Remote Sensing of Environment 168 (2015) 126–133

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

# Optimizing sample size allocation to strata for estimating area and map accuracy

John E. Wagner, Stephen V. Stehman *

*Department of Forest and Natural Resources Management, State University of New York, College of Environmental Science & Forestry, 1 Forestry Drive, Syracuse, NY 13210, United States*

ARTICLE INFO

*Article history:*
Received 17 November 2014
Received in revised form 19 June 2015
Accepted 30 June 2015
Available online 14 July 2015

*Keywords:*
Least-cost/cost effective model
Accuracy assessment
Stratified sampling
User's accuracy
Producer's accuracy

ABSTRACT

The results of a map accuracy assessment are often summarized by reporting user's accuracy and producer's accuracy for each class in the map legend. Additionally estimating the proportion of area of each class based on the best assessment of ground condition (i.e., the reference classification) of the locations selected in the sample is often of interest for monitoring status and change in land cover. Stratified random sampling is a commonly used sampling design for accuracy assessment, and an important decision for this design is the allocation of sample size to the strata. In this article, the allocation that minimizes the sum of the variances of the estimators of user's accuracy, producer's accuracy, and area of a single targeted class for a fixed total sample size is derived for stratified random sampling. For example, the targeted class might be a rare land-cover type such as wetland or in the case of a land-cover change assessment forest loss. An Excel sample allocation calculator implements the optimization and two examples illustrate the application. Practitioners can apply these optimization results to guide sample size allocation decisions when using a stratified random sampling design for accuracy assessment and area estimation.



## 1. Introduction

Map accuracy assessments are implemented to evaluate the quality of a map based on a sample of higher quality information (i.e., the reference classification) than used to create the map (Stehman & Foody, 2009). The outcome of an accuracy assessment is an error matrix (Table 1) from which various accuracy measures can be estimated. In the population error matrix shown, the rows represent the map classification, the columns represent the reference classification, and the cell entries of the error matrix ($p_{ij}$) represent the proportion of area in which the map classification is class $i$ and the reference classification is class $j$.

Overall accuracy is the sum of the diagonal entries of the Table 1 error matrix. Class-specific accuracy of class $k$ is typically reported as user's accuracy, defined as $p_{kk}/p_{k+}$, and producer's accuracy, defined as $p_{kk}/p_{+k}$. User's accuracy is the complement of commission error, and producer's accuracy is the complement of omission error. Another important use of the sample data and reference classification is to estimate the proportion of area (with associated confidence interval) of reference class $j$, denoted as $p_{+j}$, where $p_{+j}$ is the sum of the proportions in column $j$ of the error matrix (e.g., Olofsson, Foody, Stehman, & Woodcock, 2013; Olofsson et al., 2014; Stehman, 2013).

Stratified random sampling, in which a simple random sample is selected within each stratum, is recommended as a good practice option for map accuracy assessment and area estimation (Olofsson et al., 2014). Stratified random sampling is typically implemented with the map classes defined as the strata (i.e., each pixel is assigned to the stratum of the class to which it is mapped). The sample size allocated to each stratum is an important decision, and the allocation will depend on the objectives specified for the accuracy assessment. In this article we derive an optimal sample allocation to strata when the objectives of the accuracy assessment are to estimate user's accuracy, producer's accuracy, and proportion of area of a single targeted class of primary interest. Optimizing the sample allocation for the accuracy and area estimates of all classes is not addressed. The optimal allocation to strata is chosen to minimize the sum of the variances of the three estimators given a fixed total sample size. Minimizing the sum of the variances circumvents the problem that optimizing the allocation for each of the three estimates separately would yield three different sample allocations. The optimization is implemented via an Excel sample allocation calculator that is available from the authors.

The applications addressed by our sample allocation optimization represent situations in which the focus is on one class of priority interest. That is, we assume that the class listed as row $i = 1$ and column $j = 1$ in the error matrix (Table 1) is the primary class of interest, and the optimal allocation of the total sample to the strata is constructed to minimize the sum of the variances of the estimators of user's accuracy of map class 1, producer's accuracy of reference class 1, and proportion of area of reference class 1. An example of the applications we envision is a study in which forest cover loss is the targeted class of interest. The accuracy assessment objectives would then focus on estimating user's and producer's accuracies of forest cover loss, and the proportion of

* Corresponding author.
*E-mail addresses:* jewagner@esf.edu (J.E. Wagner), svstehma@syr.edu (S.V. Stehman).

http://dx.doi.org/10.1016/j.rse.2015.06.027
0034-4257/

**Table 1**
Population error matrix for a classification with $D$ classes, where the rows ($i$) represent the map classification and the columns ($j$) represent the reference classification; $p_{ij}$ is the population proportion of area with map class $i$ and reference class $j$. The row ($p_{i+}$) and column ($p_{+j}$) marginal totals are the sum of the $p_{ij}$ values in each row and column.

| | | Reference class | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | … | $k$ | … | $D$ | Total |
| Map class | 1 | $p_{11}$ | $p_{12}$ | … | $p_{1k}$ | … | $p_{1D}$ | $p_{1+}$ |
| | 2 | $p_{21}$ | $p_{22}$ | … | $p_{2k}$ | … | $p_{2D}$ | $p_{2+}$ |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | $k$ | $p_{k1}$ | $p_{k2}$ | … | $p_{kk}$ | … | $p_{kD}$ | $p_{k+}$ |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | $D$ | $p_{D1}$ | $p_{D2}$ | … | $P_{Dk}$ | … | $P_{DD}$ | $P_{D+}$ |
| | Total | $p_{+1}$ | $p_{+2}$ | … | $p_{+k}$ | … | $p_{+D}$ | 1 |

area of forest loss (based on the reference classification). An example of a map-based stratification for this application could be forest loss, forest no change, and non-forest (e.g., Olofsson et al., 2013, Table 3).

### 1.1. A review of sample size and allocation recommendations for accuracy assessment

Sample size planning has a long history in applications of accuracy assessment to remote sensing derived products. The early sample size planning publications in remote sensing focused on deciding the sample size $n$ that would achieve a specified standard error, width of confidence interval, or Type I error probability of a test regarding overall accuracy of the map (Hord & Brooner, 1976; Rosenfield, Fitzpatrick-Lins, & Ling, 1982; Stehman, 2001, Figure 1; Van Genderen, Lock, & Vass, 1978). These early publications typically assumed the design was simple random sampling. For stratified sampling, Hay (1979) and Congalton (1991) suggested a rule of thumb of a minimum sample size of 50 per class, with Congalton (1991) further recommending that this minimum sample size be increased to 75 or 100 if the classification has a large number (>12) of vegetation or land use categories. This rule of thumb appears to apply primarily to the objective of estimating user's accuracy of each class as the minimum sample size guidelines are not based on a formal optimization strategy for an explicit set of estimates.

For stratified random sampling, optimization algorithms have been developed for the allocation of the sample to strata. If the objective is to estimate overall accuracy or the proportion of area of reference class $j$ ($p_{+j}$), optimal allocation formulas can be used (Cochran, 1977, p. 108), where the optimization is constructed to minimize the variance of the estimator of overall accuracy or the estimator of $p_{+j}$. Stehman (2012) investigated various allocation options for the special case of a $2 \times 2$ error matrix (i.e., $D = 2$ classes in Table 1, for example, forest and non-forest, or change and no change classifications) when the priority objectives are specified as estimating overall accuracy, user's accuracy, and area of a targeted class. For a $2 \times 2$ error matrix, optimal allocation for estimating area of reference class $j$ is the same as the optimal allocation for estimating overall accuracy, so this effectively simplified the optimization to two criteria. Stehman (2012) evaluated power allocation (Bankier, 1988) which is a sample size allocation designed to optimize simultaneously estimates of stratum-specific parameters (e.g., user's accuracy in this application) and overall accuracy. The method has the name "power allocation" because it uses a constant "$a$" (called the power of the allocation) that is specified by the user (Särndal et al., 1992, pp. 470–471). Olofsson et al. (2014, Section 5.1.1) suggested allocating a specified minimum sample size to each stratum (e.g., 50 to 100 per stratum, depending on the total sample size) with the remainder of the sample then allocated proportional to the size of each stratum (see Olofsson et al., 2014, Table 5). The recommendation of 50 to 100 per class addresses the objective of estimating user's accuracy for each class and the recommendation to proportionally allocate the remainder of the sample addresses the objective of estimating overall accuracy and area of reference class $j$. The allocation presented in Olofsson et al. (2014) is not based on a formal mathematical optimization.

The optimal allocation we implement in this article differs from previous studies in that the optimization targets three specific estimates of primary interest, user's accuracy, producer's accuracy, and proportion of area for a single high priority class. The question addressed is not what total sample size $n$ to use, but how to allocate the $n$ sample units to strata to minimize the sum of the variances of the three estimates. In addition, our approach differs from much of the previous sample allocation work in accuracy assessment in that we implement a formal mathematical optimization to derive the optimal allocation.

## 2. Methods

The methods will be divided into three sections: i) $2 \times 2$ error matrix optimization, ii) generalized $D \times D$ error matrix optimization, and iii) optimal solution searching algorithm. The optimal allocation algorithm requires specification of the cell entries of a population error matrix (Table 1). The population values of these proportions are unknown in any given application, so the optimal allocation results used to decide the sample allocation must be based on hypothesized proportions for the population error matrix (i.e., in practice we specify $p_{ij}$ as best we can to reflect the accuracy of the map that will be produced). It is reasonable to evaluate a variety of such population error matrices to examine the sensitivity of the optimal allocation. If the optimal allocation does not substantially vary for these different hypothetical population error matrices, then we have stronger assurance that the allocation will likely be effective. If on the other hand the optimal allocation varies considerably over the range of population error matrices considered plausible for the specific application, then we are in the less enviable position of having to make the best guess of which error matrix most likely reflects the accuracy of the map to be evaluated and choosing the allocation derived from that matrix. If the error matrix actually associated with the particular map being assessed turns out to differ from the hypothetical population error matrix used to generate the optimal allocation, then the expected precision gain achieved by optimal allocation will be diminished.

While the optimal sample allocation for a $2 \times 2$ error matrix can be determined using a quadratic equation (see Section 2.1), for any error matrix greater than a $2 \times 2$, a searching algorithm is needed. The quadratic equation solution for the $2 \times 2$ error matrix provides a concise illustration of interactions among the $p_{ij}$'s and how they influence the optimal sample sizes. These interactions are more difficult to illustrate in a general $D \times D$ case. Readers not interested in the technical derivation of the optimization results may proceed to Section 3 describing the Excel sample allocation calculator.

### 2.1. 2 × 2 error matrix

We begin with the simplest case which is a $2 \times 2$ error matrix. The problem is to choose sample sizes $n_1$ and $n_2$ allocated to stratum 1 and stratum 2 (given a fixed total sample size, $n$) to minimize simultaneously the sum of the three variances of interest, which are the variances of the estimators of user's accuracy and producer's accuracy of class 1, and the proportion of area of class 1. The cell entries of the population error matrix must be specified for Table 1 with $D = 2$ classes.

For the $2 \times 2$ error matrix, the three variances for stratified random sampling that are included in the minimization are given by $VAR1$ for user's accuracy, $VAR2$ for producer's accuracy, and $VAR3$ for the proportion of area of class 1:

$$VAR1 = \frac{p_{11}p_{12}}{(p_{11}+p_{12})^2} \cdot \frac{1}{n_1} = [v_{11}] \cdot \frac{1}{n_1} \quad (1)$$