A Simulation Comparison of Three Marginal Area Estimators for Image Classification

Ding Yuan

Abstract
Obtaining correct area estimations for different land-cover or environmental categories is one of the main objectives of environmental applications of statistics. Area estimations are often obtained through classifying surveyed or remotely sensed data along with necessary adjustment. In this paper, three marginal area estimators, including the direct estimator, inverse estimator, and additive estimator for image classification, were compared using Monte Carlo simulations. The results suggested that, under minimum constraints for the acceptable image classifier and under our simulation environment, (1) both inverse and direct estimators were asymptotically unbiased and asymptotically zero-dispersed as sampling fraction increased; (2) the direct estimator normally has a smaller bias than the inverse estimator, but the inverse estimator normally had smaller dispersion than the direct estimator when the sampling fraction was small; and (3) the additive estimator was not asymptotically unbiased and was competitive with the other two methods only when sampling fraction and number of classes were both small. Simulated feasible regions for the three marginal area estimators are presented in this paper.

Introduction
Image classifications are commonly used for area estimations of land-cover types. Mapped or direct count of classified pixels, as a marginal area estimator, often contains components of bias, or classification errors. The true classifier accuracy is represented by the population confusion matrix, which is estimated by the sample confusion matrix. It is often necessary to make calibrations on the direct counts in order to obtain better estimates for the marginal areas (Dymond, 1993; Stohr *et al.*, 1994; Schriever and Congalton, 1995). The calibrations of the marginal area estimates are based on the utilization of the sample confusion matrices. Three distinctive calibration methods, the direct correction method (Card, 1982), the inverse correction method (Bauer *et al.*, 1978), and the additive correction method (Dymond, 1992), had been introduced into image classification. The objective of this research is to evaluate the application feasibility of the three marginal area correction techniques under reasonable constraints for classifiers using simulation methods.

The inverse estimator (Bauer *et al.*, 1978) makes use of the sample omission probabilities or frequencies to estimate the corresponding population omission (*a priori*) probabilities, and was named the "classic correction method" by Czaplewski and Catts (1990; 1992). Because the computation of the estimates of this method involves the inverting of the sample omission probability matrix, it is therefore called the inverse correction method by researchers and in this paper.

Desert Research Institute, University and Community System of Nevada, P.O. Box 19040, Las Vegas, NV 89132-0040. The author is currently with Lockheed Stennis Operations, Building 1110, Stennis Space Center, MS 39529.

The direct estimator of the marginal areas has been suggested by Card (1982) for image classification. It has been theoretically compared with the inverse estimator (Jupp, 1989). Using the asymptotic multinomial distribution, Card (1982) proved that the direct estimator is an asymptotic maximum-likelihood estimator for the marginal areas. The direct estimator uses the sample commission (*a posteriori*) probability matrix to estimate the population commission probability matrix, and therefore it was called the "inverse method" by Czaplewski and Catts (1992). One advantage of the direct estimator is that it does not involve matrix inversion in computation.

The inverse estimator had been widely used in remote sensing applications and has been frequently reported as giving satisfactory results (Maxim *et al.*, 1981; Prisley and Smith, 1987; Hay, 1988). However, in recent years, it has invoked strong criticisms for its instability when the sample omission probability matrix becomes singular or near-singular (Jupp, 1989). In a recent simulation study by Czaplewski and Catts (1992), the direct estimator has been shown to be systematic superior to the inverse estimator under no constraint for the simulated sample confusion matrices. Though the study by Czaplewski and Catts (1992) demonstrated that the inverse correction is inferior if the sample confusion matrix is random, it did not provide an explanation for the contrasting fact that the inverse correction had been practically acceptable in most application cases.

The apparently simplest method is the newly proposed additive method by Dymond (1992). The basic idea for this method is that the true ground area of a class can be estimated by the area classified by the classifier plus a single correction term. The additive correction method retains the asymptotic unbiased property of the correction (but not absolutely unbiased, as it was shown in our simulation). The variance of the estimation depends on the difference of the true ground area and the classified area. However, no work comparing the additive correction with other methods had been found by the author. To compare the additive correction method with the other two methods was another objective of this research.

In a recent paper, Yuan (1996) introduced the concepts of minimum practically acceptable classifier and reasonably acceptable classifier. Yuan's work theoretically proved that the inverse correction exists when the classifier is minimum practically acceptable, and the solution is stable if the classifier is also reasonably acceptable. However, no comparison with other correction methods had been made in that paper under the condition of acceptable classifiers.

Photogrammetric Engineering & Remote Sensing,
Vol. 63, No. 4, April 1997, pp. 385–392.

0099-1112/97/6304-385$3.00/0

For the purpose of comparison and demonstration, a simple proportional estimator or simple expansion was also simulated in the study. Simple expansion assumes that the true area of a ground class is strictly proportional to the area of the same class in the sample verification region. The simple proportional estimator is an unbiased but not absolutely unbiased estimator for the true ground areas. The variance of the estimation depends on the true area (proportion) of the given class in the sample verification region, and normally this variance is too high to be acceptable as demonstrated by Cochran (1977, pp. 21–22).

Preliminaries

Let r denote as number of the classes for our consideration. Assume that the true population confusion matrix for a classifier is $\mathbf{C}_P$ and the sample confusion matrix obtained from the field data is $\mathbf{C}_S$: that is,

$$\mathbf{C}_P = \begin{pmatrix} N_{11} & N_{12} & \cdots & N_{1r} \\ N_{21} & N_{22} & \cdots & N_{2r} \\ \vdots & \vdots & ::: & \vdots \\ N_{r1} & N_{r2} & \cdots & N_{rr} \end{pmatrix} \qquad \mathbf{C}_S = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1r} \\ n_{21} & n_{22} & \cdots & n_{2r} \\ \vdots & \vdots & ::: & \vdots \\ n_{r1} & n_{r2} & \cdots & n_{rr} \end{pmatrix}$$

Define

$$N_{i.} = \sum_{j=1}^{r} N_{ij} \qquad N_{.j} = \sum_{i=1}^{r} N_{ij}$$

$$N = \sum_{i=1}^{r} N_{.j} = \sum_{j=1}^{r} N_{j.} = \sum_{i=1}^{r}\sum_{j=1}^{r} N_{ij}$$

$$n_{i.} = \sum_{j=1}^{r} n_{ij} \qquad n_{.j} = \sum_{j=1}^{r} n_{ij}$$

$$n = \sum_{i=1}^{r} n_{.j} = \sum_{j=1}^{r} n_{j.} = \sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}$$

where

- N_{ij} = (true) population number of pixels being classified from class i to class j by the classifier, unknown;
- $N_{i.}$ = (true) population number of pixels in class i, unknown;
- $N_{.j}$ = (true) population number of pixels being classified into class j by the classifier;
- N = total number of the pixels in the population (whole scene);
- n_{ij} = sample number of pixels identified being classified from class i to class j by field verification;
- $n_{i.}$ = sample number of pixels in class i;
- $n_{.j}$ = sample number of pixels being classified into class j; and
- n = total number of pixels in the sample (sample size).

In order to simplify our following notations, we define two vectors as follows:

$$\mathbf{T} = (N_{1.}, N_{2.}, \ldots, N_{r.})^{\mathrm{T}} \qquad \mathbf{M} = (N_{.1}, N_{.2}, \ldots, N_{.r})^{\mathrm{T}}$$

where $\mathbf{T}$ consists of all unknown parameters for the classifier—the "true" numbers of pixels for the classes; and $\mathbf{M}$ consists of the numbers of pixels in each classes classified by the classifier ("mapped" pixels). Notice that $\mathbf{T}$ is an unknown vector whereas $\mathbf{M}$ is a known vector. The task for the area estimation correction is to obtain better area estimates of $\mathbf{T}$ based on knowledge of $\mathbf{M}$ and the sample confusion matrix $\mathbf{C}_S$.

The population omission and commission probability matrices $\mathbf{P}_P$ and $\mathbf{Q}_P$ are defined based on the population confusion $\mathbf{C}_P$ matrix: i.e.,

$$\mathbf{P}_P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1r} \\ P_{21} & P_{22} & \cdots & P_{2r} \\ \vdots & \vdots & ::: & \vdots \\ P_{r1} & P_{r2} & \cdots & P_{rr} \end{pmatrix} \qquad \mathbf{Q}_P = \begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1r} \\ Q_{21} & Q_{22} & \cdots & Q_{2r} \\ \vdots & \vdots & ::: & \vdots \\ Q_{r1} & Q_{r2} & \cdots & Q_{rr} \end{pmatrix}$$

where

$$P_{ij} = \frac{N_{ij}}{N_{i.}} \qquad Q_{ij} = \frac{N_{ij}}{N_{.j}}$$

$$P_{i.} = \frac{N_{i.}}{N} \qquad Q_{.j} = \frac{N_{.j}}{N}$$

and

- P_{ij} is the omission (*a priori*) probability for a pixel being classified as in class j given the condition it is from class i,
- Q_{ij} is the commission (*a posteriori*) probability for a pixel being from class i given the condition it is classified as in class j,
- $P_{i.}$ is the marginal *a priori* probability for class i, and
- $Q_{.j}$ is the marginal *a posteriori* probability for class j.

The sample omission and commission probability (frequency) matrices $\mathbf{P}_S$ and $\mathbf{Q}_S$ can be defined similarly based on the sample confusion matrix: i.e.,

$$\mathbf{P}_S = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & \vdots & ::: & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix} \qquad \mathbf{Q}_S = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1r} \\ q_{21} & q_{22} & \cdots & q_{2r} \\ \vdots & \vdots & ::: & \vdots \\ q_{r1} & q_{r2} & \cdots & q_{rr} \end{pmatrix}$$

$$p_{ij} = \frac{n_{ij}}{n_{i.}} \qquad q_{ij} = \frac{n_{ij}}{n_{.j}}$$

$$p_{i.} = \frac{n_{i.}}{n} \qquad q_{.j} = \frac{n_{.j}}{n}$$

where

- p_{ij} is the sample omission (*a priori*) probability for a pixel being classified as in class j given the condition it is from class i,
- q_{ij} is the sample commission (*a posteriori*) probability for a pixel being from class i given the condition it is classified as in class j,
- $p_{i.}$ is the sample marginal *a priori* probability for class i, and
- $q_{.j}$ is the sample marginal *a posteriori* probability for class j.

Marginal Area Estimators

Inverse Estimator

The inverse estimator is based on the inverse of the sample omission probability matrix (Bauer *et al.*, 1978). This method has been advocated by Hay (1988). Denote

$$\mathbf{T}^{(i)} = (N_{1.}^{(i)}, N_{2.}^{(i)}, \ldots, N_{r.}^{(i)})^{\mathrm{T}}$$

the inverse estimator for the marginal areas. Because

$$N_{.j} = \sum_{i=1}^{r} N_{ij} = \sum_{i=1}^{r} \frac{N_{ij}}{N_{i.}} N_{i.} = \sum_{i=1}^{r} P_{ij} N_{i.}$$

If $\mathbf{P}_S$ is used to estimate $\mathbf{P}_P$, then the inverse estimates can be solved from the equations

$$N_{.j} = \sum_{i=1}^{r} p_{ij} N_{i.}^{(i)} = \sum_{i=1}^{r} \frac{n_{ij}}{n_{i.}} N_{i.}^{(i)} \qquad j = 1, 2, \ldots, r$$

or, in matrix form,

$$\mathbf{M} = \mathbf{P}_S^{\mathrm{T}} \mathbf{T}^{(i)} .$$

Thus, the inverse estimates can be obtained by inverting $\mathbf{P}_S^{\mathrm{T}}$ if its inverse exists. In most application cases, $\mathbf{P}_S^{\mathrm{T}}$ is invertible and, therefore, the following equation holds:

$$\mathbf{T}^{(i)} = (\mathbf{P}_S^{\mathrm{T}})^{-1} \mathbf{M} .$$

Because $(\mathbf{P}_S^{\mathrm{T}})^{-1}$ could have some negative elements, the estimator $\mathbf{T}^{(i)}$ could also have some negativelements, but not all of them. If that happened, those negative elements can be adjusted to zeros.

Direct Estimator

The direct estimator (Card, 1982) is based on the sample commission probability matrix. Let's denote

$$\mathbf{T}^{(d)} = (N_{1.}^{(d)}, N_{2.}^{(d)}, \ldots, N_{r.}^{(d)})^{\mathrm{T}}$$

the direct estimator for the marginal areas. Because

$$N_{i.} = \sum_{j=1}^{r} N_{ij} = \sum_{j=1}^{r} \frac{N_{ij}}{N_{.j}} N_{.j} = \sum_{i=1}^{r} Q_{ij} N_{.j} \qquad i = 1, 2, ..., r .$$

If $\mathbf{Q}_S$ is used to estimate $\mathbf{Q}_P$, then the direct estimate $N_{i}^{(d)}$ for N_i can be obtained: i.e.,

$$N_{i.}^{(d)} = \sum_{j=1}^{r} q_{ij} N_{.j} = \sum_{j=1}^{r} \frac{n_{ij}}{n_{.j}} N_{.j}$$

or, in matrix form,

$$\mathbf{T}^{(d)} = \mathbf{Q}_S \mathbf{M} .$$

Additive Estimator

The additive correction (Dymond, 1992) is based on the idea that the true number of pixels in a class can be estimated by the classified number of pixels in the same class plus a correction term. Similarly, let's denote

$$\mathbf{T}^{(a)} = (N_{1.}^{(a)}, N_{2.}^{(a)}, \ldots, N_{r.}^{(a)})^{\mathrm{T}}$$

the additive estimator for the marginal areas. Because

$$\begin{aligned} N_{i.} &= N_{.i} + N_{i.} - N_{.i} \\ &= N_{.i} + P_{i.} N - Q_{.i} N \\ &= N_{.i} + (P_{i.} - Q_{.i}) N_i \end{aligned}$$

the difference $(P_{i.} - Q_{.i})N$ is the desired correction term. If the sample *a priori* and *a posteriori* marginal probabilities are used to estimate the corresponding population probabilities, the additive correction can be obtained by

$$\begin{aligned} N_{i.}^{(a)} &= N_{.i} + (p_{i.} - q_{.i}) N \\ &= N_{.i} + (n_{i.}/n - n_{.i}/n) N \\ &= N_i + (n_{i.} - n_{.i}) N/n \end{aligned}$$

or, equivalently,

$$N_{i.}^{(a)} = N_{.i} + (\sum_{j=1}^{r} n_{ij} - \sum_{j=1}^{r} n_{ji}) \frac{N}{n} \qquad i = 1, 2, ..., r.$$

Because subtraction is used in the computation, some of these estimates could be negative, but not all of them. As in the inverse correction situation, if that happens, the negative estimates can be adjusted to zeros.

Proportional Estimator

The discussion of the simple proportional correction or simple expansion is only for the purpose of comparison and demonstration. The details about simple expansion had been discussed by Cochran (1977, p. 50). We denote

$$\mathbf{T}^{(p)} = (N_{1.}^{(p)}, N_{2.}^{(p)}, \ldots, N_{r.}^{(p)})^{\mathrm{T}}$$

the proportional estimator for the marginal areas. The elements of $T^{(p)}$ have the simplest form: i.e.,

$$N_{i.}^{(p)} = n_{i.} \frac{N}{n} \qquad i = 1, 2, ..., r.$$

Classifier Constraints

The invertibility of the sample omission probability matrix has been emphasized for inverse correction by many authors. One reasonable and natural sufficient condition to guarantee such invertibility is requiring that the classifier be practically minimum acceptable (Yuan, 1996). We mean that a classifier is practically minimum acceptable if it has a larger sample probability of getting correct classification for all classes. That is,

$$p_{ii} = \mathrm{Prob}_{\mathrm{sample}}\ (\text{class } i \,|\, \text{class } i) > 0.5 \qquad i = 1, 2, ..., r.$$

Similarly, we can define a minimum acceptable classifier, though it can hardly be verified, if

$$p_{ii} = \mathrm{Prob}_{\mathrm{population}}\ (\text{class } i \,|\, \text{class } i) > 0.5 \qquad i = 1, 2, ..., r.$$

Yuan (1996) has proved the existence of the inverse correction under the minimum practically acceptable condition.

Using notations of n_{ij}, a practically minimum acceptable classifier would have

$$n_{ii} > 0.5 n_{i.} \qquad i = 1, 2, ..., r.$$

Confusion Matrix Generating Models

In order to simulate the behaviors of the minimum practically acceptable classifiers, an appropriate Monte Carlo model must be established. Yuan (1996) has shown that if the sample omission probability matrices were generated using simple uniform distribution, then the majority of the simulated matrices would not represent any feasible sample omission probability matrices of a practically minimum acceptable classifier. In other words, purely randomly generated matrices only have small chance of becoming sampling omission matrices for a practically minimum acceptable classifier. In order to model the "true" image classification situation, the random matrix generating process must first be justified.

Random Model for $\mathbf{C}_P$

Because there is no theoretical distribution model for the population confusion matrix, uniform distribution is used for generating population confusion matrices. That is, for a given population size N,

(1) Assume $N_{1.}, N_{2.}, \ldots, N_{r.}$ are uniformly distributed inside a multidimensional cone defined by $N_{1.} > 0$, $N_{2.} > 0, \ldots, N_{r.} > 0$ and $N_{1.} + N_{2.} + \cdots + N_{r.} = N$; and
(2) For each $N_{i.}$, assume $N_{i1}, N_{i2}, \ldots, N_{ir}$ are uniformly distributed inside a multidimensional cone defined by $N_{i1} > 0, \ldots, N_{ii} > 0.5N_{i}, \ldots, N_{ir} > 0$ and $N_{i1} + N_{i2} + \cdots + N_{ir} = N_{i}$, where constraint $N_{ii} > 0.5\, N_i$ is used such that the classifier is minimum acceptable.

Random Model for $\mathbf{C}_S$

(1) Asssume simple random sampling scheme is used. The distribution for $\mathbf{C}_S$ can be approximated by normal distribution. Using a binomial model under simple sampling scheme, one can prove (Cochran, 1977, pp. 50–51) that

$$E(\frac{n_{i.}}{n}) = \frac{N_{i.}}{N} ,$$

$$V(\frac{n_{i.}}{n}) = \frac{1-f}{n}\frac{N}{N-1}\frac{N_{i.}}{N}(1 - \frac{N_{i.}}{N}) ,$$

or

$$E(n_{i.}) = \frac{n}{N} N_{i.} = fN_{i.} ,$$

$$V(n_{i.}) = n(1-f)\frac{N}{N-1}\frac{N_{i.}}{N}(1 - \frac{N_{i.}}{N}) ,$$

By the central limit theory

$$\frac{n_{i.} - E(n_{i.})}{\sqrt{V(n_{i.})}} \rightarrow N(0, 1) \qquad \text{as } n \rightarrow N ,$$

or, equivalently,

$$n_{i.} \rightarrow N(E(n_{i.}), V(n_{i.})) \qquad \text{as } n \rightarrow N .$$

In practice, $n_{1.}, n_{2.}, \ldots, n_{r.}$ are confined by conditions: $0 < n_{i.} \leq N_{i.}$, $0 < n_{2.} \leq N_{2.}, \ldots, 0 < n_{r.} \leq N_{r.}$ and $n_{1.} + n_{2.} + \cdots + n_{r.} = n = fN$.

(2) Using similar binomial reasoning, we have

$$E(\frac{n_{ij}}{n}) = \frac{N_{ij}}{N} ,$$

$$V(\frac{n_{ij}}{n}) = \frac{1-f}{n}\frac{N}{N-1}\frac{N_{ij}}{N}(1 - \frac{N_{ij}}{N}) ,$$

or

$$E(n_{ij}) = \frac{n}{N} N_{ij} = fN_{ij} ,$$

$$V(n_{ij}) = n(1-f)\frac{N}{N-1}\frac{N_{ij}}{N}(1 - \frac{N_{ij}}{N}) ,$$

By central limit theory

$$\frac{n_{ij} - E(n_{ij})}{\sqrt{V(n_{ij})}} \rightarrow N(0, 1) \qquad \text{as } n \rightarrow N ,$$

or, equivalently,

$$n_{ij} \rightarrow N(E(n_{ij}), V(n_{ij})) \qquad \text{as } n \rightarrow N .$$

In practice, $n_{i1}, n_{i2}, \ldots, n_{ir}$ are also confined by conditions $0 < n_{i} \leq N_{i1}, \ldots, 0.5n_{i.} < n_{ii} \leq N_{ii}, \ldots, 0 < n_{ir} \leq N_{ir}$, where $n_{ii} > 0.5\, n_{i}$ is such that the classifier is practically minimum acceptable; and $n_{i1} + n_{i2} + \cdots + n_{ir} = n_{i.}$

Evaluation Criteria

Two evaluation criteria used by Czaplewski and Catts (1992) are adopted with minor modification for this simulation. Assume for each sampling fraction f, m_p population confusion matrices are simulated. For each simulated population confusion matrix, m_s sample confusion matrices are simulated. Let $\mathbf{T}_i$ be the "true" marginal area vector for the ith simulated population and $\mathbf{T}_{ij}^{(e)}$ be the estimate of $\mathbf{T}_i$ based on the simulated jth sample confusion matrix where $e = i$ for the inverse estimator, $e = d$ for the direct estimator, $e = a$ for the additive estimator, and $e = p$ for the proportional estimator.

Average Absolute Bias

Using above definitions, the mean estimation vector for the ith population is

$$\overline{\mathbf{T}}_i^{(e)} = \frac{1}{m_s}\sum_{j=1}^{m_s} \mathbf{T}_{ij}^{(e)} .$$

The average error vector for the given estimator and the given population is

$$\mathbf{E}_i^{(e)} = \overline{\mathbf{T}}_i^{(e)} - \mathbf{T}_i .$$

To reduce the scaling factor, the corresponding relative bias vector

$$B_i^{(e)} = \frac{1}{\|T_i\|} E_i^{(e)}$$

is introduced, where

$$\|T_i\| = \sum_{k=1}^{r} | N_{k.} | .$$

Let $b_i^{(e)} = \|B_i^{(e)}\|$ be the total relative absolute bias for the population i, then the average absolute relative bias for all m_p simulated populations is

$$b^{(e)} = \frac{1}{m_p}\sum_{i=1}^{m_p} b_i^{(e)} .$$

Average Dispersion

For the ith simulated population, the square root of the mean of the sum of errors is defined by

$$s_i^{(e)} = \sqrt{\frac{1}{m_s - 1}\sum_{j=1}^{m_s} \|T_{ij}^{(e)} - \overline{T}_i\|_2^2} .$$

where $\|x\|_2$ is the 2-norm of a vector $\mathbf{x}$.

Considering the scaling factor, the relative dispersion for the ith simulated population is given by

$$d_i^{(e)} = \frac{S_i^{(e)}}{\|T_i\|} .$$

The average dispersion for all m_p simulated populations is

$$d^{(e)} = \frac{1}{m_p}\sum_{i=1}^{m_p} d_i^{(e)} .$$

In our simulation, the measurements $b^{(e)}$ and $d^{(e)}$ were computed for four different estimators for each sampling fraction that ranged from 0.05 to 0.95 of 0.05 increment.

Simulation Procedure

The simulation procedure is as follows:

(1) Generate sample population confusion matrix $\mathbf{P}_p$:
- (1-1) Generate true population marginal areas (uniform) $\mathbf{T} = (N_{1.}, N_{2.}, \ldots, N_{r.})^T$ such that $N_{1.} + N_{2.} + \cdots + N_{r.} = N$.
- (1-2) For each $N_{i.}$, generate uniform vector $(N_{i1}, N_{i2}, \ldots, N_{ir})$ such that $N_{i1} > 0, \ldots, N_{ii} > 0.5N_{i.}, \ldots, N_{ir} > 0$ and $N_{i1} + N_{i2} + \cdots + N_{ir} = N_{i.}$

(2) Generate confusion matrix $\mathbf{P}_S$ consistent with the $\mathbf{P}_P$ generated in step (1):
- (2-1) For given sampling fraction f, let $n = f^*N$.
- (2-2) Generate true sample marginal areas normal approximations $(n_{1.}, n_{2.}, \ldots, n_{r.})$ such that $0 < n_{i.} \leq N_{i.}$ and $n_{1.} + n_{2.} + \cdots + n_{r.} = n$.
- (2-3) For each $n_{i.}$, generate normal approximations $(n_{i1}, n_{i2}, \ldots, n_{ir})$ such that $0 \leq n_{ij} \leq N_{ij}$, $n_{ii} > 0.5\, n_i$ and $n_{i1} + n_{i2} + \cdots + n_{ir} = n_{i.}$.

(3) Compute four marginal area estimators: If some of the elements of an estimator are less than zero, then adjust them to zero.

(4) Update statistics including the two accuracy measures for the given simulated population.

(5) Repeat steps (2) to (4) m_s times. Output the final statistics for the given population and update the statistics including the accuracy measures for the same sampling fraction.

(6) Repeat steps (1) to (5) m_p times. Output the final statistics for the given sampling fraction.

(7) Repeat steps (1) to (6) for $f = 0.05, 0.10, \ldots, 0.95$.

(8) Repeat steps (1) to (7) for $r = 4, 6, \ldots, 20$.

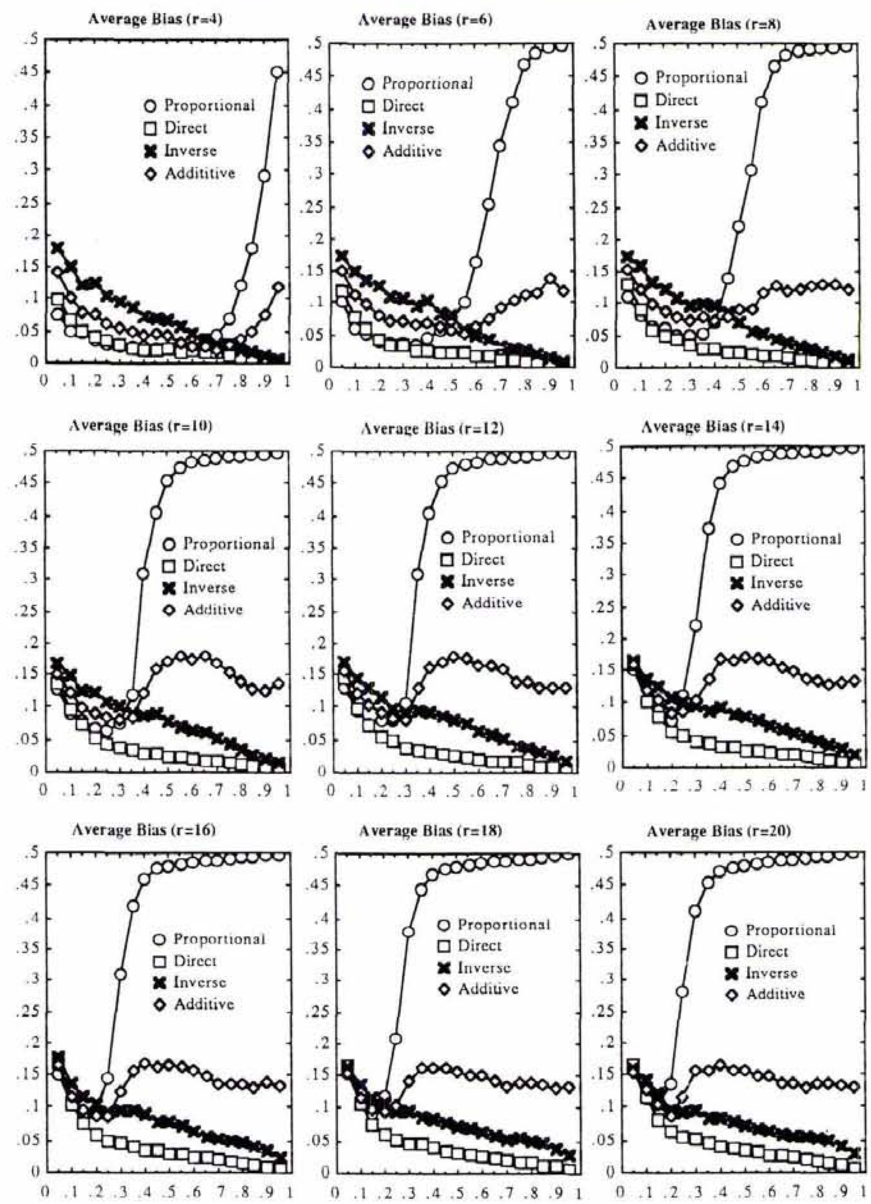

Figure 1. Average absolute bias under various sampling fraction schemes, where the *x*-axes represent the sampling fraction (f), the *y*-axes represent the normalized average absolute bias, and $r =$ the number of classes in the classification.

In the simulation, we set $m_p = m_s = 50$, that is, 2500 pairs of population and sample confusion matrices had been simulated for each sampling fraction f. For each r (number of classes), 45000 such pairs have been simulated.

The simulation process was conducted on the supercomputer Cray Y-MP2/216 in the National Supercomputing Center for Energy and the Environment. Total simulation took about 10 CPU hours.

Simulation Results

The simulation results are plotted in Figure 1 and Figure 2. In those plots, the *x*-axes represent sampling fractions that ranged from 0.05 to 0.95 by an increment of 0.05; the *y*-axes represent the average absolute bias and average dispersion, respectively. Those plots adequately demonstrated the bias, dispersion, and asymptotic characteristics of the four simulated estimators under the minimum practically acceptable constraint and changing parameters r and f.

Average Bias

The average absolute biases (in short, bias) for the four estimators under different sampling schemes are given in Figure 1. Some important observations can be made on those plots.

(1) When r is small ($r \leq 10$) and the sampling fraction is small ($f \leq 0.35$), the proportional estimator often has the smallest bias and the inverse estimator often has the largest bias. However, they are all within acceptable range and show a tendency of decreasing bias as the sampling fraction increases.

(2) When r is large ($r > 10$) and the sampling fraction is small ($f < 0.25$), there is not a significant difference among different estimators in terms of bias, and they all show a tendency to decrease as the sampling fraction increases.

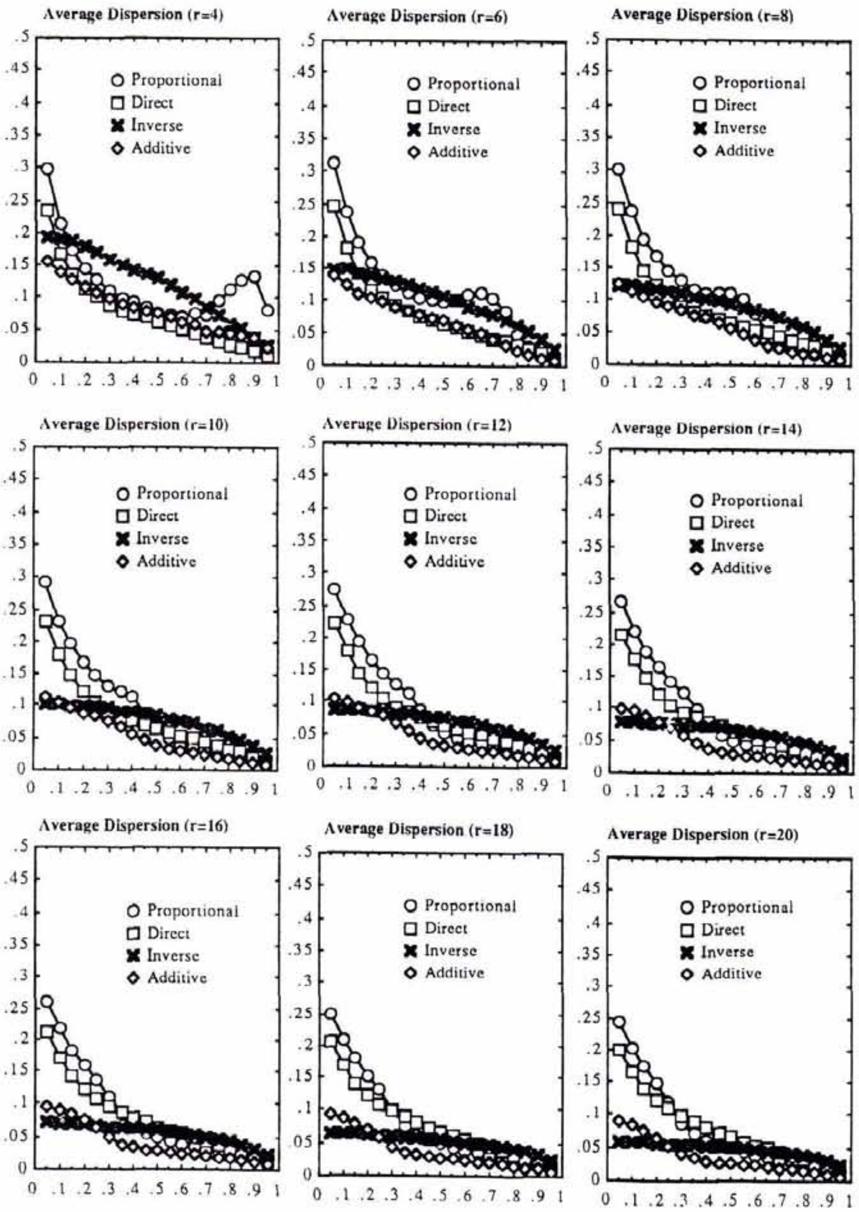

Figure 2. Average dispersion under various sampling fraction schemes, where the x-axes represent the sampling fraction (f) the y-axes represent the normalized average dispersion, and r = the number of classes in the classification.

(3) When the sampling fraction f is large, the bias in the proportional estimator increases rapidly. The bias in the additive estimator become stabilized but far-away from zero; and the bias in both the direct and inverse estimators approaches zero.

(4) The direct estimator generally has the smallest bias of all four estimators under all sampling fractions.

Average Dispersion

The average dispersions for the simulations are displayed in Figure 2. The following are the direct observations.

(1) The dispersions for all four estimators approach zero as the sampling fraction f increases. For asymptotically unbiased estimators (inverse and direct estimators), this means that they asymptotically approach the true values of **T**. For asymptotically biased estimators (proportional and additive estimators), this suggests that their estimates will be dominated by bias when f is large.

(2) When the sampling fraction f is small, proportional and direct estimators normally have significantly bigger dispersions than inverse and additive estimators. This suggests that, when the sampling fraction is small, estimates obtained by inverse and additive estimators are more stable than those obtained by others.

Estimator Evaluation

(1) The proportional estimator is not suggested for use because of its large dispersion when the sampling fraction is small, and its large bias when the sampling fraction is large.

(2) Both the inverse and additive estimators fairly behaved

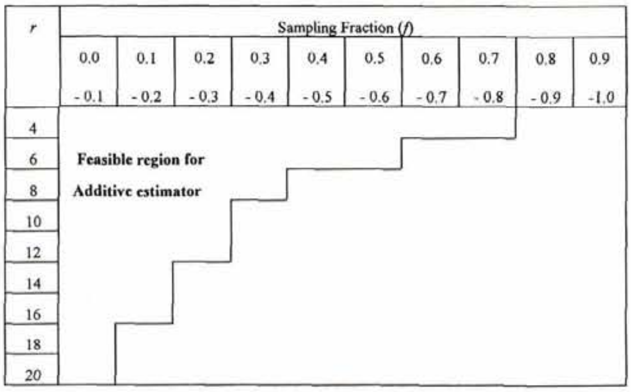

Figure 3. Feasible region for the additive estimator, where r = the number of classes.

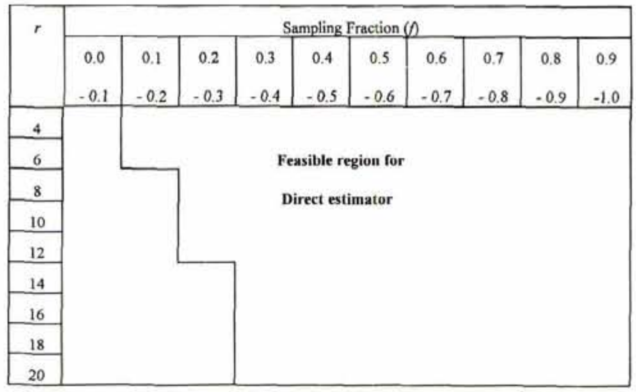

Figure 5. Feasible region for the direct estimator, where r = the number of classes.

when the sampling faction is small ($f < 3$), particularly when $r \geq 8$, because of a reasonable bias and small dispersion. However, the additive estimator is not suggested for use when the sampling fraction is large because of its strong bias.

(3) Both the inverse and direct estimators are asymptotically unbiased and zero-dispersed. However, the direct estimator has a systematic bias smaller than the inverse estimator. When the sampling fraction is large, the direct estimator is asymptotically better than the inverse estimator. This result is similar to that of Czaplewski and Catts (1992). When the sampling fraction is small, the inverse estimator has a bias similar to that of the direct estimator, but with a significantly smaller dispersion. Therefore, the inverse estimator could be more suitable for applications when the sampling fraction is small, which is the common situation in accuracy verification of image classification.

The simulated feasible regions for the three marginal area estimators are plotted in Figure 3, 4, and 5. However, those regions were obtained from the interpretation of the simulations study reported in this paper and therefore are dependent on the simulation environment. They could be used as reference, but should not be considered as strictly proven statements or universal rules for method selection.

Conclusions

Through Monte Carlo simulation, three marginal area estimators for image classification were statistically studied. The results suggest that, if the classifier is minimum practically acceptable, then different estimators have different feasible regions and no single one is systematically superior to the others in all sampling situations. When the sampling fraction is small, additive and inverse estimators are relatively better than the direct estimator. When the sampling fraction is large, the direct estimator is better than the inverse estimator. The additive estimator is strongly biased and is not suggested for use when the sampling fraction is large or when the number of classes is large.

The real situation could be more complicated than our simulation. For instance, the simulation only modeled the case when

$$\omega = \min_{1 \leq i \leq r} p_{ii} > 0.5.$$

We still don't know the statistical behavior of the marginal area estimators when

$$\omega = \min_{1 \leq i \leq r} p_{ii} > 0.7$$

which is probably more realistic in image classification. We suspect that when ω increases, the inverse estimator has more chances of getting better estimates than does the direct estimator, and when ω decreases, the direct estimator has more chances of getting better estimates than does the inverse estimator. However, more simulation work is needed in order to better characterize this situation.

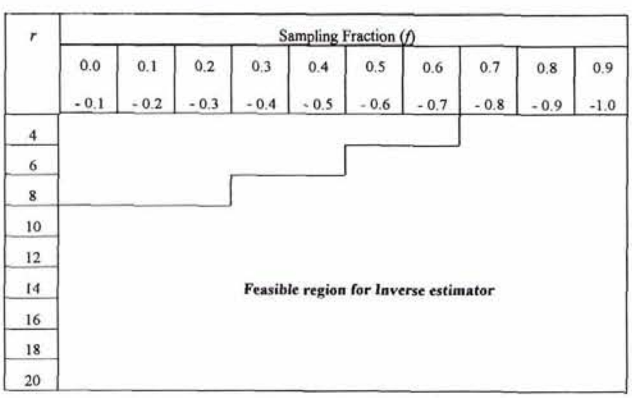

Figure 4. Feasible region for the inverse estimator, where r = the number of classes.

Acknowledgments

This research had been supported by the National Supercomputing Center for Energy and the Environment in terms of computing time. The author wishes to thank Dr. Czaplewski, Dr. Dudewicz, and Dr. Lyon for their valuable comments.

References

Bauer, M.E., M.M. Hixson, B.J. Davis, and J.B. Etheridge, 1978. Area estimation of crops by digital analysis of Landsat data, *Photogrammetric Engineering & Remote Sensing*, 44:1033–1043.

Card, D.H., 1982. Using known map categorical marginal frequencies to improve estimates of thematic map accuracy, *Photogrammetric Engineering & Remote Sensing*, 48:431–439.

Cochran, W.C., 1977. *Sampling Techniques, Third Edition*, John Wiley & Sons, New York.

Czaplewski, R.L., and G.P. Catts, 1990. Calibrating area estimates for classification error using confusion matrices, *Proceeding ACSM/ASPRS Convention*, Denver, Colorado, 4:431–440.

———, , 1992. Calibration of remotely sensed proportion or area estimates for misclassification error, *Remote Sensing of Environment*, 39:29–43.

Dymond, J.R., 1992, How accurately do image classifier estimate area? *International Journal of Remote Sensing*, 13:1735–1742.

———, , 1993. An improved Skidmore/Turner classifier, *Photogrammetric Engineering & Remote Sensing*, 59(5):623–626.

Hay, A.M., 1988. The derivation of global estimates from a confusion matrix, *International Journal of Remote Sensing*, 9:1395–1398.

Jupp, D.L.B., 1989. The stability of global estimates from confusion matrices, *International Journal of Remote Sensing*, 10:1563–1569.

Maxim, L.D., L. Harrington, and M. Kennedy, 1981. Alternative "scale-up" estimates for aerial surveys where both detection and classification errors exist, *Photogrammetric Engineering & Remote Sensing*, 47:1227–1239.

Prisley, S.P., and J.L. Smith, 1987. Using classification error matrices to improve the accuracy of weighted land-cover models, *Photogrammetric Engineering & Remote Sensing*, 53:1259–1263.

Schriever, J.R., and R.G. Congalton, 1995. Evaluating seasonal variability as an aid to cover-type mapping from Landsat Thematic Mapper data in the northeast, *Photogrammetric Engineering & Remote Sensing*, 61(3):321–327.

Stohr, C., R.G. Darmody, T.D. Frank, A.P. Elhance, R. Lunetta, D. Worthy, and K. O'Connor-Shoresman, 1994. Classification of depressions in landfill covers using uncalibrated thermal-infrared imagery, *Photogrammetric Engineering & Remote Sensing*, 60(8): 1019–1028.

Yuan, D., 1996. Natural constraints for inverse area estimate corrections, *Photogrammetric Engineering & Remote Sensing*, 62(4): 413–417.

(Received 31 July 1995; accepted 18 December 1995; revised 18 April 1996)