Natural Constraints for Inverse Area Estimate Corrections

Ding Yuan

Abstract

Though it has been used for marginal area estimate correction in image classification for years, the inverse correction technique has been the most controversial compared with several other marginal area estimate correction techniques, such as the direct and additive methods. In the reported practices, the inverse correction technique provided acceptable corrections to the marginal area estimates. In statistical simulation comparison, however, the inverse method was found unstable and systematically inferior to the direct method. Our objective in this study was to investigate what has caused this controversy. Through theoretic analysis and discussions on the characteristics of inverse correction for image classification, the author concludes that (1) the inverse correction exists if the classifier is minimum practically acceptable and (2) the inverse is not ill-conditioned (i.e., it is stable) if the classifier is reasonably acceptable.

Introduction

Remotely sensed data are often used for land-cover classification. The classification is commonly quantified by providing estimates for areas or numbers of pixels for different land-cover classes using a marginal area estimate technique such as the direct counting technique. Due to the classification error, the direct counting marginal area estimates are sometimes biased or prone to error. These estimate methods are often found unacceptable unless they are consistent with the definition and measurement protocol used for reference data (Thomas, 1986; Burk *et al.*, 1988; Poso, 1988; Czaplewski and Catts, 1992). Therefore, it is necessary to make corrections to the area estimates based on the knowledge regarding (mis)classification of the remotely sensed data. Three major correction techniques with various modifications – inverse correction (Bauer *et al.*, 1978; Maxim *et al.*, 1981; Prisley and Smith, 1987; Hay, 1988; Jupp, 1989; Hay, 1989; Czaplewski and Catts, 1990; Czaplewski and Catts, 1992; Czaplewski, 1992), direct correction (Card, 1982; Chrisman, 1982; Maselli *et al.*, 1990; Czaplewski and Catts, 1990; Czaplewski and Catts, 1992; Czaplewski, 1992), and additive correction (Dymond, 1992) – have been used or proposed for marginal area estimations.

The inverse correction, which was introduced to remote sensing for crop area estimations by Bauer *et al.* (1978), makes use of the inverse matrix of the forward conditional confusion probability matrix (for definition, see context). When the survey is based on the field classes and designed independently of image interpretation, the inverse estimate is a natural choice (Jupp, 1989). Although this method may fail because of singularity of the forward conditional confusion probability matrix, it has been the most widely used correction method for area estimate. Because it is a straightforward algorithm and produces acceptable results for applications, most of the early application work on area estimate correction was based on the inverse correction technique (Bauer *et al.*, 1978; Maxim *et al.*, 1981; Prisley and Smith, 1987).

Although applications of inverse correction have been frequently reported in remote sensing, many authors have been questioning the instability of the inverse correction method, which was induced by the singularity or near-singularity of the sample forward conditional confusion probability matrix. Jupp (1989) studied the relative stability of the inverse correction as compared with the direct correction. He found that the relative stability of the inverse correction for area estimate depended on the singular values of the sample forward conditional confusion probability matrix. The difference between two estimates (direct and inverse) is a function of the separation or contingency between the interpreted and surveyed classes. Through a theoretic analysis, Jupp concluded that the direct correction method is more stable, or less sensitive, than the inverse correction. The instability caused by inverse correction becomes serious when the singular values of a sample forward conditional confusion probability matrix approach zero. In a recent paper by Dymond (1992), an additive correction method was proposed for area estimate correction. The stability of the additive method is better than that of the inverse method when the sample confusion matrix is singular or near singular, because it does not involve the inverse of the sample confusion matrix.

In a recent numerical simulation study by Czaplewski and Catts (1992) (note: the inverse correction technique was based on the classical statistical model while the direct correction method was based on the inverse statistical model), a series of simulations for both direct and inverse corrections had been performed for a wide range of number of classes (from 4 to 21). The inverse correction (classical statistical model) was found systematically inferior to the direct correction (inverse statistical model). However, their results also suggest that the inverse correction had a better chance of getting acceptable area estimation for cases of a smaller number of classes.

Why aren't the theoretic analysis (Jupp, 1989), the simulations (Czaplewski and Catts, 1992), and the practical applications consistent with one another in the inverse correction situation? Inverse correction, though it may not necessarily be superior to the others in every case, is still an acceptable area estimate correction method. The objective of this study was to investigate this confusion created by theoretic analysis, mathematical simulation and practical applications.

Basic Definitions

Let r denote the 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 survey data is $\mathbf{C}_S$. As usual, we have the following

Photogrammetric Engineering & Remote Sensing,
Vol. 62, No. 4, April 1996, pp. 413–417.

0099-1112/96/6204–413$3.00/0

Desert Research Institute, University
and Community College System of Nevada,
P.O. Box 19040, Las Vegas, NV 89132-0040.

notations and definitions, where rows represent reference classes and columns represent classified classes:

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

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

$$N = \sum_{i=1}^{r} N_{i.} = \sum_{j=1}^{r} N_{.j} = \sum_{i=1}^{r} \sum_{j=1}^{r} N_{ij} \tag{3}$$

$$n_{i.} = \sum_{j=1}^{r} n_{ij} \qquad n_{.j} = \sum_{i=1}^{r} n_{ij} \tag{4}$$

$$n = \sum_{j=1}^{r} n_{.j} = \sum_{j=1}^{r} n_{j.} = \sum_{i=1}^{r} \sum_{j=1}^{r} n_{ij} \tag{5}$$

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

We also define

$$\mathbf{A}_1 = (N_{1.},\ N_{2.},\ldots,N_{r.})^{\mathrm{T}} \qquad \mathbf{A}_2 = (N_{.1},\ N_{.2},\ldots,N_{.r})^{\mathrm{T}} \tag{6}$$

where $\mathbf{A}_1$ is the vector of true unknown number of pixels (marginal areas) of the classes in the whole scene and $\mathbf{A}_2$ is the vector of the numbers of pixels in each classes obtained from digital classification. Notice that $\mathbf{A}_1$ is an unknown vector whereas $\mathbf{A}_2$ is a known vector. The goal for marginal area estimation correction is to obtain a reasonable marginal area estimate for $\mathbf{A}_1$ based on knowledge of $\mathbf{A}_2$ and the sample confusion matrix $\mathbf{C}_S$.

The population forward conditional confusion probability matrices $\mathbf{P}_P$ and sample forward conditional confusion frequency matrix $\mathbf{P}_S$ are defined based on the population and sample confusion matrices $\mathbf{C}_P$ and $\mathbf{C}_S$, respectively: i.e.,

$$\mathbf{P}_P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1r} \\ P_{21} & P_{22} & \cdots & P_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ P_{r1} & P_{r2} & \cdots & P_{rr} \end{pmatrix} \quad \mathbf{P}_S = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix} \tag{7}$$

$$P_{ij} = \frac{N_{ij}}{N_{i.}} \qquad p_{ij} = \frac{n_{ij}}{n_{i.}} \tag{8}$$

$$P_{i.} = \frac{N_{i.}}{N} \qquad p_{i.} = \frac{n_{i.}}{n} \tag{9}$$

where P_{ij} is the conditional probability for a pixel being classified as in class j given the condition that it is from class i; $P_{i.}$ is the marginal *a priori* probability for class i; p_{ij} is the sample conditional probability for a pixel being classified as in class j given the condition that it is from class i; and $p_{i.}$ is the sample marginal *a priori* probability for class i.

Inverse Correction

The inverse correction method for marginal area estimates was proposed by Bauer *et al.* (1978) and has been advocated by Hay (1988; 1989). Denote

$$\hat{\mathbf{A}}_1 = (\hat{N}_{1.},\ \hat{N}_{2.},\ldots,\hat{N}_{r.})^{\mathrm{T}} \tag{10}$$

the inverse correction for the population area estimations. Because

$$N_{.j} = \sum_{i=1}^{r} N_{ij} = \sum_{i=1}^{r} \frac{N_{ij}}{N_{i.}} N_{i.} = \sum_{i=1}^{r} P_{ij}\ N_{i.} \tag{11}$$

or, in matrix form,

$$\mathbf{A}_1 = \mathbf{P}_P^{\mathrm{T}}\ \mathbf{A}_1 \tag{12}$$

If $\mathbf{P}_S$ is used to estimate $\mathbf{P}_P$, then the inverse corrections $\hat{\mathbf{A}}_1$ can be obtained from the equations

$$\mathbf{A}_2 = \mathbf{P}_S^{\mathrm{T}}\ \hat{\mathbf{A}}_1 \tag{13}$$

Thus, the inverse corrections can be solved by inverting $\mathbf{P}_S^{\mathrm{T}}$ if its inverse exists. In most application cases, $\mathbf{P}_S^{\mathrm{T}}$ is invertible. For instance, if a classifier satisfies the constraints discussed in the next section, then $\mathbf{P}_S^{\mathrm{T}}$ is invertible. Therefore, the following equation holds:

$$\hat{\mathbf{A}}_1 = (\mathbf{P}_S^{\mathrm{T}})^{-1}\ \mathbf{A}_2. \tag{14}$$

Apparently, the possibility and stability of the solutions of inverse correction depend on the invertibility of the sample forward conditional confusion probability matrix. Jupp (1989) correctly criticized the instability of the solutions of inverse correction when the singularity values of the sample forward confusion probability matrix approach zero.

Acceptable Classifiers

It should be noted that, when the inverse correction is applied to the remote sensing classification results, the confusion matrices are no longer arbitrary ones. Instead, the classifiers used for classifications are presumably "accurate" and "acceptable." That is, they have been tested or verified in some way such that the probabilities for obtaining correct classifications for all the classes are larger than those for getting errors. Mathematically, the minimum acceptability for a classifier would mean

$$P_{ii} = \mathrm{Prob}_{\mathrm{population}}\ (\text{Map class } i \mid \text{True class } i) > 0.5,\ i = 1,2,\ldots,r. \tag{15}$$

Because the population forward conditional confusion probability matrix is often estimated by a sample forward conditional probability confusion matrix, condition (15) can be replaced by the corresponding condition about sample forward conditional confusion probabilities: i.e.,

$$p_{ii} = \mathrm{Prob}_{\mathrm{sample}}\ (\text{Map class } i \mid \text{True class } i) > 0.5,\ i = 1,2,\ldots,r. \tag{16}$$

A classifier that satisfies condition (15) can be said to be minimum theoretically acceptable; while condition (16) is satisfied, it is said to be minimum practically acceptable. If a classifier is minimum practically acceptable, it can be shown that it has an inverse correction. In other words, condition (16) is a sufficient condition for invertibility required by the inverse correction.

To prove the above statement, consider a minimum practically acceptable classifier, because

$$\sum_{j \neq i,\ 1 \leq j \leq r} p_{ij} = 1 - p_{ii} < 1 - 0.5 = 0.5 < p_{ii},\ i = 1,2,\ldots,r. \quad (17)$$

the sample forward conditional confusion probability matrix $\mathbf{P}_S$ for the classifier is a non-negative, strictly diagonally dominant matrix. Therefore, it is invertible (Horn and Johnson (1985), Theorem 6.1, p. 349) and so is $\mathbf{P}_S^T$.

It can be observed that condition (16) is not normally acceptable for applications. A more reasonable constraint for a successful classifier would be

$$p_{ii} = \text{Prob}_{\text{sample}}\ (\text{Map class } i \mid \text{True class } i) \geq 0.7, \quad i = 1,2,\ldots,r. \quad (18)$$

Therefore, a classifier satisfying condition (18) will be called a "reasonably acceptable" classifier in this study.

Stability of Inverse Correction

Jupp (1989) has discussed the stability of inverse correction with respect to the singular values of the sample forward confusion probability matrix. Based on our definitions on acceptable classifiers, the stability of the solutions of inverse corrections can be much simpler. Suppose $\mathbf{P}_S$ is the sample forward conditional confusion probability matrix for a minimum acceptable classifier; therefore, it is non-negative and strictly diagonally dominant. Denote

$$\omega = \min_{1 \leq i \leq r} p_{ii} \qquad (>0.5) \quad (19)$$

One can prove that all eigenvalues of $\mathbf{P}_S$ are contained in the complex disc (Minc (1988), Theorem 6.14):

$$|\lambda - \omega| \leq 1 - \omega < 0.5. \quad (20)$$

Because

$$||\lambda| - \omega| \leq |\lambda - \omega| \leq 1 - \omega; \quad (21)$$

therefore,

$$1 \geq |\lambda| > 2\omega - 1 \qquad (> 0) \quad (22)$$

Our first conclusion about a minimum practically acceptable classifier is that its sample forward conditional confusion probability matrix $\mathbf{P}_S$ is invertible (equivalent to all its singular values being non-zero). All of the eigenvalues (or singular values in Jupp's paper (Jupp, 1989)) of $\mathbf{P}_S$ have positive real parts (Horn and Johnson (1985), Theorem 6.1.10). Actually, 1 is one of the eigenvalues corresponding to eigenvector $\mathbf{1} = (1,1,\ldots,1)^T$ of $\mathbf{P}_S$. $\mathbf{P}_S^T$ is also invertible because $\mathbf{P}_S$ and $\mathbf{P}_S^T$ have the same eigenvalues. Therefore, Equation 14 holds. If the given classifier is a reasonably acceptable classifier (in most application cases), one would further have

$$\omega = \min_{1 \leq i \leq r} p_{ii} \geq 0.7 \quad (23)$$

and

$$1 \geq |\lambda| \geq 2\omega - 1 \geq 0.4. \quad (24)$$

Therefore, according to Jupp (1989), the solutions of the inverse corrections would be quite stable. This probably could explain why, in most reported application cases, the solutions of inverse corrections were stable.

Conditional number is a common measure of error propagation rate (stability) for solving a linear system (Golub and Loan, 1989). The 2-norm condition number for $\mathbf{P}_S^T$ is

$$\kappa_2(\mathbf{P}_S^T) = \kappa_2(\mathbf{P}_S) = \|\mathbf{P}_S\|_2 \|\mathbf{P}_S^{-1}\|_2 = \frac{\lambda_{\max}(\mathbf{P}_S)}{\lambda_{\min}(\mathbf{P}_S)} \leq \frac{1}{2\omega - 1}. \quad (25)$$

The 2-norm of a matrix $\mathbf{M}$ is defined (Horn and Johnson, 1985) by

$$\|\mathbf{M}\|_2 = \sqrt{\lambda_{\max}(\mathbf{M}^T\mathbf{M})}. \quad (26)$$

For a reasonably acceptable classifier, one would have

$$\kappa_2(\mathbf{P}_S^T) \leq \frac{1}{2\omega - 1} \leq \frac{1}{2 \times 0.7 - 1} = 2.5. \quad (27)$$

In numerical analysis, a linear system with a conditional number falling into this range would normally be considered very stable.

However, if the classifier is merely minimum acceptable or even not acceptable, the error propagation may be out of control, and sometimes error propagation could be a disaster. It is recommended that the classifier always be checked in order to determine that it is reasonably acceptable before using the inverse correction.

Simulation Controversy

The above discussion obviously interpreted the controversy between practical application results and theoretical analysis by Jupp (1989). The interpretation for the controversy between the application results and the simulation results by Czaplewski and Catts (1992) was more subjective, because the $\mathbf{P}_S$ generating procedure used by Czaplewski and Catts (1992) was not fully presented in the paper. However, the following are the possibilities that might have resulted in the controversial results in their simulation:

(1) A regular transition matrix in a normal Marcovian chain may not necessarily represent a forward conditional confusion probability matrix of an image classifier. That is, if the transition matrix is not strictly diagonally dominant, normally it does not represent any acceptable classification. Only a small portion of Marcovian chain transition matrices can be considered as a $\mathbf{P}_S$ for an image classifier.

(2) If the generated $\mathbf{P}_S$ had zero elements, the Bayesian method had been used for modifying the zero elements (Czaplewski and Catts, 1992). If these zero elements happened to be in the diagonal positions, the smoothed matrix would be very likely ill-conditioned. In this case, the solutions of the modified linear system may not present the true solution at all. There are numerous examples about this problem in standard text books for numerical analysis such as the one by Golub and Loan (1989).

(3) If the random variable generating process was based on [0 1]-uniform distributions for $\mathbf{P}_S$ generating, one would have

$$\text{Prob (generate } p_{ii} \leq 0.7) = 0.7, \quad \text{Prob (generate } p_{ii} > 0.7) = 0.3. \quad (28)$$

If the generating of p_{ii}'s is independent, one would further have

$$\text{Prob (generate all } p_{ii} > 0.7) = 0.3^r. \quad (29)$$

Or, in another form,

$$\text{Prob (generate a } \mathbf{P}_S \text{ that is not reasonably acceptable)} = \text{Prob (generate at least one } p_{ii} < 0.7) = 1 - 0.3^r \quad (30)$$

Thus, if the $\mathbf{P}_S$ generating process was not controlled by the reasonably acceptable condition (Equation 18), the probability of producing a reasonably acceptable $\mathbf{P}_S$ would be only 0.3^r (where r is the number of classes). It is more likely that the simulated transition matrix could not be used as a forward conditional confusion probability matrix for a reasonably acceptable classifier. In Czaplewski and Catts' simulation (Czaplewski and Catts, 1989), r had been set to range from 4 to 21. If $r = 4$, then $0.3^r = 0.008$; if $r = 21$, then $0.3^r = 1.046$

$\times 10^{-11}$. One may observe that even if r is small, the chance of getting a reasonable $\mathbf{P}_S$ is already very small; if r is large, the chance of getting a reasonable $\mathbf{P}_S$ is almost zero! Therefore, if condition (18) is not adopted in generating $\mathbf{P}_S$, the simulation results would definitely not be in favor of inverse correction. This observation might also help in interpreting the phenomena presented in Figure 1 of Czaplewski and Catts' paper (Czaplewski and Catts, 1989): the smaller the r, the better the simulation results for inverse correction.

However, the discussion here is only a conjecture. We do think the inverse correction method needs further evaluation. And, we do hope the natural constraints for image classifiers such as conditions (16) or (18) would be considered in future evaluations.

Multiscene Corrections

The constraints in conditions (15), (16), or (18) are consistent in terms of multi-scene classification assemblies. Therefore, the marginal area estimation correction for a large region that consists of multiple scenes can be achieved through the correction for individual scenes. One way to do this is assuming that there is a "super" composite classifier that acts on all scenes respectively. This composite classifier has the effect that it is equal to the local classifier whenever it is applied to a particular scene. For the composite classifier, one can still perform area estimation and correction for the whole region if the sampling design is the same.

Assume that there are m scenes in the considered region. The population confusion matrix for scene k ($k = 1, 2, ..., m$) is

$$\mathbf{C}_P^{(k)} = \begin{pmatrix} N_{11}^{(k)} & N_{12}^{(k)} & \cdots & N_{1r}^{(k)} \\ N_{21}^{(k)} & N_{22}^{(k)} & \cdots & N_{2r}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ N_{r1}^{(k)} & N_{r2}^{(k)} & \cdots & N_{rr}^{(k)} \end{pmatrix} \qquad k = 1,2,\ldots,m. \tag{31}$$

Then the total population confusion matrix $\mathbf{C}_P$ is

$$\mathbf{C}_P = C_P^{(1)} + C_P^{(2)} + \ldots + C_P^{(m)} \tag{32}$$

Let $N_{i.}^{(k)}$ be the total number of pixels in class i in scene k, and $N_{i.}$ be the total number of pixels in class i for the whole region (m scenes). We will have

$$N_{i.}^{(k)} = \sum_{j=1}^{r} N_{ij}^{(k)} \qquad N_{.j}^{(k)} = \sum_{i=1}^{r} N_{ij}^{(k)} \tag{33}$$

$$N_{i.} = \sum_{k=1}^{m} N_{i.}^{(k)} \qquad N_{.j} = \sum_{k=1}^{m} N_{.j}^{(k)} \tag{34}$$

$$\sum_{i=1}^{r} N_{i.} = \sum_{j=1}^{r} N_{.j} = N \tag{35}$$

where N = the total number of pixels in m scenes in the whole region.

The estimate for $N_{i.}$ can be obtained by taking the sum of the estimates for the individual $N_{i.}^{(k)}$s. The corrections for these estimates may also be done based on the confusion matrices of the "super" composite classifier.

In the case of inverse correction, the inverse of the forward conditional confusion probability matrix is involved. The invertibility of the forward confusion probability matrices for the composite classifiers needs to be confirmed. Assume $\mathbf{P}_P^{(k)}$ and $\mathbf{P}_S^{(k)}$ are the forward population and sample confusion probability matrices for scene k; $\mathbf{P}_P$ and $\mathbf{P}_S$ are the corresponding population and sample forward confusion probability matrices for the composite classifier:

$$\mathbf{P}_P^{(k)} = \begin{pmatrix} P_{11}^{(k)} & P_{12}^{(k)} & \cdots & P_{1r}^{(k)} \\ P_{21}^{(k)} & P_{22}^{(k)} & \cdots & P_{2r}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{r1}^{(k)} & P_{r2}^{(k)} & \cdots & P_{rr}^{(k)} \end{pmatrix} \quad \mathbf{P}_S^{(k)} = \begin{pmatrix} p_{11}^{(k)} & p_{12}^{(k)} & \cdots & p_{1r}^{(k)} \\ p_{21}^{(k)} & p_{22}^{(k)} & \cdots & p_{2r}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1}^{(k)} & p_{r2}^{(k)} & \cdots & p_{rr}^{(k)} \end{pmatrix} \tag{36}$$

$$\mathbf{P}_P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1r} \\ P_{21} & P_{22} & \cdots & P_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ P_{r1} & P_{r2} & \cdots & P_{rr} \end{pmatrix} \quad \mathbf{P}_S = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix} \tag{37}$$

The population and sample probabilities that a pixel in the given class is from scene k are

$$\frac{N_{i.}^{(k)}}{N_{i.}} \quad \text{and} \quad \frac{n_{i.}^{(k)}}{n_{i.}}. \tag{38}$$

The population and sample (mis)classification probabilities for the composite classifier would be

$$\begin{aligned} P_{ij} &= \text{Prob}_{\text{population}} \text{ (Map class} j \mid \text{True class} i) = \sum_{k=1}^{n} \frac{N_{i.}^{(k)}}{N_{i.}} P_{ij}^{(k)} \\ p_{ij} &= \text{Prob}_{\text{sample}} \text{ (Map class} j \mid \text{True class} i) = \sum_{k=1}^{n} \frac{n_{i.}^{(k)}}{n_{i.}} p_{ij}^{(k)} \end{aligned} \tag{39}$$

As usual, the population forward conditional confusion probability matrix for the composite classifier is estimated by the sample forward conditional confusion probability matrix for the composite classifier.

To guarantee the sample forward confusion probability matrix for the composite classifier obtained this way is invertible, some conditions need to be added to the local classifiers. A reasonable sufficient condition would be requiring that all the local classifiers be minimum practically acceptable (or reasonably acceptable).

It is easy to prove that, if all $\mathbf{P}_S^{(k)}$s are minimum practically acceptable (i.e., strictly diagonally dominant), then $\mathbf{P}_S$ is also minimum practically acceptable (strictly diagonally dominant), because

$$\begin{aligned} \sum_{j \neq i,\ 1 \le j \le r} p_{ij} &= \sum_{j \neq i,\ 1 \le j \le r} \sum_{k=1}^{n} \frac{N_{i.}^{(k)}}{N_{i.}} p_{ij}^{(k)} \\ &= \sum_{k=1}^{n} \frac{N_{i.}^{(k)}}{N_{i.}} \sum_{j \neq i,\ 1 \le j \le r} p_{ij}^{(k)} \qquad i=1,2,\ldots,r. \\ &< \sum_{k=1}^{n} \frac{N_{i.}^{(k)}}{N_{i.}} p_{ii}^{(k)} \\ &= p_{ii} \end{aligned} \tag{40}$$

This suggests that, if all the local classifiers are acceptable, then the composite classifier is also acceptable. This guarantees that the inverse correction can be performed for the composite classifier. Similar results can be obtained if the classifiers are required to be reasonably acceptable. Furthermore, it indicates that the constraints for the classifiers in conditions (15) through (18) are "natural."

Summary

Through our discussion we can conclude that (1) the inverse correction exists if the classifier is minimum practically acceptable and (2) the inverse is not ill-conditioned (i.e., it is stable) if the classifier is reasonably acceptable.

Marginal area estimate corrections have long been used to obtain accurate area estimates for land-cover types. Techniques used for corrections are commonly application practice dependent. Jupp (1989) pointed out that which method is more "natural" depends on the circumstances of the interpretation and survey and involves the fundamentals of survey design.

The inverse area estimate correction has been practically used in remote sensing applications for some time. Though large errors in classification scheme could result in large uncertainties in inverse correction, for most reasonably acceptable classification schemes, inverse correction still provides reliable and stable area estimates. However, care should be taken when the accuracy of the classifier is particularly low.

Comparison work for different correction methods is still needed. However, a meaningful comparison should be based on the fact that the inverse corrections are largely based on reasonably acceptable classifiers; and, at least, based on the minimum practically acceptable classifiers.

Acknowledgment
The author wishes to thank the anonymous reviewers for their valuable comments on the first draft of the paper.

References

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

Burk, T.E., M.E. Bauer, A.R. Ek, and S.C. Ahearn, 1988. Satellite inventory of Minnesota's forest resource, *Proceedings of the IUFRO S4.02.05 Meeting,* Hyytiala, Finland, 29 August-2 September, pp. 43–52.

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

Chrisman, N.R., 1982. Beyond accuracy assessment: correction of misclassification, *Proceedings 5th International Symposium on Computer-Assisted Cartography,* Crystal City, Virginia, 22–28 August, pp.123–132.

Czaplewski, R.L., 1992. Misclassification bias in areal estimates, *Photogrammetric Engineering & Remote Sensing,* 58:189–192.

Czaplewski, R.L., and G.P. Catts, 1990. Calibrating area estimates for classification error using confusion matrices, *Proceedings 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 classifiers estimate area? *International Journal of Remote Sensing,* 13:1735–1742.

Golub, G.H., and C.F.V. Loan, 1989. *Matrix Computations,* The Johns Hopkins University Press, Baltimore, Maryland, pp. 79–81, p. 120.

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

———, 1989. Global estimates from a confusion matrix; a reply to Jupp, *International Journal of Remote Sensing,* 10:1571–1573.

Horn, R.A., and C.R. Johnson, 1985. *Matrix Analysis,* Cambridge University Press, New York, p. 349, p. 295.

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

Maselli, F., C. Conese, G. Zipoli, and M.A. Pittau, 1990. Use of error probabilities to improve area estimate based on maximum likelihood classifications, *Remote Sensing of Environment,* 31:155–160.

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

Minc, H., 1986. *Nonnegative Matrices,* John Willey & Sons, New York, 146 p.

Poso, S., 1988. Seeking for an optimal path for using satellite imageries for forest inventory and monitoring, *Proceedings of IUFRO S4.02.05 Meeting,* Hyytiala, Finland, 29 August–2 September.

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.

Thomas, R.W., 1986. Utility of remote sensing data in renewable resource sample survey, *Proceedings of the 1986 International Geoscience and Remote Sensing Symposium,* Zurich, 8–11 September, pp. 755–758.

(Received 1 March 1994; accepted 29 November 1994; revised 17 February 1995)