Performance of Adaptive Sampling Design with a Nested Area Sampling Frame for Binary Maps

G.P. Patil, S. Ray and C. Taillie

Center for Statistical Ecology and Environmental Statistics
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

Abstract: With the growing use of remotely sensed data, accuracy assessment of classified maps based on such data has become a pressing issue. The assessment is based on site-specific comparison of the classified map with reference information (e.g., "ground truth"). But wall-to-wall reference data is seldom feasible so the assessment employs a sample of locations ("pixels"). This paper is concerned with the choice of efficient sampling design and, in particular, the effectiveness of an adaptive sampling strategy.

Several measures of thematic accuracy have been proposed but the most widely used is the overall accuracy, which is the proportion p of pixels that are in error.For purposes of estimating p, the classified map can be conceptually replaced by a binary map in which each pixel carries one of two possible values: the value 0 indicates a correct classification for that pixel while the value 1 indicates an error. The overall accuracy is therefore the proportion of ones across the conceptual binary map. It is expected that p will be relatively small but that the error pixels are spatially clustered. If one had some prior information about the probable location of error pixels, then a stratified design would be appropriate. When such prior information is not available but where the errors are spatially clustered, then an adaptive strategy might be effective in identifying and intensively sampling the high error regions.

This paper studies the effectiveness of an adaptive strategy using a two-level nested grid design. The initial sample of N_A pixels draws one pixel from each top-level grid cell. A secondary sample of N_B pixels is drawn from each top-level grid cell whose initial sample indicates the presence of error. Pixels in the secondary sample are then assessed for error/non-error.Then we find the unbiased estimator of the population proportion via Rao-Blackwellization of the first stage estimator.

CSEES Technical Report 2000-0720.
Prepared with partial support from the United States Environmental Protection Agency Cooperative Agreement Number CR-825506. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of Agency and no official endorsement should be inferred.

The values of 1 in the binary map are replaced by numerical responses drawn from a skewed distribution with mean 1 but variance σ^2. We study the effectiveness of the adaptive strategy as a function of σ^2. The new map with continuous responses reduces to the original binary map in the limit as σ^2 goes to zero. The results clearly show that it is not possible to gain from the second stage of adaptive sampling in the error-nonerror (binary) case, but adaptive sampling becomes very efficient if the non-zero units of the population have large dispersion. We investigate the results in a general post-classification versus pre-classification set up for continuous spectral response with varying coefficient of variation.

Keywords: Accuracy assessment, global and local clustering, minimum variance unbiased estimator, neighborhood sampling, post-clasiification vs pre-classification setup, Rao-Blackwell, statistical sufficiency, two-stage adaptive sampling.

1. Introduction

With the growing use of remotely sensed data, accuracy assessment of classified maps based on such data has become a pressing issue. The assessment is based on site-specific comparison of the classified map with reference information (e.g., "ground truth"). But wall-to-wall reference data is seldom feasible so the assessment employs a sample of locations ("pixels"). This paper is concerned with the choice of efficient sampling design and, in particular, the effectiveness of an adaptive sampling strategy.

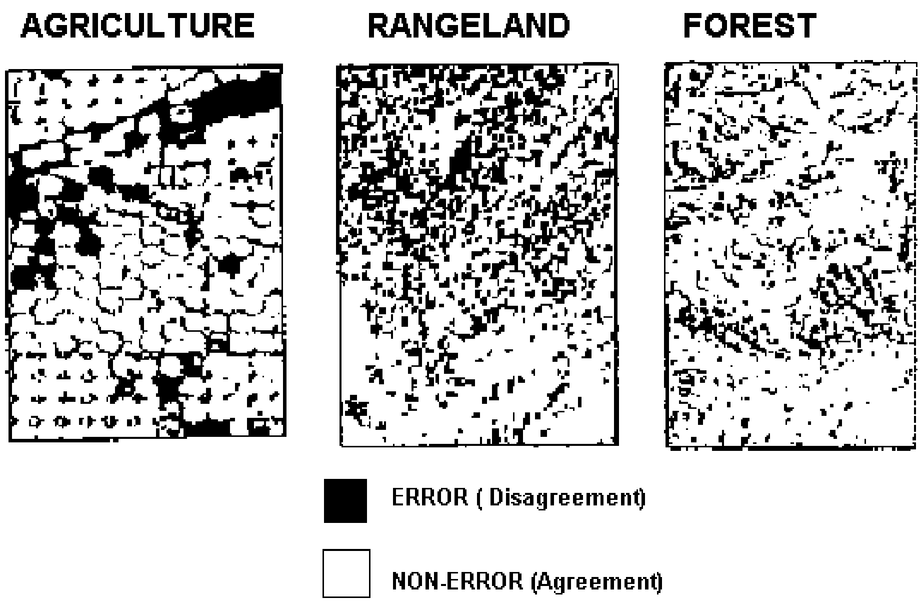

Figure 1: Patterns of misclassified(error) pixels in different eco-regions

Several measures of thematic accuracy have been proposed but the most widely used is the overall accuracy, which is the proportion p of pixels that are in error.For purposes of estimating p, the classified map can be conceptually replaced by a binary map in which each pixel carries one of two possible values: the value 0 indicates a correct classification for that pixel while the value 1 indicates an error. The overall accuracy is therefore the proportion of ones across the conceptual binary map. It is expected that p will be relatively small but that the error pixels are spatially clustered. If one had some prior information about the probable location of error pixels, then a stratified design would be appropriate. When such prior information is not available but where the errors are spatially clustered, then an adaptive strategy might be effective in identifying and intensively sampling the high error regions.

This paper studies the effectiveness of an adaptive strategy using a two-level nested grid design. The initial sample of N_A pixels draws one pixel from each top-level grid cell. A secondary sample of N_B pixels is drawn from each top-level grid cell whose initial sample indicates the presence of error. Pixels in the secondary sample are then assessed for error/non-error.Then we find the unbiased estimator of the population proportion via Rao-Blackwellization of the first stage estimator.

The values of 1 in the binary map are replaced by numerical responses drawn from a skewed distribution with mean 1 but variance σ^2. We study the effectiveness of the adaptive strategy as a function of σ^2. The new map with continuous responses reduces to the original binary map in the limit as σ^2 goes to zero. The results clearly show that it is not possible to gain from the second stage of adaptive sampling in the error-nonerror (binary) case, but adaptive sampling becomes very efficient if the non-zero units of the population have large dispersion. We investigate the results in a general post-classification versus pre-classification set up for continuous spectral response with varying coefficient of variation.

Most of the work in sampling of a rare trait has involved people with a concept of referral invoked (Christman, 2000a). This has given rise to network sampling, neighborhood sampling and similar adaptive sampling procedures. But a survey of a rare classification error in a geographical region is a completely different scenario.

2. Sampling procedure

Consider a population consisting of N units/pixels. Let p be the true proportion of pixels which have the rare trait of being in error.

Let $N = N_A N_B N_C$ and let the population be partitioned into N_A strata, each having $N_B N_C$ units. Let each primary or top-level stratum a be further divided into N_B sub-strata each consisting of N_C units.

The sampling design is a *two stage adaptive sampling design* (Thompson, 1991a,b, Thompson and Seber, 1996). One unit is selected at random from each top level stratum and the response on that unit is noted. If the response is 0, there is no further sampling within the top level stratum. Otherwise, one unit is selected at random from each of its N_B sub-strata. Note that if the spatial pattern within the top level stratum is more or less homogeneous, one can take N_B samples from the stratum instead of dividing it into N_B substrata and taking one sample from each substratum.

3. Assessing the effectiveness of the second stage sampling in a two stage design

We want to estimate a population parameter θ on the basis of a two-stage sampling design. Let $\hat{\theta}_I$ be an estimator of θ using only the first stage information and $\hat{\theta}_{I,II}$ the estimator using the information of both the stages. We want to verify whether the second stage gives extra information or we should engage all our resources to the first stage sampling.

Thus we check the information contributed by each stage of sampling. Let I_I and I_{II} be the information provided by the first and second stage of sampling. We define information as inverse of the variance or MSE of the estimator so that

$$\begin{aligned} N_I.I_I &= \frac{1}{MSE(\hat{\theta}_I)} &\quad (1)\\ N_I.I_I + N_{II}.I_{II} &= \frac{1}{MSE(\hat{\theta}_{I,II})} &\quad (2) \end{aligned}$$

where N_I and N_{II} are the expected sample size of the first and the second stage of sampling

Here $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$ Then we define the relative effectiveness of the second stage

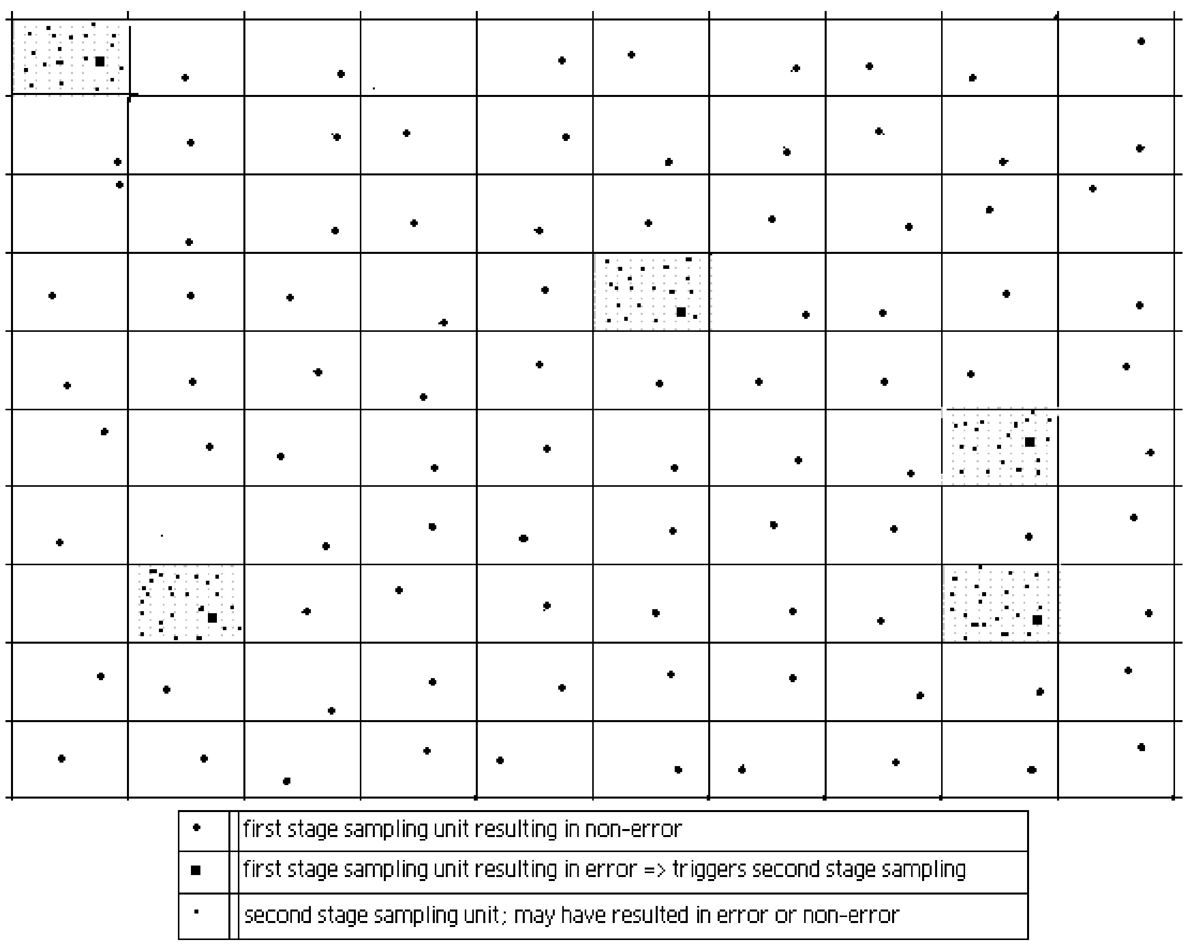

Figure 2: Simulation result showing a possible two stage adaptive sample

sampling as

$$e = \frac{I_{II}}{I_I} \quad (3)$$

Also if we have a quantitative cost structure then we can incorporate the cost efficiency in (3) by

$$e' = \frac{I_{II}/C_{II}}{I_I/C_I} \quad (4)$$

Solving(1) and (2) we get

$$e = \frac{N_I}{N_{II}} \left[\frac{MSE(\hat{\theta}_I)}{MSE(\hat{\theta}_{I,II}}) - 1 \right] \quad (5)$$

$$e' = \frac{N_I.C_I}{N_{II}.C_{II}} \left[\frac{MSE(\hat{\theta}_I)}{MSE(\hat{\theta}_{I,II}}) - 1 \right] \quad (6)$$

$$\text{or} \quad e' = \frac{TC_I}{TC_{II}} \left[\frac{MSE(\hat{\theta}_I)}{MSE(\hat{\theta}_{I,II})} - 1 \right] \quad (7)$$

where TC_i denotes the total cost in $i^{\underline{th}}$ stage of sampling.

We will investigate the performance of two stage adaptive sampling in two different ways. Still our main goal is to estimate the proportion of errors. So in the binary map situation, at first we will estimate the proportion without any confounding. In the second case, we will replace the 1's by an observation from a log-normal ($LN(\mu, \sigma^2)$) where $\mu + \frac{\sigma^2}{2} = 0$, so that the mean of the log-normal is 1. The 0's in the image are left as 0's. Thus the response in the new map follows a mixture distribution and the mean pixel response is

$$\theta = 0 + \tilde{p}.\bar{Y}$$

where $\tilde{p}$ is the proportion of 1's in the generated map and $\bar{Y}$ is the average of the generated LN values. The effectiveness parameter $e = \frac{I_{II}}{I_I}$ depends on the proportion and location of 1's as well as the generated log-normal values. The super-population ξ has two components; the generation of the binary image and generation of the log-normal values. Note that

$$\begin{aligned} E_\xi[\,\theta\,] &= E[\,\tilde{p}\,]E[\bar{Y}] \\ &= p \qquad \text{since} \qquad E[\bar{Y}] = e^{\mu+\frac{\sigma^2}{2}} = 1. \end{aligned}$$

Now $E_\xi[\frac{I_{II}}{I_I}]$ should depend on σ^2 and p and the spatial dependence parameters. We will investigate the dependence on σ^2. Note that the case of not changing the 1's into a log-normal is the limiting case when $\sigma^2 \to 0$. Thus in particular we will investigate,

1. Is $E_\xi[\frac{I_{II}}{I_I}]$ a monotone increasing function in σ^2 or some other parameters of the distribution?
2. When is $E_\xi[\frac{I_{II}}{I_I}] > 1$?
3. Is $\lim_{\sigma^2 \to 0} E_\xi[\frac{I_{II}}{I_I}] < 1$?

4. Rao-Blackwell Two-stage Adaptive Estimator

As each of the first stage unit is chosen at random from each of the top-level stratum, the estimation of p reduces to the estimation of p_a, the proportion of 1's in the top level stratum a. Thus the estimate and variance of p can be obtained from the estimate and variance of p_a's.

We use the following notation to study the estimation of p_a within a given top-level stratum a:

$Y_{1,a}$	response on the primary sampling unit of stratum a
δ_a	trigger for secondary sampling within stratum a. Note that for our case $Y_{1,a}$= δ_a.But this would not be the case in more complex multi-category settings.
$\mathbf{Y}_{2,a}$	(Virtual) vector of responses of secondary sample in stratum a. Note that there are N_B elements in $\mathbf{Y}_{2,a}$.

We now find the optimum estimator by Rao-Backwellizing the unbiased estimator given the sufficient statistics (Thompson and Seber, 1996).For a particular stratum a the first stage sampling $Y_{I,a}$ is an unbiased estimator of the stratum mean (proportion). But if that triggers the second stage of sampling and we investigate the N_B second stage samples, the sufficient statistics are the unordered $N_B + 1$ samples which we denote by

$$\begin{aligned} \mathbf{Y}_{R,a} &= Y_{I,a} \cup \mathbf{Y}_{2,a} \\ &= \left\{ Y_{R,a_1}, Y_{R,a_2}, \ldots, Y_{R,a_{N_B+1}} \right\} \quad . \end{aligned}$$

We have

$$\hat{p_a}^{RB} = E[Y_{I,a}|\mathbf{Y}_{R,a}]$$

$$= \frac{\sum_{i=1}^{N_B+1} Y_{R,a_i} I_{[Y_{R,a_i}>0]}}{\sum_{i=1}^{N_B+1} I_{[Y_{R,a_i}>0]}} \tag{8}$$

$= $ mean of non-zero elements in the sample

where $I_{[x>0]}$ is the indicator of x being positive.

Note that, for the zero-one case, $\hat{p_a}^{RB} = 1$, which is same the as the first stage estimate. Let

$$\begin{aligned} n_a &= \text{\# non-zero pixels in stratum } a, \\ \tilde{p} &= \text{proportion of non-zero pixels in stratum } a, \\ \nu_a^* &= \text{\# non-zero pixels in secondary samples when triggered.} \end{aligned}$$

The estimate of the mean pixel response in stratum a is

$$\hat{\theta}_a = \begin{cases} 0 & \text{if no secondary sampling,} \\ \\ \frac{y_0+y_1+\ldots+y_{\nu_a^*}}{1+\nu_a^*} & \text{if the initial sample triggers secondary sampling.} \end{cases} \tag{9}$$

Here $y_0, y_1, \ldots, y_{\nu_a^*}$ are the nonzero numerical responses in the combined sample.
Let μ_a and σ_a^2 be the mean and variance of the non-zero responses in stratum a. The quantities μ_a and σ_a^2 are undefined when $n_a = 0$, in which case we take $\tilde{p}\mu_a = 0\tilde{p}\sigma_a^2$.
Let t_a indicate the triggering of secondary sampling.
Now

$$\begin{aligned} E[\hat{\theta}_a] &= E_{t_a} E[\hat{\theta}_a | t_a] \\ &= (1 - \tilde{p_a}) E[\hat{\theta}_a | t_a = 0] + \tilde{p_a} E[\hat{\theta}_a | t_a = 1] \\ &= \tilde{p_a} E_{\nu_a^*} E[\hat{\theta}_a | t_a = 1, \nu_a^*] \\ &= \tilde{p_a} E_{\nu_a^*}[\mu_a] \\ &= \tilde{p_a} \mu_a. \end{aligned}$$

Thus $\hat{\theta}_a$ is indeed unbiased.
The overall estimate of the mean pixel response is

$$\hat{\theta} = \frac{1}{N_A} \sum_a \hat{\theta}_a \quad ,$$

and this is also unbiased, and

$$\begin{aligned} MSE[\hat{\theta}] &= Var[\hat{\theta}] \\ \\ &= \frac{1}{N_A} \sum_a Var[\hat{\theta}_a] \quad . \end{aligned}$$

Thus

$$Var[\hat{\theta}_a] \quad = \quad \tilde{p_a} E\left[\frac{1}{1+\nu_a^*}\right]\sigma_a^2 + \mu_a^2 \tilde{p_a}(1-\tilde{p_a}) \quad . \tag{10}$$

Note that for a purely binary response, $\sigma_a^2 = 0$ and $\mu_a^2 = 1$, so that $Var[\hat{\theta}_a] = \tilde{p_a}(1-\tilde{p_a})$, which is the same as the first stage estimator.
Now

$$Var[\hat{\theta}_a] \quad = \quad C_a\sigma_a^2 + D_a\mu_a^2 \quad ,$$

$$\text{where} \quad C_a = \tilde{p_a} E\left[\frac{1}{1+\nu_a^*}\right] = \frac{1-(1-\tilde{p_a})^{N_B+1}}{N_B+1} \quad ,$$

$$\text{and} \quad D_a = \tilde{p_a}(1-\tilde{p_a}) \quad .$$

Thus the C_a and D_a depend only on the binary image or to be specific on $\tilde{p_a}$ and N_B. So the overall MSE for the two-stage sampling is

$$MSE[\hat{\theta}] \quad = \quad \frac{1}{N_A}\sum_a C_a\sigma_a^2 + \frac{1}{N_A}\sum_a D_a\mu_a^2 \quad , \tag{11}$$

and the MSE of the first-Stage sampling

$$MSE[\hat{\theta}] \quad = \quad \frac{1}{N_A}\sum_a C_a^*\sigma_a^2 + \frac{1}{N_A}\sum_a D_a\mu_a^2 \quad , \tag{12}$$

$$\text{where } C_a^* = \tilde{p_a}$$

Conditional on the binary image and the numerical log-normal values which replace the $1's$, we get for the effectiveness measure

$$e = \frac{I_{II}}{I_I} \quad = \quad \frac{1}{\tilde{p}N_B}\left[\frac{MSE(\hat{\theta}_I) - MSE(\hat{\theta})}{MSE(\hat{\theta})}\right]$$

$$= \quad \frac{1}{\tilde{p}N_B}\left[\frac{\sum_a (C_a^* - C_a)\sigma_a^2}{\sum_a C_a\sigma_a^2 + \sum_a D_a\mu_a^2}\right] \quad , \tag{13}$$

$$\text{where } \tilde{p} = \frac{1}{N_A}\sum_a \tilde{p_a} \quad .$$

$$\text{Since } (C_a^* - C_a) \quad = \quad \tilde{p_a} - \tilde{p_a}E\left[\frac{1}{1+\nu_a^*}\right] \geq 0 \quad ,$$

the effectiveness measure is always non-negative.

So we will assume that the non-zero pixel values are iid draws from a common continuous distribution with mean μ and variance σ^2. But the expected value of the effectiveness measure, $E[e]$ is not tractable, so we obtain a "typical" value for $\frac{I_{II}}{I_I}$ by replacing $MSE(\hat{\theta})$ and $MSE(\hat{\theta}_I)$ by their expected values and recomputing e.
Since

$$\begin{aligned} E[\sigma_a^2] &= \frac{n_a}{n_a - 1}\sigma^2 \quad , \\ E[\mu_a^2] &= \mu^2 + \frac{\sigma^2}{n_a} \quad , \end{aligned}$$

we get

$$e = \frac{I_{II}}{I_I} \; = \; \frac{1}{\tilde{p}N_B}\left[\frac{CV^2\sum_a(C_a^* - C_a)}{CV^2\sum_a C_a + CV^2\sum_a \frac{D_a}{n_a} + \sum_a D_a}\right] \quad , \tag{14}$$

$$\text{where } CV^2 = \frac{\sigma^2}{\mu^2} = \text{ squared coefficient of variation} \quad .$$

This is the "typical" effectiveness conditional on the binary image. It only depends on the Y-distribution through the CV^2. Further it is a monotone increasing function of CV^2 and the limiting cases are

$$\begin{aligned} \lim_{CV^2\to 0} e &= 0. \\ \lim_{CV^2\to\infty} &= \frac{1}{\tilde{p}N_B}\left[\frac{\sum_a(C_a^* - C_a)}{\sum_a[C_a + \frac{D_a}{n_a}]}\right]. \end{aligned}$$

Note that $\frac{D_a}{n_a} = \frac{\tilde{p_a}(1-\tilde{p_a})}{n_a}$ is taken to be zero when $n_a = 0$.
From the second limit above, we have the typical effectiveness statistic

$$e \;\leq\; \frac{1}{\tilde{p}N_B}\left[\frac{\sum_a(C_a^* - C_a)}{\sum_a[C_a + \frac{D_a}{n_a}]}\right].$$

It is not clear when this upper-bound becomes less than unity, stating that the second stage is not that much effective. But we can note that the upper bound is a decreasing function of C_a, and thus by putting $C_a = 0$, we get a conservative upper-bound given by

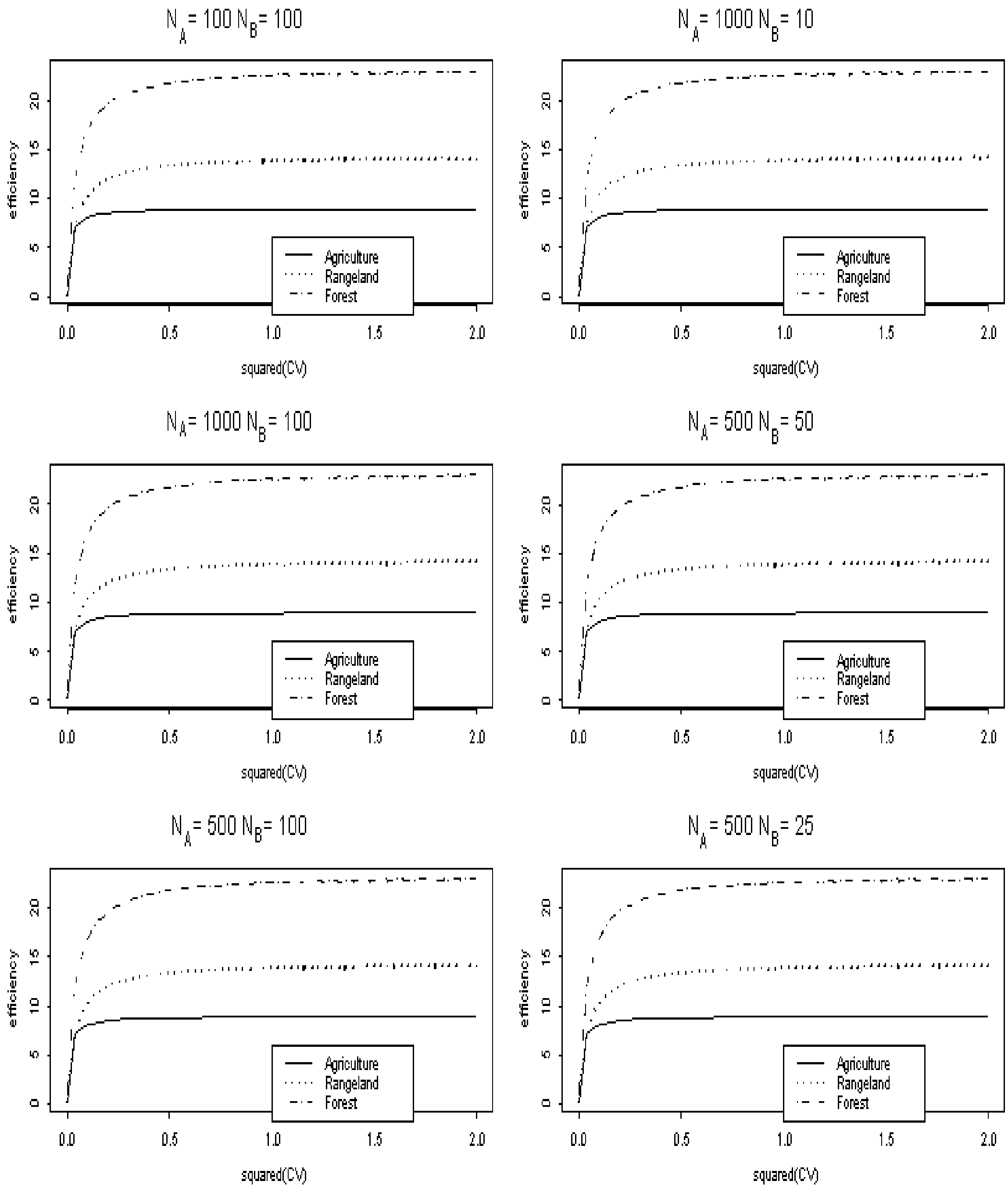

Figure 3: Efficiency of the second stage of adaptive sampling for different ecosystems as a function of CV^2 of the non-zero elements in the image with different sampling parameters

$$e \leq \frac{1}{\tilde{p}N_B}\left[\frac{\sum_a C_a^*}{\sum_a \frac{D_a}{n_a}}\right] = \frac{1}{\tilde{p}N_B}\left[\frac{\sum_a \tilde{p_a}}{\sum_a \frac{\tilde{p_a}(1-\tilde{p_a})}{n_a}}\right]$$

$$\leq \frac{N_A}{N_B \sum_a' \frac{n_a(1-\tilde{p_a})}{N_B N_C n_a}} = \frac{N_A N_B N_C}{N_B \sum_a'(1-\tilde{p_a})} = \frac{N}{N_B \sum_a'(1-\tilde{p_a})} \quad ,$$

where N is the total number of pixels in the image and $\sum_a'$ is the sum over all strata for which $\tilde{p_a} > 0$.

5. Post-Classification vs Pre-Classification for Adaptive Sampling

Adaptive sampling together with the Rao-Blackwell estimator gives us an efficient sampling scheme. It is most effective when the trait of interest has larger coefficient of variation. Simulation results and theory clearly show that, in the zero-one case or the error non-error case, the second stage of the adaptive sampling doesn't provide any extra information. Thus, for *post-classified errors* , the adaptive sampling scheme is not effective as in that case the nonzero elements of the image have a degenerate distribution. Figure 3 shows the change of effectiveness of the second stage of adapative sampling with the change of the coefficient of variation of the non-zero units in different eco-systems. Among the different eco-systems agriculture has the most heterogeneous set of p_a values and forest has the most homogeneous set of p_a values (see Figure 1). In all the cases however, the effectiveness calculated by Equation 14 increases as the coefficient of variation of the non-zero units increase. A striking feature in all the eco-systems and with all the sampling parameters is that the effectiveness rises very sharply near $CV^2 = 0$ but stabilizes very fast.

In the pre-classified case, however the error is not zero-one situation. The non-zero elements have a non-degenerate distribution. Thus if we work with pre-classified errors, we can gain efficiency from the second stage of the adaptive sampling scheme. Examples may be cited from the *"Love Canal"* data,where different regions have different amounts of contamination, but we may post-classify into contaminated or not contaminated according to some threshold. Instead of reducing the image into a zero-one map, we may work with the nondegenerate discrete or continuous response in the grayscale, resulting in a pixel-specific measure of error.

References

Christman, M. C. (2000a). A review of quadrat-based sampling of rare, geographically clustered populations. *J. Agricultural, Biological, and Environmental Statistics*, **5(2)**, 168–201.

Christman, M. C. (2000b). Adaptive two-stage one-per-stratum sampling. *Environmental and Ecological Statistics* (submitted).

Cochran, W. J. (1946). *Sampling Techniques*. Wiley, New York.

Thompson, S. K. (1991a). Stratified adaptive cluster sampling. *Biometrika*, **78(2)**, 389–397.

Thompson, S. K. (1991b). Adaptive cluster sampling: Designs with primary and secondary units. *Biometrics*, **47**, 1103–1115.

Thompson, S. K., Ramsey, F. L., and Seber, G. A. F. (1992). An adaptive procedure for sampling animal populations. *Biometrics*, **48**, 1195–1199.