Published as:

Edwards, T. C., Jr., G. G. Moisen, and D. R. Cutler. 1998. Assessing map uncertainty in remotely-sensed, ecoregion-scale cover-maps. Remote Sensing and Environment 63:73-83.

ASSESSING MAP ACCURACY IN A REMOTELY-SENSED, ECOREGION-SCALE COVER-MAP

Thomas C. Edwards, Jr.
USGS Biological Resources Division
Utah Cooperative Fish and Wildlife Research Unit
Utah State University, Logan, UT 84322-5210 USA

Gretchen G. Moisen
USDA Forest Service, Intermountain Research Station
507 25th Street, Ogden, UT 84401 USA

D. Richard Cutler
Department of Mathematics and Statistics
Utah State University, Logan, UT 84322-3900 USA

Abstract.--*Landscape- and ecoregion-based conservation efforts increasingly use a spatial component to organize data for analysis and interpretation. A challenge particular to remotely-sensed cover-maps generated from these efforts is how best to assess the accuracy of the cover-maps, especially when they can exceed 1,000s km^2 in size. Here we develop and describe a methodological approach for assessing the accuracy of large-area cover-maps, using as a test case the 21.9 million ha cover-map developed for Utah Gap Analysis. As part of our design process, we first reviewed the effect of intracluster correlation and a simple cost function on the relative efficiency of cluster sample designs to simple random designs. Our design ultimately combined clustered and subsampled field data stratified by ecological modelling unit and accessibility (hereafter a mixed design). We next outline estimation formulae for simple map accuracy measures under our mixed design and report results for 8 major cover-types and the 3 ecoregions mapped as part of the Utah Gap Analysis. Overall percent correct accuracy of the map was 83.2% (SE=1.4). Within ecoregions, accuracy ranged from 78.9% to 85.0%. Accuracy by cover-type varied, ranging from a low of 50.4% for barren to a high of 90.6% for man-modified. In addition, we examined gains in efficiency of our mixed design compared to a simple random sample approach. In terms of precision, our mixed design was more precise than a simple random design given fixed sample costs. We close with a discussion of the logistical constraints facing attempts to assess the accuracy of large area, remotely-sensed cover-maps.*

Key words: *accuracy assessment, Gap Analysis, cluster sampling, cover-maps, relative efficiency, sample design, remote sensing*

INTRODUCTION

Conservation efforts are increasingly focused at the landscape or ecoregion level rather than at localized sites. While the goals of these efforts vary, most use a spatial component to organize data for analysis and interpretation. One such effort, the National Biological Service's (NBS) Gap Analysis Program (Scott et al., 1993; Edwards et al., 1993; Edwards and Scott, 1995), uses land-cover/land-use classes and terrestrial vertebrate species as indicators of biological diversity and is designed to identify "gaps" in the protection of biological diversity. Once identified, gaps can be filled through land acquisition or changes in existing land-use practices.

Central to the goal of Gap Analysis is the creation of a thematic map depicting land-cover/land-use cover-types. In general, cover-types are derived by clustering reflectance values from spatially explicit locations (pixels) obtained from satellite imagery (Richards, 1986) and assigning them to cover classes based on field observations, aerial photographs, and existing maps (Scott et al., 1993). Uses of these types of cover-maps range from bioregional conservation planning (Merrill et al. in press) to more finely focused management efforts aimed at specific species (Homer et al., 1993; Stoms et al., 1993). Because Gap Analysis data are organized by states and ecoregions, such as the Great Basin, maps can easily cover thousands of square kilometers. The sheer size of areas modelled in Gap Analysis poses immense field and fiscal logistical difficulties that complicate any sample design for accuracy assessment. Nonetheless, estimates of the spatial and classification accuracy of Gap Analysis coverages are needed to assist land managers confronted with conflicting demands from user groups and to provide a defensible basis for use of the coverages in conservation decisions (Karieva, 1993; Edwards et al., 1996).

Assessing the accuracy of spatial data bases has received considerable attention in the literature, as witnessed by Veregin's (1989a) extensive annotated bibliography. Bolstad and Smith (1992) discuss both positional and attribute accuracy in a resource management setting. Lunetta et al. (1991) discussed how error propagates through the map-making process, beginning with data acquisition and flowing through analysis, conversion, and presentation of the final product. The taxonomy of error discussed by Veregin (1989b) also reflects the complexity of the issue. Various research agendas have been proposed for more creative error modelling strategies for geographic information systems (GIS). Chrisman (1989), Goodchild (1989), Openshaw (1989), and others have recognized the unique problems posed by combining data of a diverse nature collected at different scales and at various levels of error and uncertainty. Improved models of map error must be joined by better techniques to visualize spatial data base accuracy. Beard et al. (1991) discuss the research needed to develop these techniques.

Although a general conclusion of these studies is that more comprehensive models of thematic map error are needed, simpler measures of map accuracy are most common in the remote sensing literature. Frequently, sample data are used to construct a contingency table or "error matrix" from which many measures of thematic accuracy may be derived, including total percent of pixels (or other sample units) classified correctly, percent commission and omission error by class, and the Kappa statistic (Rosenfield, 1986; Story and Congalton, 1986; Congalton, 1991; Monserud, 1992; Green et al, 1993). Numerous studies have addressed the problem of choosing appropriate sample sizes and designs to assess classification accuracy (Berry and Baker, 1968; Hord and Brooner, 1976; Ginevan, 1979; Hay, 1979; Rosenfield, 1982; Congalton, 1991).

Congalton (1988a, b) discussed simple random, systematic and cluster sampling and the effect of spatial autocorrelation (dependency between neighboring "pixels", or units) on the efficiency of sample designs used for classification assessment. Stehman (1992) followed with a discussion of common misconceptions about systematic designs and appropriate criteria for evaluating various sampling schemes. In addition, Moisen et al. (1994) evaluated the relative efficiency of various sample designs for map accuracy assessment, applying a realistic cost function to data collection on a large-area map.

While a growing body of literature is available on assessing map accuracy, little seems to have been done to assimilate the wide variety of techniques into a single, broad-scale application. Issues of accuracy assessment include not only how to design and perform statistically sound sampling protocols and analyses, but also include consideration of how the map was modelled, travel time between sample units, logistical constraints in locating sample points, existing sources of ground truth data, and the potential uses of the map. Further, little attention has been paid to the obvious linkage between user needs and accuracy assessment. Because the technology supporting the construction of broad-scale vegetation maps is evolving so quickly, the problem itself is dynamic. Our goal, then, was not to come up with the ultimate solution to a rapidly changing problem, but to begin bridging some of the gaps between existing tools for assessing map accuracy and the unique logistical challenges posed by large-scale applications.

We briefly describe the building of a cover-map of Utah developed for the Utah Gap Analysis program (Edwards et al., 1995; Homer et al. in press). As part of our design process, we first reviewed the effect of intracluster correlation and a simple cost function on the relative efficiency of cluster sample designs to simple random designs. Our design is described next, coupled with points on how we attempted to work around logistical difficulties. This design ultimately combined clustered and subsampled field data stratified by ecological modelling unit and accessibility (hereafter a mixed design). We next outline estimation formulae for simple map accuracy measures under our mixed design and report results for cover-types and ecoregions mapped as part of the Utah Gap Analysis (Edwards et al., 1995). In addition, we examine gains in efficiency of our mixed design compared to a simple random sample (SRS) approach. We close with a discussion of the strengths and weaknesses of our mixed design, and a discussion of the unique logistical challenges faced when conducting an accuracy assessment of large area cover-maps.

METHODS

Cover-map development

A cover-map of Utah, ~219,000 km^2 in size, was developed from a state-wide Landsat Thematic Mapper (TM) mosaic created from 24 scenes at 30 m resolution. A total of 38 cover-types were modelled using a four step modelling approach. Steps included: (1) the creation of a statewide seamless mosaic of 14 TM images plus 10 pieces used to fill in cloud cover; (2) the subsetting of the mosaic into 3 ecoregions, the Basin and Range, Wasatch-Uinta and Colorado Plateau (after Omernik, 1987); (3) the association of 1,758 state-wide field training sites to spectral classes; and (4) the use of ecological parameters based on elevation, slope, aspect and location to further refine spectral classes representing multiple cover-types (see Homer at al., in press).

Once the mosaic was created, each ecoregion was independently classified, modelled and subsequently edgematched into a state-wide coverage. Spectral modelling was based on statistical relationships between field data and spectral information. A second phase of the modelling incorporated ancillary data to clarify cover-type associations by spectral class. Ancillary data focused on topographic and regional patterns of vegetation. Ancillary data used to support modelling included 3 arc-second resolution digital elevation data, slope, aspect, and region-specific vegetation polygons. Region-specific vegetation polygons were developed from existing literature and maps, and used to limit the geographic extent of some cover-types. Information from field data, literature, localized maps, and personal communications provided the input for specific ancillary modelling parameters. Environmental parameters collected from field training sites were summarized to enhance and verify published vegetation parameters. Further details on development of the Utah cover-map can be found in Homer et al. (in press) and Edwards et al. (1995).

Map scale is dependent on user needs. In Utah, for example, vegetation was modelled at a 1 ha base minimum map unit (MMU) and ecologically based aggregation routines used to create larger MMUs as needed (Bassett et al.; in review). Nationwide, the NBS Gap Analysis Program desires a 100 ha MMU since its primary focus is at ecoregion scales. In contrast, many regional and local conservation planning efforts require finer-grained mapping resolutions. As a consequence, the Utah 1 ha base modelling MMU represents a convenient starting point for assessing uncertainty that is independent of user needs.

Efficiency of cluster sampling

In general, cluster samples contain less information per unit sampled than simple random samples (SRS). On the other hand, it is often easier or less expensive to obtain samples from cluster sampling. This trade-off between sample unit cost and information content for large area cover-maps was previously illustrated by Moisen et al. (1994), who concluded that information return was sensitive to the cost of collecting data under a clustered versus a SRS.

The ratio of the variance of the proportion of misclassified sample units obtained under a SRS to the variance obtained from a sample of equivalent size under another design is known as "relative efficiency". Under cluster sample designs, relative efficiency is driven by intracluster correlation, ρ. Roughly speaking, ρ is a measure of the similarity of sample units within clusters. It determines the relative efficiency of cluster sampling to a simple random design; the larger the value of ρ, the more redundant the information content of the cluster is and the less efficient cluster sampling becomes.

Suppose within a strata or subpopulation there are N clusters each comprising M sampling units. The observation on the jth sampling unit in the ith cluster is denoted by y_{ij} and takes on the value of 1 if the sample unit is correctly classified and 0 if it is incorrectly classified. Finally, let P denote the true proportion of correctly classified sample units. We now define the intracluster correlation coefficient, ρ, by

$$\rho = \frac{2\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=j+1}^{M}(y_{ij}-P)(y_{ik}-P)}{(M-1)(NM-1)\sigma^2} \tag{1}$$

where

$$\sigma^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij}-P)^2}{(NM-1)} \quad . \tag{2}$$

Now consider a SRS of n' clusters within the stratum and compare it to a SRS of size $n'M$ in each stratum. Letting p denote the sample proportion of correctly classified sample units, and $V_{srs}(p)$ and $V_{clus}(p)$ the variances of p under the two sampling schemes, then the relative efficiency of the cluster sample to the SRS is given by

$$\frac{V_{srs}(p)}{V_{clus}(p)} = \frac{1}{(1+(M-1)\rho)} \tag{3}$$

This quantity is always <1 for positive ρ and the dependency of the efficiency on the intracluster correlation is clear.

Instead of comparing cluster and simple random samples of the same size, consider comparing samples of the same total cost according to the following cost functions. For a SRS, total cost was set equal to a product of the number of sample units and cost of traveling to, and collecting data on, each of those units. That is,

$$\text{cost} = n\,c_1 + n\,c_3 \tag{4}$$

where c_1 is the cost of traveling between sample units, c_3 is the cost of collecting data on the sample unit, and n is the number of sample elements in a SRS. For a sample of n' linear clusters, or primary units, each containing M units within a cluster, or secondary units, total cost can be expressed as, $\text{cost} = n_h' c_1 + n_h'(M_h - 1)c_2 + n_h'(M_h - 1)bc_2 + n_h' M_h c_3$ where c_2 is the cost of locating and traveling between units within a cluster and b is the proportion of c_2 it costs to travel back across the unit.

Multiplying (3) by the ratio $n'M/n$ gives a cost-sensitive relative efficiency (RE_c) that weighs redundancy of information in a cluster against larger quantities of data that become affordable through clustering. Specifically:

$$RE_c = \frac{V_{srs}(p)}{V_{clus}(p)} \; \frac{n'M}{n} \quad , \tag{6}$$

where

$$\frac{n'M}{n} = \frac{M(1+c_3/c_1)}{(1+(1+b)(M-1)c_2/c_1 + Mc_3/c_1)} \; . \tag{7}$$

When $n'M$ is >> n for a given cost, the variance of a cluster sample may become smaller than that of a SRS, making the cluster design more efficient for a fixed cost (see Moisen et al., 1994).

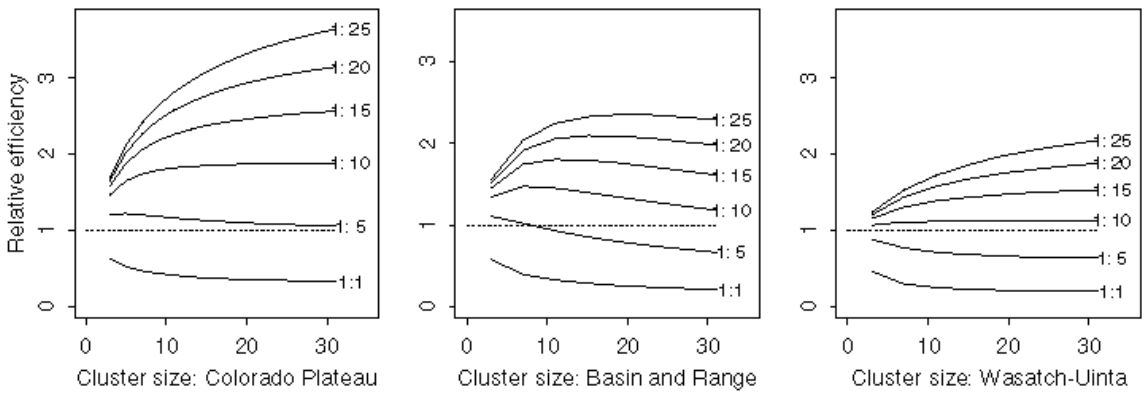

Figure 1. Intracluster correlation based on simulated error and relative efficiency as a function of different ratios of cost of traveling between points within a cluster (c_2) and costs of travleing between clusters (c_1).

To estimate RE_c for our specific application, we determined that the time to travel between and accurately locate sample points, c_2, within a linear cluster would be approximately 30 min, and that the time to identify and record existing cover types on the sample points, c_3, would be <10 min. Travel back across points in the cluster when finished, b, expressed as a proportion of c_2, was also expected to take <10 min. The cost of traveling between clusters in the state, c_1, was allowed to range from 30 min to over 12 h. Using these cost parameters and intracluster correlations obtained from simulated error patterns described in Moisen et al. (1994) for the Wasatch-Uinta, Basin and Range, and Colorado Plateau ecoregions, intracluster correlation and relative efficiency were plotted by linear cluster size in Figure 1. This figure illustrates for what cluster sizes and c_2 to c_1 ratios cluster sampling is more cost-effective than SRS (RE_c>1). While these three simulated error patterns do not necessarily yield the same level of intracluster correlation as that in the true error population, they serve as a tool for considering a range of scenarios. Given that the c_2 to c_1 ratio was expected to be well above 1:15 in Utah, we believed that logistical constraints, described later, would be more the limiting factors than intracluster correlation in determining a sensible cluster size.

Sample design

Vegetation in Utah was modelled by three ecoregions, each of which had different modelling algorithms and hence required separate assessment. For accuracy assessment, each ecoregion was divided into road, defined as a 1 km-wide corridor centered on a road, and offroad strata. Because of financial constraints and length of field season, we sought a design that maximized the amount of information that a 2-person field crew could collect in approximately 4.5 months. Preliminary work determined that in an average day, a crew could collect information on 20 sample points, 10 of those being clustered in the offroad stratum and 10 randomly distributed in

the road stratum. Of course, this would only hold if travel time between sample points was not excessive. To ensure relatively close proximity of sample points, 100 of 1435 Utah USGS quadrangle maps distributed proportionally among the 3 ecoregions were randomly selected for subsampling and placement of clusters.

Utah lands used for Department of Defense (DOD) maneuvers were excluded from the population upfront, and those land areas were not included the stratum weights. In addition, 185 of the 1435 (13%) Utah quadrangles were unavailable in orthophoto form and were precluded from the sample of 100. However, the precluded quadrangles followed what appeared to be a random pattern and we have assumed the 100 selected quadrangles are a random sample from the population. The choice of 10 linear clustered points was believed to be conservative based on earlier relative efficiency simulations by Moisen et al. (1994), and logistically sensible in terms of distance a crew would have to walk. The choice of the additional 10 road points provided a realistic target for work that could be completed in one day. Accordingly, a total of 20 sample points were evaluated on each quadrangle, 10 randomly selected within the 1 km-wide corridor and one randomly selected starting point for a linear cluster of 10 off-road sample points (Figure 2). The sample unit was defined as a 1 ha block, the base MMU for the cover-map.

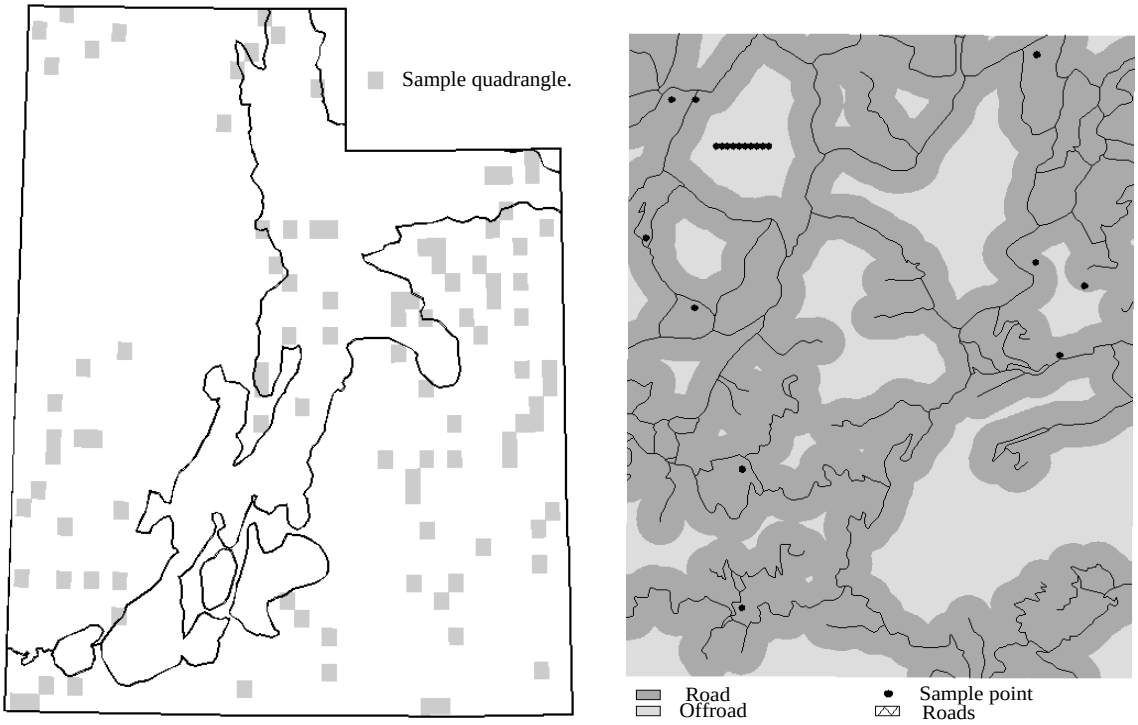

Figure 2. Location of sample quadrangles by ecoregion. Representative depiction of road and offroad strata, with random sample points located in the road strata and a linear cluster of 10 points in the offroad strata.

All sample points were randomly generated in the lab prior to field work. Selection of sample points was accomplished by creating a numbered 500 m grid and overlaying orthophoto quadrangles on top. Sample points were randomly selected from the grid, digitized into

coverages, and noted on the orthophoto quadrangles. The orthophoto quadrangles were carried into the field during data collection for aid in field orienteering.
Sample points were located with the aid of a GPS unit. Information collected at each sample point included: (1) cover-type within which sample point fell (Table 1); (2) primary and secondary cover-types within a 1 ha block surrounding sample point; (3) primary and secondary cover-types within 200 m of the sample point; and (4) a subjective measure of certainty with which cover-types were assigned in the field. In order to be considered a separate cover type, an area on the ground had to be at least 1 ha in size (base MMU of the cover-map) and 30 m wide (base pixel size of the TM data).

Table 1. Cover-types modelled and evaluated for accuracy in the Utah cover-map. Although the Utah cover-map includes a total of 38 cover-types, only forest, woodland, shrubland, herbaceous, man-modified, aquatic and barren classes are analyzed here.

Forest	Shrubland (cont.)
Aspen	Mountain Shrub
Aspen/Conifer	Oak
Lodgepole	Pickleweed Barrens
Lodgepole/Aspen	Sagebrush
Mountain Fir	Sagebrush/Perennial Grass
Mountain Fir/Mountain Shrub	Salt Desert Scrub
Spruce-Fir	Herbaceous community
Spruce-Fir/Mountain Shrub	Alpine
Woodland	Desert Grassland
Juniper	Dry Meadow
Lowland Riparian	Grassland
Mountain Riparian	Man-modified
Pinyon	Agriculture
Pinyon-Juniper	Urban
Ponderosa Pine	Aquatic community
Ponderosa Pine/Mountain Shrub	Water
Shrubland	Wetland
Blackbrush	Wet Meadow
Creosote-Bursage	Barren
Greasewood	Barren (<5% vegetation)
Maple	Lava

Measures of accuracy

We estimated the proportion of correctly classified units within each of the three ecoregions, six strata, and seven cover-types, as well as for the entire State map. The 6 strata, *h*, were based on the 3 ecoregions and the road and off-road categories and are defined as:

h =	1	Basin and Range ecoregion, road;
	2	Basin and Range ecoregion, offroad;
	3	Wasatch Uinta ecoregion, road;
	4	Wasatch Uinta ecoregion, offroad;
	5	Colorado Plateau ecoregion, road; and
	6	Colorado Plateau ecoregion, offroad.

On each of the 100 quadrangles, ten sample points, j, of 1 ha size were randomly selected in the road stratum, and ten were aligned in a linear cluster in the offroad stratum. Our design most closely resembles a stratified version of what Cochran (1977:303-305) calls subsampling with units of unequal sizes. However, it differs in two ways. First, quadrangles, or primary units, were randomly selected for subsampling then road and off-road strata identified within each of those quadrangles. Here, strata are not independent as would have been the case had quadrangles been selected independently for the road and off-road strata. Second, subsampling usually involves random selection of secondary units within each primary unit. While this is what was done in the road strata, a single cluster was selected in the off-road strata. We define y_{hij} to be the sample data collected within stratum h and quadrangle i, at point j. Here, y_{hij} has a value of one if the sample point is correctly classified. For this analysis, a point was considered correctly classified if the mapped cover-type matched the field call of primary cover-type within one ha of sample point. Misclassified points have a value of zero. The proportion of correctly classified units within stratum h, p_h, is estimated by a mean of all n_h quadrangle means, weighted by the area of stratum h within quadrangle i (M_{hi}), where

$$p_h = \frac{\sum_{i=1}^{n} M_{hi} p_{hi}}{n_h \sum_{i=1}^{n} M_{hi}}, \tag{8}$$

where

$$p_{hi} = \frac{1}{m} \sum_{j=1}^{m} y_{hij} \quad . \tag{9}$$

The formula for the variance of p_h under subsampling with units of unequal size accounts for variation both between and within quadrangles. In order to estimate within-quadrangle variation in the off-road strata, data would have had to been collected on more than one cluster. However, the contribution of within-quadrangle variation is weighted by the ratio of quadrangles selected for subsampling to the number of quadrangles in the population. Because this ratio was quite small, and because between-quadrangle variation was much larger than that within quadrangles, the variance for p_h was estimated using the formula

$$v\left(p_h\right) = \frac{\sum_{i=1}^{n_h} M_{hi}^2\left(p_{hi} - p_h\right)^2}{n_h(n_h - 1)\overline{M_h^2}}. \tag{10}$$

where M_h^2 is the mean area of stratum h within a quadrangle.

Given these within strata estimates and their variances, an estimate of the proportion of correctly classified units in the entire state map, p, is

$$p = \sum_{h=1}^{6} W_h p_h \tag{11}$$

where

W_h = ratio of the area in stratum h to total area in the state.

Caution must be exercised in computing the variance of p because data collected in the road and offroad strata within an ecoregion were collected from the same quadrangle maps and are not independent. Accounting for covariance between strata, the variance of p is

$$v\left(p\right) = \sum_{h=1}^{6} W_h^2 v\left(p_h\right) + 2\sum_{h=1,3,5} W_h W_{(h+1)} \mathrm{cov}\left(p_h, p_{(h+1)}\right) \quad , \tag{12}$$

where

$$\mathrm{cov}\left(p_h, p_{(h+1)}\right) = \frac{1}{\left(n_h^2\right)\left(n_h - 1\right)} \sum_{i=1}^{n_h}\left(p_{hi} - p_h\right)\left(p_{(h+1)i} - p_{(h+1)}\right). \tag{13}$$

Estimates of the proportion of correctly classified units in each of the three ecoregions are obtained by summing equations (11) through (13) over the two strata within each ecoregion. Estimates of proportion of correctly classified area by each of seven cover-types and three ecoregions are obtained by replacing y_{hij} in (9) with what we will label y_{khij} sample data collected within map class k, stratum h, and quadrangle i, at point j. As above, (11) through (13) are then summed over the two strata within each ecoregion.

RESULTS

Six of the initial 100 quadrangles were eliminated because of problems with access (i.e., private lands) and were replaced with other randomly selected quadrangles. Although every effort was made to ensure that the full 20 sample points (10 road, 10 off-road) were evaluated in every quadrangle, logistical constraints occasionally precluded field personnel from visiting every

point. Problems encountered were principally access, either an impassable road or uncertainties regarding ownership and our desire to avoid trespass on private lands without permission. An additional 4 quadrangles were eliminated because of inaccessibility, leaving a total of 96 quadrangles containing 960 road and 960 off-road points in 96 clusters (n=1920) (Appendix I). Like lands on quadrangles that were unavailable in orthophoto form, these inaccessible quads were assumed not to differ from the population at large.

Table 2. Percent accuracy (SE) by ecoregion.

Ecoregion	*% correct*	*(SE)*	n^a
Basin and Range	85.0	(3.2)	380
Wasatch-Uinta	78.9	(3.1)	540
Colorado Plateau	84.5	(1.6)	1000

[a]Includes observations in clusters.

Table 3. Percent accuracy (SE) by cover-type.

Cover-type	*% correct*	*(SE)*	n^a
01-Forest	74.2	(7.3)	114
02-Woodland	73.3	(4.3)	330
03-Shrubland	77.2	(2.8)	1009
04-Herbaceous community	75.0	(4.3)	292
05-Man-modified	90.6	(3.5)	118
06-Aquatic community	52.4	(11.9)	43
07-Barren	50.4	(17.1)	14

[a]Includes observations in clusters.

Although our design optimized field logistics, sample sizes were too small to adequately estimate accuracy for each ecoregion by cover-type combination. Instead, results are presented separately by ecoregion and cover-type. Within ecoregions, percent correct accuracy ranged from 78.9% to 85.0% (Table 2). Accuracy by cover-type varied, ranging from a low of 50.4% for barren to a high of 90.6% for man-modified (Table 3). Overall accuracy for the entire cover-map was 83.2% (SE=1.5).

To illustrate the gain realized through stratification and clustering, we considered how precise our estimates would have been had we performed our assessment using a SRS. Optimistically, a field crew could have visited an average of three randomly located sample points per day, yielding a total of 300 proportionally distributed among the three ecoregions. Given these sample sizes we estimated variance under a SRS and computed the ratio of v^2_{SRS} to v^2_{Mixed} (Table 4). Ratios were >1 in all cases, indicating that the estimates from our mixed design were more precise.

DISCUSSION

Several lessons were learned from this exercise. One is the importance of a clear articulation of the objective of the accuracy assessment. As noted by Hunter and Goodchild (1995), treatment of the kind of spatial error we modelled must be expanded to include a user-oriented perspective. As an example, Moisen et al. (1996) extended the analysis presented in this paper by fitting generalized linear mixed models to our original sample data. They explore the relationship between classification error in the blackbrush cover type of the vegetation cover-map of Utah, and elevation, aspect, and slope. Numerous approaches to modelling error exist (see

introduction), but it is important to consider needs of the end-user prior to selecting any single approach. Although meeting user needs is a primary concern, one consequence of the application of unique solutions to modelling error is the associated difficulty in comparing error statements among different maps evaluated for different purposes.

Table 4. Ratio of V^2_{SRS} to V^2_{Mixed} for the overall map and by ecoregion accuracy values. Ratios >1.0 indicate a more precise estimate of error from the mixed design.

Comparison	p	n^a	$V^2_{SRS}{}^b$	V^2_{Mixed}	V^2_{SRS} / V^2_{Mixed}
Overall	0.832	300	0.000467	0.000228	2.05
Basin and Range	0.850	102	0.001262	0.001026	1.23
Wasatch-Uinta	0.789	78	0.002162	0.000980	2.21
Colorado Plateau	0.845	120	0.001101	0.000285	3.86

[a] n=the number of samples realized under a SRS, distributed proportionally among the ecoregions.

[b] Estimated from $\frac{p(1-p)}{n-1}$, where p=estimated proportion of correct classifications.

In our case, we were interested in overall state, state by ecoregion, and state by ecoregion by cover-type estimates of accuracy. This necessarily complicated our design because each estimate could legitimately necessitate a different design. If our only objective was accuracy by cover-type, for example, it would have been more sensible to equally distribute sample points among cover-types. However, another realized problem with cover-maps of this size is computational limitations that can preclude prestratification and allocation of sample points to each cover-type. Basically, we did not have the computational resources to prestratify 21.9 million spatially explicit sample units and randomly select an equal sample size within each stratum. Moreover, this approach would have immediately compromised any attempt to maximize logistical efficiency. Hence, it was necessary to randomly distribute points within the framework of our design and examine the distribution of sample points within cover-types. We echo Congalton's (1991) recommendation that a primary objective of any accuracy assessment is to balance statistical recommendations with practical limitations such as those we encountered during our assessment of the Utah cover-map.

Another problem of concern is undersampling of certain cover-types. In our example, sample sizes for two cover-types, aquatic community and barren, were <100, a rule-of-thumb sample size proposed by Congalton (1991). In cases where sample sizes for specific cover-types are low, supplementary visits could be made to these cover-types and the additional samples required could be collected. Unfortunately, this information would be restricted to these cover-types only, and could not be combined with other cover-types to estimate accuracy in higher level groupings like ecoregions (Aronoff, 1982). An attempt to use these supplementary data changes the fundamental design to one stratified by cover-type, which requires that we be able to

prestratify the 21.9 million possible sample units by cover-type. As noted previously, we were unable to prestratify by cover-types for computational reasons.

The inability to use the supplementary samples for higher level accuracy estimates does not, however, pose significant problems. Although additional samples never hurt, sample sizes for higher level accuracy estimates from our mixed design were large enough for our estimates to be considered sufficiently precise. Collection of supplementary data that is focused on specific cover-types also helps to maintain the overall efficiency of our design. In most cases, low sample cover-types are likely to be localized and relatively rare components of any large area map. Consequently the effort required to travel to and sample these cover-types will be a small portion of the overall sample effort. This modification for dealing with undersampled cover-types is easy to make without compromising our analytical approach, suggesting our mixed design is flexible in terms of user objectives and its ability to deal with low sample cover-types.

Yet another lesson involved the choice of quadrangles as primary sample units. We had thought that sample units within quadrangles would be easier to collect than sample units spread over the entire state. In hindsight, quadrangle maps may not have been the best surrogate for assuring proximity of sample points. For example, in a topographically rugged state like Utah, it is not unusual for sample points within a quadrangle to be quite close linearly but distant in terms of access time because of geographic features and road layout. A more sensible approach might have been to have defined accessibility zones based on travel route. Instead of randomly selecting quadrangles, these travel zones could have been established a priori and the road and offroad strata distributed among zones. Alternatively, both clusters in the offroad strata, and single points in the road strata could have simply been randomly distributed throughout the state then later grouped for data collection efforts. Irrespective of the choice, a quadrangle- or travel zone-based selection of primary sample units, our mixed design is sufficiently flexible and provides a solid statistical basis for analysis.

The logistical constraints we encountered were varied but nontrivial. Consequently they cannot be dismissed, but must be considered when designing any sampling scheme for large-area cover-maps. For fieldwork, they included the cost of movement between sample points, the time and money necessary to support field assessment, and access, an increasingly difficult proposition as private landholders become more vocal in their opposition to resource surveys for fear of new-found information that might limit options on their land. In the case of access, use of data collected from clusters located on lands where permission for access has been granted should reduce the number of private landowners from whom permission must be acquired. By treating private/public lands as analogous to our offroad/road strata, relatively fewer parcels of private lands need be sampled to achieve reliable estimates of accuracy.

Additionally, accuracy of field geographic position systems (GPS) is variable, even when linked with use of orthophoto quadrangles or other types of photographs (August et al., 1994). GPSs also face a host of problems in remote, hilly areas and deep forests, with the signals necessary to locate position frequently difficult to obtain. Last, mapped classes are not necessarily mutually exclusive when viewed on the ground, creating an additional level of uncertainty (see Gopal and Woodcock, 1994).

The major constraint we considered in assessing the accuracy of ecoregion-scale vegetation cover-maps was excessive time of movement between sample sites. Although tempting to apply a SRS for selection of sample units, the cost of collecting information at locations randomly distributed across the entire state is, intuitively, substantially higher than collecting information at clustered locations. The mix of an off-road linear cluster and randomly selected points within a road corridor returned greater information per unit sampling cost for this broad-scale map for two reasons. First, sample units next to roads were easy to access, rapidly assessed, and cheap to collect, thereby boosting sample size. Second, use of a linear cluster for off-road points reduced sampling effort and reduced the per sample unit cost (see Moisen et al., 1994). Field personnel needed only to find a single starting sample point and locate subsequent sample points along a 1 km transect. When coupled with the flexibility described above, our mixed design provides a reasonable strategy for assessing accuracy of large area cover maps.

We thank T. Crowl, R. Czaplewski, S. Durham, T. B. Murphy, D. Roberts and S. Stehman, for their thoughtful reviews. Funding for this project was provided by the USDA Forest Service, Region 4, and the USDA Forest Service Intermountain Research Station, Ogden, Utah. We thank D. Born and D. Meyers, USDA Forest Service, Region 4, for their help in obtaining funding for this effort and clearing away bureaucratic barriers. D. Early deserves special mention for field work above and beyond the call of duty, having taken our truck to places where no person had gone before (and coming back, albeit with 11 fewer tires than he started with!). We also thank S. Bassett who can program anything, fast. Last, we thank C. G. Homer for continually reminding us that the best laid statistical plans fall apart as soon as they hit the road. He was right. This manuscript was prepared as part of the USGS Biological Resources Division's Gap Analysis Program, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, Utah.

REFERENCES

Aronoff, S. (1982), The map accuracy report: a user's view, *Photogram. Eng. Remote Sens.* 48:1309-1312.

August, P., J. Michaud, C. Labash, and C. Smith. (1994), GPS for environmental applications: accuracy and precision of locational data, *Photogram. Eng. Remote Sens.* 60:41-45.

Bassett, S. D., T. C. Edwards, Jr., and C. G. Homer. An attribute based algorithm for generalizing thematic maps. In review, *Int. J. GIS.*

Beard, K. M., S. Clapham, and B. P. Buttenfield. (1991), Visualization of spatial data quality. Technical Paper 91-26, Nat. Center Geographic Infor. Analysis, Univ. Calif, Santa Barbara, Calif.

Berry, B. J. L., and A. M. Baker. (1968), Geographic sampling. In *Spatial Analysis: a Reader in Statistical Geography* (B. J. L. Berry and D. F. Marble, Eds.) Prentice-Hall, Englewood Cliffs, New Jersey, pp 91-100.

Bolstad, P. V., and J. L. Smith. (1992), Errors in GIS: assessing spatial data accuracy, *J. Forest.* 90:21-29.

Chrisman, N. R. (1989), Modelling error in overlaid categorical maps. Pages 21-34 In *Accuracy of Spatial Databases* (M. Goodchild and S. Gopal, Eds.), Taylor and Francis, Philadelphia, Penn., pp21-34.

Congalton, R. G. (1988a), Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data, *Photogram. Eng. Remote Sens.* 54:587-592.

Congalton, R. G. (1988b), A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data, *Photogram. Eng. Remote Sens.* 54:593-600.

Congalton, R. G. (1991), A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sens. Environ.* 37:35-46.

Edwards, T. C., Jr., E. Deshler, D. Foster, and G. G. Moisen. (1996), Adequacy of wildlife habitat relation models for estimating spatial distributions of terrestrial vertebrates, *Conserv. Biol.* 10:263-270.

Edwards, T. C., Jr., J. M. Scott, C. G. Homer, and R. D. Ramsey. (1993), Gap analysis: a geographic approach for assessing national biological diversity, *Nat. Res. Environ. Issues* 2:65-72.

Edwards, T. C., Jr., C. G. Homer, S. D. Bassett, A. Falconer, R. D. Ramsey, and D. W. Wight. (1995), Utah Gap Analysis: an environmental information system, Technical Paper 95-1, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, Utah.

Edwards, T. C., Jr., and J. M. Scott. (1994), Use of Gap Analysis as a tool for the management of biodiversity, In *Proc. XXI Congr. Int. Union Game Biologists* (I. D. Thompson, Ed.), Vol. 1, pp 82-86.

Ginevan, M. E. (1979), Testing land-use map accuracy: another look, *Photogram. Eng. Remote Sens.* 45:1371-1377.

Green, E. J., W. E. Strawderman, and T. M. Airola. (1993), Assessing classification probabilities for thematic maps, *Photogram. Eng. Remote Sens.* 59:635-639.

Goodchild, M. F. (1989), Modelling error in objects and fields, In *Accuracy of Spatial Databases* (M. Goodchild and S. Gopal, Eds.), Taylor and Francis, Philadelphia, Penn., pp 107-113.

Gopal, S., and C. Woodcock. (1989), Theory and methods for accuracy assessment of thematic maps using fuzzy sets, *Photogram. Eng. Remote Sens.* 60:181-188.

Hay, A. M. (1979), Sampling designs to test land-use map accuracy, *Photogram. Eng. Remote Sens.* 45:529-533.

Homer, C. G., T. C. Edwards, Jr., D. H. Ramsey, and K. H. Price. (1993), Use of remote sensing methods in modeling sage grouse winter habitat, *J. Wild. Manage.* 57:78-84.

Homer, C. H., Ramsey, R. D., T. C. Edwards, Jr., and A. Falconer. 1997. Landscape cover-type modelling using a multi-scene TM mosaic. *Photogram. Eng. Remote Sens. 62.*

Hord, R. M., and W. Brooner. (1976), Land-use map accuracy criteria, *Photogram. Eng. Remote Sens.* 42:671-677.

Hunter, G. J., and M. F. Goodchild. (1995), Dealing with error in spatial databases: a simple case study, *Photogram. Eng. Remote Sens.* 61:529-537.

Karieva, P. (1993), No shortcuts in new maps. *Nature* 365:292-293.

Merrill, T., R. G. Wright, and J. M. Scott. In press. Using ecological criteria to evaluate wilderness planning options in Idaho, *Environ. Manage*.

Moisen, G. G., T. C. Edwards, Jr., and D. R. Cutler. (1994), Spatial sampling to assess classification accuracy of remotely sensed data, In *Environmental Information Management and Analysis: Ecosystem to Global Scales* (J. Brunt, S. S. Stafford, and W. K. Michener, Eds.), Taylor and Francis, Philadelphia, Penn., pp 161-178.

Moisen, G. G., D. R. Cutler, and T. C. Edwards, Jr. (1996), Generalized mixed models for analyzing error in a satellite-based vegetation map of Utah, In *Spatial Accuracy*

Assessment in Natural Resources (H. T Mower, R. L. Czaplewski, and R. H. Hamre, Eds.), USDA Forest Service General Technical Report RM-GTR-277, Ft. Collins, Co., pp 459-466.

Monserud, R. A., and R. Leemans. (1992), Comparing global vegetation maps with the Kappa statistic, *Ecol. Model.* 62:275-293.

Omernik, J. M. (1987), Map supplement: ecoregions of the conterminous United States, *Ann. Assoc. Am. Geogr.* 77:118-125 (map).

Openshaw, S. (1989), Learning to live with errors in spatial databases, In *Accuracy of Spatial databases* (M. Goodchild and S. Gopal, Eds.), Taylor and Francis, Philadelphia, Penn., pp 21-34.

Richards, J. A. (1986), *Remote Sensing Digital Image Analysis: n Introduction*, Springer-Verlag, New York.

Rosenfield, G. H. (1982), Sample design for estimating change in land use and land cover, *Photogram. Eng. Remote Sens.* 48:793-801.

Rosenfield, G. H. (1986), Analysis of thematic map classification error matrices, *Photogram. Eng. Remote Sens.* 52:681-686.

Rosenfield, G. H., K. Fitzpatrick-Lins, and H. Ling. (1986), Sampling for thematic accuracy testing, *Photogram. Eng. Remote Sens*. 52:397-399.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, S. Caicco, C. Groves, T. C. Edwards, Jr., J. Ulliman, H. Anderson, F. D'Erchia, and R. G. Wright. (1993), Gap analysis: a geographic approach to protection of biological diversity, *Wild. Monogr*. No. 123.

Stehman, S. V. (1992), Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data, *Photogram. Eng. Remote Sens.* 58:1343-1350.

Stoms, D. M., F. W. Davis, C. B. Cogan, M. O. Painho, B. W. Duncan, J. Scapan, and J. M. Scott. (1993), Geographic analysis of California Condor siting data, *Conserv. Biol.* 7:148-159.

Story, M., and R. G. Congalton. (1986), Accuracy assessment: a user's perspective, *Photogram. Eng. Remote Sens.* 52:397-399.

Veregin, H. (1989a), Accuracy of spatial databases: annotated bibliography, Technical Paper 89-9, Nat. Center Geographic Infor. Analysis, Univ. Calif, Santa Barbara, Calif.

Veregin, H. (1989b), A taxonomy of error in spatial databases, Technical Paper 89-12, Nat. Center Geographic Infor. Analysis, Univ. Calif, Santa Barbara, Calif.

Appendix I. Number of observations[a] by ecoregion and cover-type and corresponding confusion.

Ecoregion	Cover-type 01		02		03		04		05		06		07		Row
Cover-type	R	OR	R	OR	R	OR	R	OR	R	OR	R	OR	R	OR	Total
Basin and Range															
01-Forest	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2
02-Woodland	0	0	33	52	4	10	0	0	1	0	0	0	0	0	100
03-Shrubland	0	0	3	11	104	85	9	4	2	0	0	0	0	0	218
04-Herbaceous community	0	0	2	0	4	0	15	26	1	0	0	0	0	0	48
05-Man-modified	0	0	0	0	1	0	0	0	9	0	0	0	0	0	10
06-Aquatic community	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
07-Barren	0	0	0	0	1	0	0	0	0	0	0	0	1	0	2
Ecoregion total	0	1	38	64	105	90	23	25	13	0	0	0	1	0	380
Wasatch-Uinta															
01-Forest	37	47	2	1	4	3	5	2	0	0	0	0	1	0	102
02-Woodland	1	1	29	11	5	3	0	2	0	0	0	0	0	0	52
03-Shrubland	8	2	8	21	62	102	3	4	3	0	1	0	0	3	217
04-Herbaceous community	0	1	0	1	3	2	19	11	2	0	0	0	0	0	39
05-Man-modified	0	0	2	0	4	1	1	0	63	19	1	0	0	0	91
06-Aquatic community	0	0	0	0	1	2	0	1	1	0	2	30	0	0	37
07-Barren	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2
Ecoregion total	46	51	41	34	79	113	28	20	69	19	4	30	3	3	540
Colorado Plateau															
01-Forest	6	3	0	0	0	0	1	0	0	0	0	0	0	0	10
02-Woodland	5	1	75	75	12	7	2	0	1	0	0	0	0	0	178
03-Shrubland	4	2	18	23	231	258	12	15	3	0	0	0	6	2	574
04-Herbaceous community	0	0	12	3	10	5	72	96	3	0	0	0	1	3	205
05-Man-modified	0	0	0	0	1	0	0	0	16	0	0	0	0	0	17
06-Aquatic community	0	0	0	0	0	1	0	0	0	0	4	0	0	1	6
07-Barren	0	0	1	0	3	0	0	1	0	0	0	0	1	4	10
Ecoregion total	15	6	106	101	266	276	88	117	23	0	4	0	8	10	1000
Column total	61	58	185	199	450	479	139	162	105	19	8	30	12	13	1560

[a]Number of observations includes samples within clusters.