{i=1}^{m} P(A_i \mid B_1) = 1$	$\sum_{i=1}^{m} P(A_i \mid B_2) = 1$

Figure 1. The "error matrix," or "confusion matrix," describes classification accuracy with remotely sensed data. The goal is estimation of the prevalence or size of each subpopulation, i.e., $P(A_i)$. Poststratification uses the distribution of subpopulation proportions in each stratum, i.e., $P(A_i|B_i)$, and the size of each stratum, i.e., $P(B_i)$ to improve statistical estimates of $P(A_i)$.

us to make generalizations in terms that are familiar to remote sensing specialists.

Error Matrix for Remote Sensing

Assume the population is subdivided into m subpopulations, such as forest and nonforest cover ($m = 2$). Each unit of the population is in one and only one subpopulation. Our goal is estimation of the proportion, $P(A_i)$, of each subpopulation A_i in the population. An estimated proportion is readily converted into a percentage or an area (e.g., number of hectares). Let $P(B_j)$ be the proportion of the total population in stratum B_j. Using the notation of a joint probability, $P(A_i \cap B_j)$ denotes the proportion of the population that is jointly within subpopulation A_i and remotely sensed stratum B_j. Using the notation of a conditional probability, $P(A_i|B_j)$ is the proportion of subpopulation A_i within remotely sensed stratum B_j, where $P(A_i \mid B_j) = P(A_i \cap B_j)/P(B_j)$. The remote sensing literature refers to $P(A_i \mid B_i)$ as "user's accuracy" for category i (Congalton 1991). Statistical notation for probabilities is used to avoid introduction of a new mathematical lexicon for proportions in the error matrix.

Figure 1 gives the mathematical notation for the error matrix when there are two subpopulations and two strata ($m = 2$). The sum of diagonal elements in the first matrix of Figure 1 is termed "overall accuracy," and each diagonal element in the second matrix is termed the "user's accuracy" by Congalton (1991).

Simple Random Sampling

With simple random sampling, Cochran (1977, p. 51–52) gives the estimated proportion of subpopulation A_i and its variance (see Cochran, eqs. 3.3 and 3.6, respectively). Because of the relative size of the sample (n) compared to the total population (N), the term $(N–n)/(N–1)$ can be ignored, yielding the following simplified approximation:

$$
\begin{aligned}
&\hat{P}_{SRS}(A_i) = \frac{1}{n}\sum_{a=1}^{n} x_a \\
&V_{SRS}\left[\hat{P}(A_i)\right] \cong \frac{1}{n} P(A_i)\left[1 - P(A_i)\right]
\end{aligned}
\tag{1}
$$

where

$$
x_a = \begin{cases} 1 & \text{if sample unit } a \text{ is in subpopulation } A_i \\ 0 & \text{otherwise} \end{cases}
$$
$$
n = \left\{ \text{the sample size, } n << N \right\}
$$

Stratification

Consider a simple random sample of field plots. Each field plot is assigned to one and only one remotely sensed stratum, and remote sensing measures the exact area of each stratum. Since the plots are assigned to the stratum after sampling, this is poststratified sampling. Cochran (1977, p. 107) gives the estimated proportion of subpopulation A_i in the total population. Poststratification complicates the variance estimator (see Cochran, eq. 5.A.42). However, ignoring terms of size $1/n^2$ and assuming $N>>n$ yields the following simplified approximation:

$$
\hat{P}_{STR}(A_i) = \sum_{j=1}^{m} P(B_j)\hat{P}(A_i \mid B_j)
$$

$$
\begin{aligned}
V_{STR}[\hat{P}(A_i)] &\cong \frac{1}{n}\sum_{j=1}^{m} P(B_j)S_j^2 \\
&= \frac{1}{n}\sum_{j=1}^{m} P(B_j)P(A_i \mid B_j)[1-P(A_i \mid B_j)]
\end{aligned} \tag{2}
$$

Design Effect

The improvement in statistical efficiency with stratification is typically quantified by the ratio of variances with and without stratification, which is defined as the "design effect" and denoted by k (Särndal et al. 1992). This scalar index is a performance measure that compares efficiency of a "complex" sampling design to a familiar benchmark, namely, simple random sampling. In our case, the design effect is approximated with Equations (1) and (2) as:

$$
\begin{aligned}
k &= \frac{V_{STR}\left[\hat{P}(A_i)\right]}{V_{SRS}\left[\hat{P}(A_i)\right]} \\
&= \frac{\sum_{j=1}^{m} P(B_j)P(A_i \mid B_j)[1-P(A_i \mid B_j)]}{P(A_i)[1-P(A_i)]}
\end{aligned} \tag{3}
$$

If stratification greatly improves the estimate (i.e., reduces the variance of the estimate compared to simple random sampling), then k will approach the value of 0. If stratification has little beneficial effect, then k will be nearly equal to 1. In the following sections, the design effect is used to further simplify the mathematics, reduce the number of *a priori* assumptions needed to construct an expected error matrix, and produce broad generalizations that are useful in survey planning.

Two Subpopulations

Czaplewski and Patterson (*in press*) rearrange Equation (3) for $m = 2$ into:

$$
1-k = \left[\frac{P(A_1 \mid B_1)-P(A_1)}{1-P(A_1)}\right]\left[\frac{P(A_2 \mid B_2)-P(A_2)}{1-P(A_2)}\right] \tag{4}
$$

and define

$$
\frac{P(A_i \mid B_i)-P(A_i)}{1-P(A_1)} = \begin{cases}\text{relative classification accuracy} \\ \text{for subpopulation } i\end{cases}
$$

If the relative accuracies in Equation (4) are identical for both strata, then:

$$
\frac{P(A_1 \mid B_1)-P(A_1)}{1-P(A_1)} = \frac{P(A_2 \mid B_1)-P(A_2)}{1-P(A_2)} = \sqrt{1-k}
$$

and

$$
P(A_i \mid B_i) = \sqrt{1-k}\left[1-P(A_i)\right]+P(A_i) \tag{5}
$$

We assume $P(A_i \mid B_i) > P(A_i)$ to assure that the design effect remains the positive ratio of two variances.

The "relative classification accuracy" in Equation (4) helps develop the generalizations that follow. For example, Czaplewski and Patterson (2001) show that $P(A_i) = P(B_i)$ when relative accuracies are identical in both strata. This produces a symmetry in the error matrix that simplifies the algebra and reduces the number of assumptions. Furthermore, this symmetry forces "user's accuracy" to equal "producer's accuracy" within each category, which reduces the number of metrics that must be considered by remote sensing specialists when setting performance standards for classification accuracy. Figure 2, which is a special case of the error matrix in Figure 1, incorporates relative classification accuracies that are identical in both strata [Equation (5)].

Required Accuracy

Table 1 defines five different levels of design effect k and the corresponding degree of statistical efficiency at each level of k. Table 2 gives the classification accuracy required to achieve these various levels of gain, assuming the error matrix has the structure given in Figure 2.

For example, assume that the study area is covered with 30% forest and 70% nonforest. Assume that a "substantial" gain in statistical efficiency is required to justify the extra expense of stratification with remotely sensed data. Using Table 1, these assumptions require that $k = 0.50$. Given these assumptions and specifications, and using Table 2, we recommend that classification accuracy for the remotely sensed "map" must be at least 79% for forest cover and 91% for nonforest cover. The recommended accuracies in Table 2 simultaneously apply to both user's and producer's accuracy (Congalton 1991) because they

$P(A_1 \cap B_1) = P(A_1)\left[P(A_1)+P(A_2)\sqrt{1-k}\right]$	$P(A_1 \cap B_2) = P(A_1)P(A_2)\left[1-\sqrt{1-k}\right]$	$P(A_1)$
$P(A_2 \cap B_1) = P(A_1)P(A_2)\left[1-\sqrt{1-k}\right]$	$P(A_2 \cap B_2) = P(A_2)\left[P(A_2)+P(A_1)\sqrt{1-k}\right]$	$P(A_2)$
$P(B_1) = P(A_1)$	$P(B_2) = P(A_2)$	1

Figure 2. Special case of an error matrix for two subpopulations and their corresponding strata when the relative classification accuracy in each stratum is identical [Equation (5)]. These assumptions permit generalizations that are useful in setting accuracy standards for remotely sensed classifications (e.g., Table 2).

Table 1. Definition of five levels of gain in statistical efficiency from stratification.

Design effect $k = V_{STR}/V_{SRS}$	Gain in efficiency through poststratification	Increase in effective number of plots* gained through stratification	Relative variance of stratified sampling compared to simple random sampling $100 \times V_{STR} / V_{SRS}$	Relative standard error[†] of stratified sampling compared to simple random sampling $100 \times \sqrt{V_{STR} / V_{SRS}}$
			(%)	
$k = 1.00$	No Gain	None	100	100
$k = 0.83$	Minimal gain	1.2-fold	83	91
$k = 0.67$	Moderate gain	1.5-fold	67	82
$k = 0.50$	Substantial gain	2-fold	50	71
$k = 0.25$	Excellent gain	4-fold	25	50

* The increase in sample size *n* that would be required to achieve the same variance without stratification.

† Approximately proportional to the confidence interval.

are based on the special case of the symmetrical error matrix in Figure 2. If these accuracy standards are met, then Table 1 indicates that there would be a two-fold increase required in the number of plots in a simple random sample to achieve this same level of statistical efficiency without remotely sensed data; stratification with remote sensing would yield an expected 50% decrease in estimation variance relative to simple random sampling; and the confidence interval for the estimate of forest area would be only 71% as wide as the interval without remotely sensed data.

Three or More Subpopulations

The number of cells in an error matrix increases geometrically with the number of remotely sensed strata and their corresponding subpopulations. Therefore, further assumptions are needed to extend generalizations to more complex classification systems. Later, we will discuss an alternative to these assumptions.

Consider two subpopulations that have the error matrix in Figure 2. Next, form a total of three subpopulations by subdividing subpopulation A_1 into two parts. The size of the first part is $\tau P(A_1)$, $0 < \tau < 1$, and the size of the second part is $(1 - \tau)P(A_1)$. For example, start with forest and nonforest subpopulations. Then, subdivide forest into hardwood and softwood stands. Assume that the relative classification accuracies [Equation (5)] in each subdivision remain equal. Under this assumption, absolute accuracies decrease in the hardwood and softwood strata because their sizes $\tau P(A_1)$ and $(1 - \tau)P(A_1)$ are smaller than the total size of entire forested subpopulation $P(A_1)$, as shown in Table 2. Assume the size of each stratum equals the size of its corresponding subpopulation for each of the three categories (similar to the error matrix in Figure 2). Finally, assume that misclassification errors between forest and nonforest are independent of whether or not the forest is a hardwood or softwood stand. Figure 3 gives the 3×3 error matrix under these assumptions.

Next, regroup the softwood forest category with the nonforest category, leaving two subpopulations: hardwood forest and nonhardwood forest. The resulting 2×2 error matrix retains the same structure as the error matrix in Figure 2, where the relative classification accuracies for hardwood and nonhardwood forest equal those for the original forest and nonforest categories. Previous generalizations, which apply to two subpopulations, now apply to the hardwood forest category. Alternatively, hardwood forest could be grouped with nonforest to produce two categories: softwood forest and nonsoftwood forest. The resulting error matrix also has the same structure as Figure 2. The sequential bifurcation of subpopulations can proceed many times to form numerous categories, thus producing a "binary-tree" classification system. Therefore, the generalizations for two subpopulations in Table 2 apply to any single category under these assumptions.

As the number of subpopulations increase, the size $P(A_i)$ of each subpopulation tends to become smaller, and the accuracy $P(A_i \mid B_i)$ required for a given design effect decreases (see Table 2). Assume the prevalence of each of the *m* subpopulations is identical, i.e., $P(A_i)=1/m$; all strata have the same relative classification accuracy [Equation (5)]; and

Table 2. User's and producer's accuracies required from remotely sensed classifications to achieve given gains in statistical efficiency with stratification.

Prevalence of subpopulation in study area $P(A_i)$	Excellent gain k = 0.25	Substantial gain k = 0.50	Moderate gain k = 0.67	Minimal gain k = 0.83	No gain k = 1.00
			(%)		
1	87	71	58	41	1
5	87	72	60	44	5
10	88	74	62	47	10
20	89	77	66	53	20
30	91	79	70	59	30
40	92	82	75	64	40
50	93	85	79	70	50
60	95	88	83	76	60
70	96	91	87	82	70
80	97	94	92	88	80
90	99	97	96	94	90
95	99	99	98	97	95
99	100	100	100	99	99

$\tau P(A_1)\left\{\begin{matrix}\tau P(A_1)\\+\left[1-\tau P(A_1)\right]\sqrt{1-k}\end{matrix}\right\}$	$\tau(1-\tau)P(A_1)^2\left[1-\sqrt{1-k}\right]$	$\tau P(A_1)P(A_2)\left[1-\sqrt{1-k}\right]$	$\tau P(A_1)$
$\tau(1-\tau)P(A_1)^2\left[1-\sqrt{1-k}\right]$	$(1-\tau)P(A_1)\left\{\begin{matrix}(1-\tau)P(A_1)\\+\left[1-(1-\tau)P(A_1)\right]\sqrt{1-k}\end{matrix}\right\}$	$(1-\tau)P(A_1)P(A_2)\left[1-\sqrt{1-k}\right]$	$(1-\tau)P(A_1)$
$\tau P(A_1)P(A_2)\left[1-\sqrt{1-k}\right]$	$(1-\tau)P(A_1)P(A_2)\left[1-\sqrt{1-k}\right]$	$P(A_2)\left\{\begin{matrix}P(A_2)\\+\left[1-P(A_2)\right]\sqrt{1-k}\end{matrix}\right\}$	$P(A_2)$
$\tau P(B_1)=\tau P(A_1)$	$(1-\tau)P(B_1)=(1-\tau)P(A_1)$	$P(B_2)=P(A_2)$	1

Figure 3. Error matrix for three subpopulations and corresponding strata, assuming relative classification accuracies are the same for all strata, the stratum sizes equal the subpopulations sizes, and other assumptions. The first two categories were originally grouped into a single category in Figure 2.

a remotely sensed stratum corresponds to each subpopulation (i.e., an $m \times m$ error matrix). Figure 4 gives the resulting accuracy standards for given gains in statistical efficiency as a function of the number of categories (m) in the classification system. When there are ten or more categories, classification accuracy in remotely sensed strata should exceed 40% to attain even minimal gains in statistical efficiency. Substantial gains require accuracies exceeding 70% in each stratum. We offer Table 2 and Figure 4 as approximate guidance for any classification system, although our derivations strictly apply to binary-tree systems with the given assumptions.

Change in Efficiency Over Time

Landscapes change over time through shifts in land use, land management, succession, and disturbance. Field plots at time t may be stratified using remotely sensed data that were acquired many years in the past (i.e., time 0). Some portion of the total "classification error" is caused by changes in the landscape between time 0 and t, and not by the accuracy of the remotely sensed classifications at time 0. To explore the question "How old can the remotely sensed data become before their value for stratification becomes seriously degraded?," we introduce a transition model between two subpopulations.

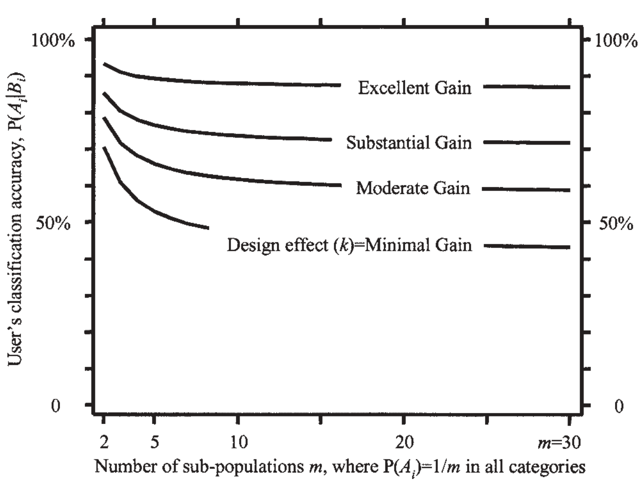

Figure 4. Accuracy standards at different levels of classification detail and given assumptions.

Assume subpopulation size $P(A_1)_t$ at time t equals some fraction φ_1 of its size at time 0, where $0 < \varphi_1 < 1$. The transition rate φ_1 spans an unspecified time interval t, where exact interval length must be specified for each unique application. The corresponding transition rate from subpopulation A_2 to A_1 is φ_2. Since the remotely sensed data were acquired only at time 0, the size of each stratum, $P(B_i)$, remains unchanged between times 0 and t. Assume the transition rates for every subpopulation are independent of the remotely sensed strata, and both strata share the same initial design effect k_0 at time 0. Figure 5 gives the error matrix under these assumptions. Define c as the relative loss in statistical efficiency caused by transitions among the two subpopulations, where the design effect at time t is $k_t = c(1 - k_0) + k_0$. This definition facilitates generalizations, as shown in the next paragraph.

There is a broad range of possible transition rates in various landscapes. Table 3 considers a few special cases from which useful generalizations are possible. First, assume losses occur from subpopulation $A_1(\varphi_1 > 0)$, but not from subpopulation $A_2(\varphi_2 = 0)$. For example, let subpopulation A_1 represent forestland use, and the stratification is based on 10-yr-old remotely sensed data. Assume 1% per year of the forest is converted to another land use, such as agriculture or urban, and the rate of change over 10 yr is $\varphi_1 \approx 0.10$. Assume there are no conversions from urban or agricultural uses back to forestland use ($\varphi_2 = 0$). If the original forestland occupies 60% of the landscape, then the stratification based on 10-yr-old remotely sensed data loses only c = 22% of its gain in statistical efficiency (Table 3). If the original design effect were $k_0 = 0.50$, the design effect at time t would be $k_t = 0.22(1 - 0.50) + 0.50 = 0.61$. With this definition of c, Table 3 can quantify the relative loss in statistical efficiency without specifying the exact value of the original design effect k_0.

Although it might be counter-intuitive, the loss in efficiency over time in a steady-state landscape is approximately twice as fast as this first example. Assume the loss from subpopulation A_1 is exactly offset by the reverse transition from subpopulation A_2, i.e., $P(A_1)\ \varphi_1 = P(A_2)\ \varphi_2$ for the error matrix in Figure 5. In this case, the net change in the size of each subpopulation is zero. However, classification accuracy decreases over time because some sites that were correctly

$P(A_1 \cap B_1)_t = \begin{bmatrix} (1-\varphi_1)P(A_1 \cap B_1)_0 \\ +\varphi_2 P(A_2 \cap B_1)_0 \end{bmatrix}$	$P(A_1 \cap B_2)_t = \begin{bmatrix} (1-\varphi_1)P(A_1 \cap B_2)_0 \\ +\varphi_2 P(A_2 \cap B_2)_0 \end{bmatrix}$	$P(A_1)_t = \begin{bmatrix} (1-\varphi_1)P(A_1)_0 \\ +\varphi_2 P(A_2)_0 \end{bmatrix}$
$P(A_2 \cap B_1)_t = \begin{bmatrix} \varphi_1 P(A_1 \cap B_1)_0 \\ +(1-\varphi_2)P(A_2 \cap B_1)_0 \end{bmatrix}$	$P(A_2 \cap B_2)_t = \begin{bmatrix} \varphi_1 P(A_1 \cap B_2)_0 \\ +(1-\varphi_2)P(A_2 \cap B_2)_0 \end{bmatrix}$	$P(A_2)_t = \begin{bmatrix} \varphi_1 P(A_1)_0 \\ +(1-\varphi_2)P(A_2)_0 \end{bmatrix}$
$P(B_1)_t = P(B_1)_0 = P(A_1)_0$	$P(B_1)_t = P(B_1)_0 = P(A_1)_0$	1

Figure 5. The error matrix from Figure 2 that includes transition rates (φ_i) from subpopulation A_i to A_j between time 0 and t, where remotely sensed data, which are used to specify strata B_1 and B are acquired at time $t = 0$.

classified at time 0 have been converted to a different category at time t. For example, subpopulation A_1 is forest cover in a steady-state landscape, where clearcutting is the sole harvesting practice, the average rotation age is 30 yr, and clearcuts return to forest cover after 3 yr. At any point in time, approximately 3/30 = 10% of all forested lands exists as clearcuts, with no detectable forest cover. Therefore, 30% of the forested acreage at time 0 will be nonstocked clearcuts at time t, i.e., $\varphi_1 = 0.10$, and an equal number of acres in clearcuts at time 0 have regenerated into forest cover at time t, i.e., $P(A_1)\cdot_1 = P(A_2)\ \varphi_2$. Assume the total forest cover is 60% of the landscape. From Table 3, the estimated loss in statistical efficiency is c = 44%.

Discussion

Remote sensing specialists frequently ask questions such as "What accuracy is required from remotely sensed products?" and "When do remote sensed images become so old as to lose their value?" The answer depends on the specific application; therefore, users of the remotely sensed products have the responsibility to provide the answer. When the application is stratification to improve statistical efficiency, then we can provide some generalities and "rules of thumb" that help provide useful answers.

Consider the following example. The user needs estimates of area for each of 30 types of forest conditions and nonforest cover types. These 30 categories can be accurately identified on field plots, but this is expensive. A stratified sampling strategy could reduce costs. Remote sensing could produce a map of these 30 categories, and each category in the map could be a stratum for statistical estimates. Unfortunately, classification accuracy is expected to be 40% or less for this detailed level of thematic resolution. Based on Figure 4 with m = 30, a user's accuracy of 40% is not expected to produce noticeable gain in statistical efficiency. However, the user believes that statistical estimates for a few major categories would remain useful, and the cost of remote sensing would be justified if the resulting stratification yields a "moderate" gain in statistical efficiency (i.e., k = 0.67 from Table 1). The user guesses that the 1,000,000 ha study area has approximately the following distribution of broad cover types: 600,000 ha of nonforest (60%), 200,000 ha of hardwood forest (20%), 150,000 ha of mixed forest (15%), and 50,000 ha of softwood forest (5%). Using these guesses and Table 2 (k = 0.67), the user specifies the following standards for both user's and producer's accuracies with remote sensing: 83% for nonforest, 66% for hardwood forest, 64% for mixed forest, and 60% softwood forest. Several months later, the remotely sensed map is delivered, and it is used for stratification. The map accuracy surpasses the original standards set by the user. At this final stage of the project, there are sufficient data to use the precise estimators in Cochran (1977) for stratified designs to produce the required statistical estimates. Ten years later, new statistical estimates are needed, but the value of the original remotely sensed map for stratification is unknown. Of the original 400,000 ha of forest cover, the user guesses that 40,000 ha have been converted to other land uses (φ_1 = 0.10) during those 10 yr. The user further assumes that there has been no reversion of non-forest lands back to forest (φ_2 = 0), and the transition among different forest types has been negligible. Using Table 3, the user decides that the loss in efficiency with the original remotely sensed map will be about 16%, and the cost of developing a new remotely sensed map for stratification is not justified. Then, field crews

Table 3. Relative loss of statistical efficiency (c) caused by changes in the landscape since acquisition of remotely sensed imagery using error matrix in Figure 5.

Prevalence of subpopulation at time 0	Transition rate from subpopulation 1 into subpopulation 2 over time t, where the reverse conversion rate is zero (φ_2 = 0%)				Steady-state transition rates between subpopulations 1 and 2 over time t, where $\varphi P(A_1) = \varphi P(A_2)$			
	φ = 1%	φ = 2%	φ = 5%	φ =10%	φ =1%	φ =2%	φ =5%	φ =10%
				(%)				
1	1	2	5	10	2	4	10	19
5	1	2	5	10	2	4	10	20
10	1	2	6	11	2	4	11	21
20	1	2	6	12	2	5	12	23
30	1	3	7	14	3	6	14	27
40	2	3	8	16	3	7	16	31
50	2	4	10	18	4	8	19	36
60	2	5	12	22	5	10	23	44
70	3	6	15	27	7	13	31	56
80	5	9	21	36	10	19	44	75
90	9	17	34	53	19	36	75	100
95	17	29	51	69	36	64	100	100

reclassify the sample of field plots, the original remotely sensed map is reused for stratification, and new statistics are generated using estimators from Cochran (1977).

In this example, the user makes certain educated guesses during the planning stage of a project (e.g., the approximate prevalence of each land cover category). These guesses are then used with our tools, which are based on additional assumptions (e.g., identical relative accuracies). As an alternative in the above example, the user could forgo the assumptions that underlie our tools and predict each user's accuracy in the 4x4 error matrix. These 16 additional guesses could be put directly into equations in Cochran (1977) to predict sampling errors with stratification. The user must weigh the risk of this uncertain information compared to the cost of better information on which planning decisions are made.

We have not made careful distinctions among different stratified sampling designs. The approximations used in Equation (2) obviate the differences between poststratification and prestratification with proportional allocation. If there is sufficient sampling intensity at the first phase of double sampling for stratification, then the effects of sampling error at that phase are relatively negligible, and recommendations for stratified sampling will be approximately correct. Therefore, we offer these tools as a reasonable basis for accuracy standards, regardless of the specific stratification strategy.

Table 2 shows a strong dependence between accuracy and prevalence. If a subpopulation is common, then the classification accuracy must be very high. If the subpopulation is rare, then less accuracy is sufficient. This pattern also appears in Table 3, where changes in a common category over time quickly degrade effectiveness of old remotely sensed data. For example, estimates are needed for the area of rare old growth forest, which might cover only 5% of a study area. In order to achieve "excellent" gains through stratification, Table 2 indicates that accuracy for old growth need only be 87%, whereas accuracy for the remaining landscape must be at least 99%. Therefore, the remote sensing specialist should strive for perfect classification accuracy in the non-old-growth stratum, and accept less accuracy in the old-growth stratum. This stratification, when used with the field sample, will produce unbiased estimates of old-growth area, even though the accuracy of remote sensing for the old-growth stratum is less than the other stratum.

The process for setting objective *a priori* accuracy standards for remotely sensed products is difficult. When the application is stratification for statistical efficiency, the tools presented here are presented as a potential improvement over current ad hoc techniques.

Literature Cited

Cochran, W.G. 1977. Sampling techniques. Wiley, New York. 428 p.

Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Rem. Sens. Environ. 37:35–46.

Czaplewski, R.L. 1999. Multistage remote sensing: Toward an annual national inventory. J. For. 97:44–48.

Czaplewski, R.L., and P.L. Patterson. In press. Accuracy of remotely sensed classifications for stratification of forest and nonforest. P. 32 *in* Proc. of the Sec. Annu. For. Inventory and Anal. Symp., Reams, G.A., et al. (eds.). For. Serv. Gen Tech. Rep. SRS-GTR-47.

Särndal, C.E., B. Swensson, and J. Wretman. 1992. Model assisted survey sampling. Springer-Verlag, New York. 694 p.