A Methodological Study for Accuracy Assessment of GAP Land Cover Maps

SARAH M. NUSSER[1], ERWIN E. KLAAS[2], CARSTEN H. BOTTS[1], AND ROBIN MCNEELY[3]
[1] Department of Statistics and Statistical Laboratory, Iowa State University, Ames, Iowa
[2] USGS Biological Resources Division, Iowa Cooperative Fish and Wildlife Research Unit, Iowa State University, Ames, Iowa
[3] Department of Animal Ecology, Iowa State University, Ames, Iowa

Introduction

Quantifying the accuracy of a GAP land cover map involves comparing the thematic content of the digital map with corresponding thematic reference data (i.e., some form of "truth") obtained from the field. Typically, assessment locations are selected from the target area, and reference data are gathered from field visits or photo-interpretation (Congalton 1991). Methods of selecting assessment locations vary widely from purposive sampling, in which areas are intentionally selected for observation without applying a randomization mechanism, to selecting statistical samples from the entire target area or from some portion of the target area (e.g., roadsides). Sampling units may be areas (polygons) or points on the land. To analyze assessment data, a number of accuracy measures are available to compare the reference data and land cover maps (Stehman 1997). The choice of accuracy assessment methodologies is influenced by scientific, statistical, and operational concerns.

Ideally, accuracy estimates are based on unbiased samples and statistical estimation methods that provide a measure of the precision of the estimated accuracy rate. However, practical considerations such as targeting sample locations while maintaining geographic spread, choosing the appropriate observational unit, obtaining access to sampled locations, and minimizing travel costs all present challenges when designing such studies. Sample survey methodologies provide a design and estimation framework that balances statistical and operational considerations with study objectives (Cochran 1977, Salant and Dillman 1994, Thompson 1992). Probability sample designs can be developed to target areas requiring more intensive study, avoid areas that are difficult to access, or select clusters of observation units to reduce study costs. Contact methods used in survey sampling provide an effective method of gaining access to private land and minimizing bias from nonresponse. Just as a questionnaire provides a rigorous basis for repeatability in telephone surveys, field observation methods are based on protocols that encourage well-defined observations at the correct location while minimizing the effort required to collect reference data. Estimators that take into account survey methods used in a study are readily available from this framework.

In response to a request from EPA Region 7 for an integrated accuracy assessment plan in the region, we designed and conducted a pilot study using a sample survey approach to assess the accuracy of GAP land cover maps. The goal was to produce a statistically sound and operationally feasible design that meets GAP's accuracy assessment objectives. In particular, we were interested in protocols for gaining permission to sample on private land, protocols for observing reference land cover in the field, appropriate sample design and estimation strategies, and quantifying the operational resources required to do a full accuracy assessment.

In this paper, we focus on the Iowa pilot study. We briefly summarize the methods we used to address scientific, statistical, and operational considerations, and present pilot study results. Further details are available in Nusser and Klaas (2001). Finally, we discuss the implications of this design for future accuracy assessment efforts.

Sample Design

The pilot study was conducted during the summer of 1999 in four northeast counties in Iowa: Allamakee, Clayton, Fayette, and Winneshiek.

A stratified two-stage cluster sample design (Lohr 1999) was used to select sample pixels for field visits from the four-county study area. We first selected USGS 7.5 degree quadrangles (or combinations of partial quads that fell on the border of the study area) as primary sampling units (PSUs) (Figure 1). Five strata of 8-12 PSUs each were created to ensure geographic spread of the PSUs and coverage of all land cover categories. Two PSUs were randomly selected from each stratum using systematic sampling, for a total of ten PSUs.

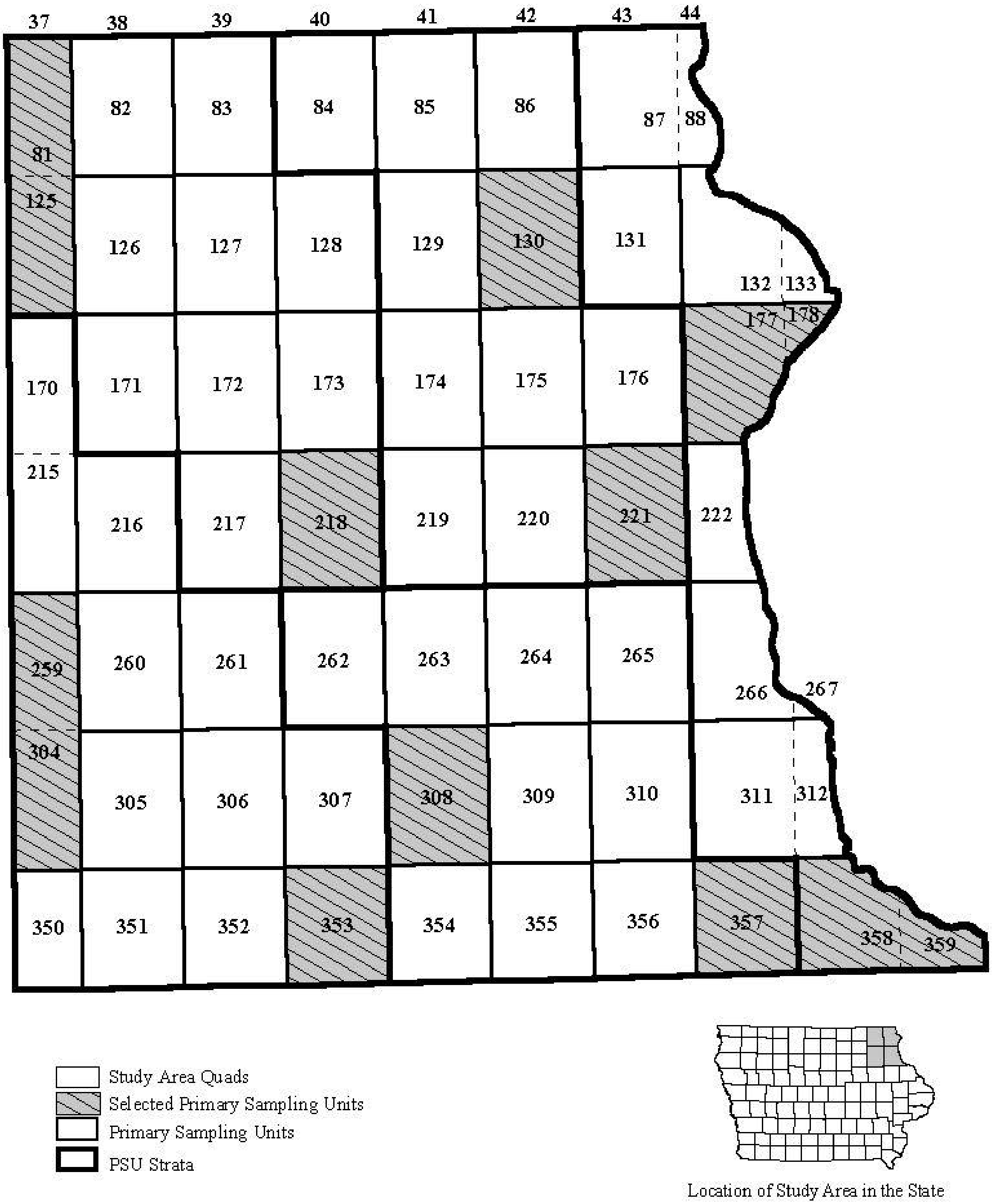

Figure 1. Accuracy assessment study area in Iowa, partitioned into quads and primary sampling units (PSUs), which are quads or combinations of partial and/or whole quads. Sampled PSUs are shaded.

Individual pixels were selected from PSUs in a second stage of sampling. Resource constraints dictated sample size. Iowa staff had a goal of visiting 200 points within the study area. Since we expected that access would be denied for approximately 15% of the sample points, 236 sample points were selected to achieve 200 responses. Pixel samples were selected from the ten PSUs using a stratified design. The pixel sample was stratified according to nine relatively homogeneous land cover categories, collapsed from the original 29 vegetation classes defined for Iowa (Table 1).

Table 1. Estimated accuracy rates by land cover category using nine-pixel cluster data.

Land Cover Category[a] (s)	Total Area with Consistent Field and Map Classifications (ha)	Estimated Field Area (ha)	Producer's Accuracy (%)			Map Area (ha)
			$\hat{PA}(s)$ [b]	(se)	n	
Coniferous Forest	326	5,464	5.9	(1.9)	83	1,362
Deciduous Forest	91,902	128,660	71.4	(3.7)	381	146,846
Mixed Forest	153	1,204	12.7	(8.7)	23	2,635
Coniferous Woodland	0	43	0.0	-	1	0
Deciduous Woodland	0	32,890	0.0	0.0	57	0
Mixed Woodland	0	3,376	0.0	0.0	11	0
Shrubland	0	13,610	0.0	0.0	8	5,202
Grass	7,795	13,659	57.1	(7.4)	55	112,282
Sparsely Vegetated/Barren	0	1,381	0.0	0.0	13	1,723
Artificial (roads, urban)	3,456	32,432	10.7	(3.5)	136	3,678
Cropland	402,789	499,237	80.6	(2.1)	536	451,658
Open Water	9,700	10,700	90.7	(4.6)	73	17,270
Total	516,121	742,656				742,656

[a] Land cover categories were defined by combining Iowa vegetation classes as follows: *coniferous forest* = pine forest, eastern red cedar forest, evergreen forest; *deciduous forest* = upland deciduous forest, temporarily flooded forested wetland, seasonally flooded forested wetland; *mixed forest* = mixed evergreen and deciduous forest; *coniferous woodland* = eastern red cedar woodland; *deciduous woodland* = upland deciduous woodland, temporarily flooded deciduous woodland, seasonally flooded deciduous woodland; *mixed woodland* = mixed evergreen and deciduous woodland; *shrubland* = upland shrub, temporarily flooded shrub, seasonally flooded shrub, semi-permanently flooded shrub, saturated shrub; *grass* = warm season grass/perennial forbs, temporarily flooded wetland, seasonally flooded wetland, semi-permanently flooded wetland, saturated wetland, permanently flooded wetland; grassland with sparse shrubs and trees; *sparsely vegetated/barren* = a single vegetation class that includes open bluff/cliff, talus slopes, mud, sand, soil; *artificial* = artificial with high vegetation, artificial with low vegetation; *agriculture* = cool season grass, cropland; *open water* = a single vegetation class. The woodland land cover categories were not present on the land cover map, but were observed in the field during the study.

[b] Producer's Accuracy is the probability that a pixel observed in the field is correctly depicted on the map.

[c] User's Accuracy is the probability that a pixel on the map correctly identifies the land cover category as it exists in the field.

To determine the allocation of sample pixels across land cover categories, we used a square root rule that balanced the need for estimates corresponding to the entire study area with the desire to obtain estimates for the defined land cover categories. We incorporated an adjustment factor for increased sample size in challenging land covers, and reduced sample size for land covers that were easier to classify. We then applied minimum (n=16) and maximum (n=44) sample sizes per stratum. The full list of pixels for a given land cover category was sorted by PSU, latitude, and longitude (to encourage geographic spread of the sample pixels), and a systematic sample was selected (Figure 2).

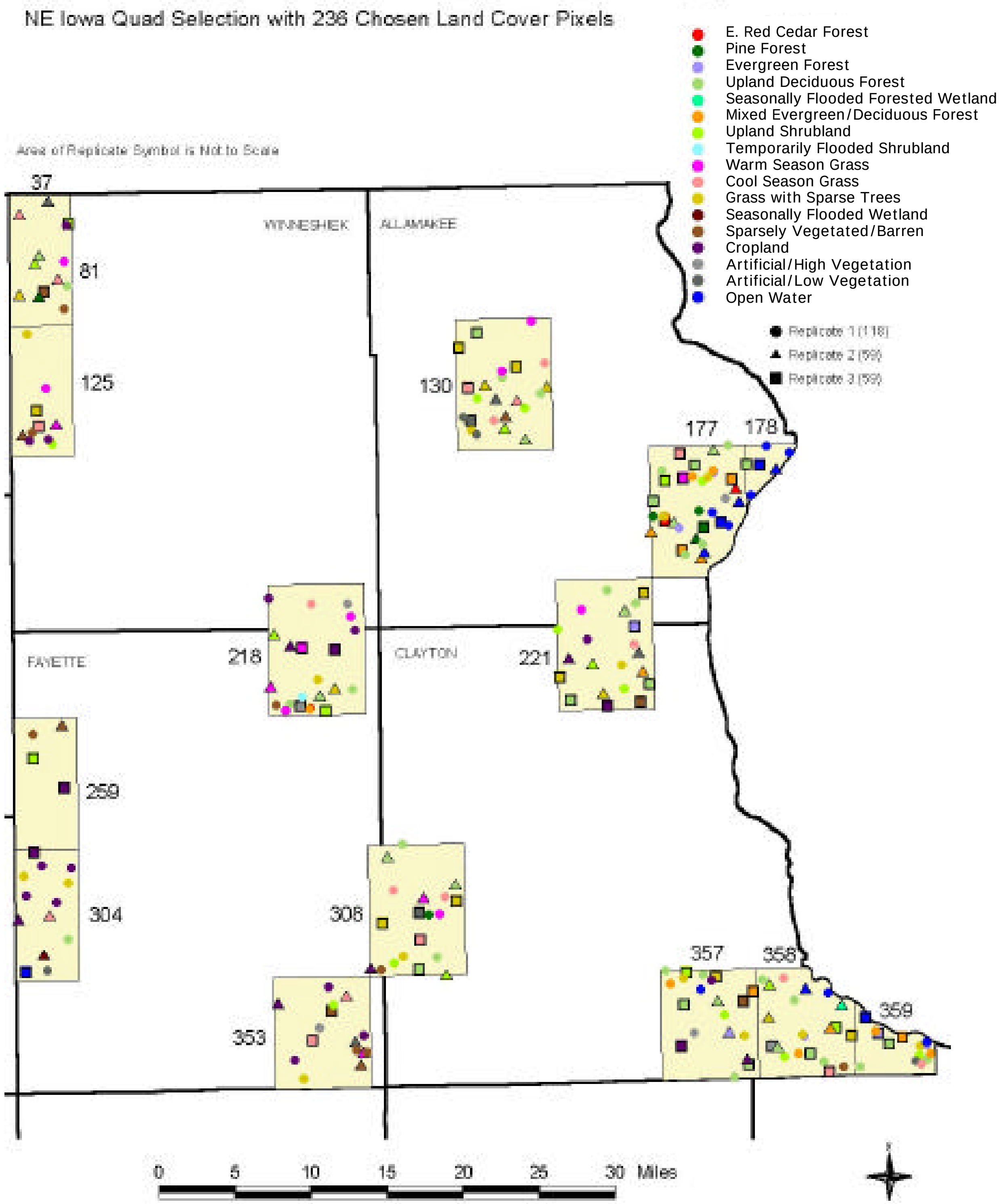

Figure 2. Sampled primary sampling units and sampled pixels by land cover. Numeric labels denote quad identification. Subsamples are denoted by symbols, as shown in the legend.

Because the time required to collect field data was not well known, the sample was divided into three balanced subsamples, corresponding to 50%, 25%, and 25% of the full sample, so that each balanced fraction of the sample could be completed and a decision made about resources availability for completing the next subsample. Field observers were instructed to complete samples from subsample 1 (50% sample) prior to collecting data on subsample 2, and were given similar instructions for subsample 3. In practice, these guidelines were implemented within county boundaries.

Obtaining Permission to Access Land

Owner information and the Public Land Survey (PLS) location for each sample pixel were obtained from offices of the County Auditor or Assessor. These offices are responsible for assessing property taxes and thus have the most recent information on land ownership. Plat directories and local phone directories were used to determine addresses and phone numbers for each landowner. Less than 10 of 236 addresses and ownerships were incorrect or had changed between the time of determination and the start of field work.

Of the 236 sample pixels, 198 were located on private property and 38 were on state or federal lands or were within city limits of towns. Letters requesting access to land were prepared using Iowa State University letterhead and mailed to each of the 198 private landowners along with a color land cover map of their county as a gift. Landowners returned 90 letters (45.4%) and 87 of these granted permission to enter their property. The day prior to visiting a site, a follow-up phone call was made to the landowner, regardless of whether a letter had been received or not, resulting in an additional 58 landowners who granted access and 8 who denied access. Due to insufficient time and resources, no follow-up calls or visits were made to 42 landowners in subsamples 2 and 3 in Fayette County and subsample 3 in Clayton County.

Field Assessment

Selected target pixels were located in the field by orienteering to the general vicinity of a point using the prepared topographic maps and then navigating to the exact coordinates of a point using a geographical positioning system (GPS) receiver with automatic differential correction capabilities. The GPS displayed a confidence interval from the desired coordinates that was usually less than five meters. Land cover was assessed for the target pixel (30 x 30 m) and the eight adjoining pixels using a list of codes for the 29 mapped vegetation classes in Iowa. A total of 18 points located on the floodplain of the Mississippi River were accessed with an air boat provided by the U.S. Fish and Wildlife Service.

Analysis

Field and map land cover data were used to estimate standard accuracy assessment rates (Congalton 1991), including the overall accuracy rate and the producer's and user's rates for each of 12 land cover categories. These corresponded to the nine preselected strata plus three additional woodland categories identified in the field but not present on the map. Two sets of analyses were performed to consider trade-offs in data collection effort and precision, one using all nine pixels from each of the 153 clusters (nine-pixel data) and a second based only on center pixels (center-pixel data).

Because an unequal probability sample design was used, and nonresponse occurred for some sample pixels, two sets of sample weights were calculated for use with center-pixel data and nine-pixel cluster data, respectively. A ratio adjustment was used to create weights that generate the map area for each land cover category when weights for points in the map land cover category are summed (Nusser and Klaas 2002).

To compare field-observed and map-determined land cover categories, weighted estimates of standard accuracy measures were calculated using estimators that were modified to incorporate sampling weights (Nusser and Klaas 2002). Variance estimates were obtained using PROC SURVEYMEANS in SAS (http://www.sas.com/rnd/app/da/new/802ce/stat/chap14/sect3.htm), accounting for pixel clusters and map land cover category strata. Domain estimation was used for estimating user's and producer's accuracy rates.

Results

Overall accuracy was estimated to be 69.5% (s.e. = 2.0) using the nine-pixel cluster data. The estimated accuracy rates for nine-pixel data varied greatly across land cover categories (Table 1). For example, the producer's accuracy is quite high for artificial and cropland categories but is poor for coniferous forest and especially for shrubland and sparse vegetation, all of which have relatively small map surface areas. A similar level of variation was observed in estimates of user's accuracy; water had a high accuracy rate, and smaller land cover classes had relatively poor accuracy. Three woodland land cover categories (coniferous, deciduous, mixed) were found in the field but were not present on the map. Mismatches between the field and map land cover categories were often associated with related land cover categories (Table 2). For example, pixels classified as woodland in the field were usually classified as forest on the land cover map. Pixels classified in the field as shrubland and sparse vegetation were often classified as herbaceous on the map.

Analyses using data from center pixels reflected similar estimates relative to the nine-pixel data but typically generated larger standard errors. The estimated overall accuracy of 64.0% (s.e. = 6.3) is not statistically different from the nine-pixel estimate but has an estimated standard error three times that of the nine-pixel estimate. Most single-pixel accuracy rate estimates (Table 3) were within ten percentage points of the nine-pixel estimates. The largest differences were found with smaller land cover categories, where a reduction in sample size had a relatively large effect. The center-pixel producer's accuracy estimate for mixed forest was 0%, because map and field-determined mixed forest pixels were never in agreement at a center pixel, whereas field and map matches for mixed forest were observed with nine-pixel data.

Nine-pixel cluster data clearly provides additional information for rare cover classes, as shown by the greater number of nonzero cells in the nine-pixel map by field matrix relative to the center-pixel matrix (Table 4). Standard errors for center-pixel estimates generally ranged from 1.5 to 4.5 times higher than the nine-pixel standard errors, with most being about triple the size of the nine-pixel estimates. For producer's accuracy estimates, one standard error (coniferous forest) was over ten times higher than the corresponding nine-pixel estimate, while one other (grass, water) was half of the nine-pixel standard error. This may be due in part to the dependence of the variance estimate on the estimated percentage. These results indicate that substantial gains in precision were generally obtained by observing additional data.

Table 2. Observed number of pixels in nine-pixel data, by field and map land cover category. [a]

Field Land Cover Category	Map Land Cover Category: Conif. Forest	Decid. Forest	Mixed Forest	Conif. Wdlnd	Decid. Wdlnd	Mixed Wdlnd	Shrub -land	Grass	Sparse Veg.	Artifi- cial	Crop- land	Open Water	Total
Coniferous Forest	39	29	15	0	0	0	0	0	0	0	0	0	83
Deciduous Forest	17	235	44	0	0	0	2	36	0	0	19	28	381
Mixed Forest	6	6	4	0	0	0	0	5	1	0	0	1	23
Coniferous Woodland	0	0	1	0	0	0	0	0	0	0	0	0	1
Deciduous Woodland	4	36	1	0	0	0	0	11	1	0	3	1	57
Mixed Woodland	2	8	0	0	0	0	0	0	0	0	1	0	11
Shrubland	0	1	0	0	0	0	0	3	0	0	4	0	8
Grass	1	10	0	0	0	0	0	23	0	0	3	18	55
Sparsely Vegetated/ Barren	0	0	0	0	0	0	0	8	0	4	0	1	13
Artificial (roads, urban)	0	4	2	0	0	0	1	40	3	41	44	1	136
Cropland	3	38	2	0	0	0	72	118	28	0	273	2	536
Open Water	0	4	0	0	0	0	0	3	3	0	0	63	73
Total	72	371	69	0	0	0	75	247	36	45	347	115	1,377

[a] Examining the table across rows shows how a land cover category observed in the field is categorized on the map (related to Producer's Accuracy). Examining the table by columns shows how map land cover categories are categorized in the field (related to User's Accuracy).

Table 3. Estimated accuracy rates by land cover category using center-pixel data.

Land Cover Category[a] (s)	Total Area with Consistent Field and Map Classifications (ha)	Estimated Field Area (ha)	Producer's Accuracy (%)			Map Area (ha)	User's Accuracy (%)		
			$P\hat{A}(s)$ [b]	(se)	n		$U\hat{A}(s)$ [c]	(se)	n
Coniferous Forest	599	5,957	10.1	(9.2)	9	1,362	43.9	(13.5)	14
Deciduous Forest	86,268	137,375	62.8	(12.3)	43	146,846	58.7	(9.1)	30
Mixed Forest	0	310	0.0	(0.0)	2	2,635	(0.0)	(0.0)	14
Coniferous Woodland	0	187	0.0	-	1	0	-		0
Deciduous Woodland	0	42,397	0.0	(0.0)	6	0	-		0
Mixed Woodland	0	5,081	0.0	(0.0)	2	0	-		0
Shrubland	0	21,827	0.0	-	1	5,202	0.0	(0.0)	17
Grass	13,111	19,986	65.6	(19.9)	6	112,282	11.7	(6.4)	26
Sparsely Vegetated/Barren	0	365	0.0	-	1	1,723	0.0	(0.0)	9
Artificial (roads, urban)	3,313	37,267	8.8	(6.1)	15	3,678	90.1	(9.5)	10
Cropland	364,349	463,759	78.6	(5.6)	60	451,658	80.7	(8.5)	20
Open Water	7,971	8,145	97.8	(2.2)	7	17,270	46.1	(13.9)	13
Total	516,121	742,656				742,656			153

[a] Land cover categories were defined by combining Iowa vegetation classes as follows: *coniferous forest* = pine forest, eastern red cedar forest, evergreen forest; *deciduous forest* = upland deciduous forest, temporarily flooded forested wetland, seasonally flooded forested wetland; *mixed forest* = mixed evergreen and deciduous forest; *coniferous woodland* = eastern red cedar woodland; *deciduous woodland* = upland deciduous woodland, temporarily flooded deciduous woodland, seasonally flooded deciduous woodland; *mixed woodland* = mixed evergreen and deciduous woodland; *shrubland* = upland shrub, temporarily flooded shrub, seasonally flooded shrub, semi-permanently flooded shrub, saturated shrub; *grass* = warm season grass/perennial forbs, temporarily flooded wetland, seasonally flooded wetland, semi-permanently flooded wetland, saturated wetland, permanently flooded wetland; grassland with sparse shrubs and trees; *sparsely vegetated/barren* = a single vegetation class that includes open bluff/cliff, talus slopes, mud, sand, soil; *artificial* = artificial with high vegetation, artificial with low vegetation; *agriculture* = cool season grass, cropland; *open water* = a single vegetation class. The woodland land cover categories were not present on the land cover map, but were observed in the field during the study.

[b] Producer's Accuracy is the probability that a pixel observed in the field is correctly depicted on the map.

[c] User's Accuracy is the probability that a pixel on the map correctly identifies the land cover category as it exists in the field.

Table 4. Observed number of pixels in center-pixel data, by field and map land cover category.[a]

Field Land Cover Category	Map Land Cover Category												Total
	Conif. Forest	Decid. Forest	Mixed Forest	Conif. Wdlnd	Decid. Wdlnd	Mixed Wdlnd	Shrub-land	Grass	Sparse Veg.	Artifi-cial	Crop-land	Open Water	
Coniferous Forest	6	1	2	0	0	0	0	0	0	0	0	0	9
Deciduous Forest	5	18	9	0	0	0	0	5	0	0	1	5	43
Mixed Forest	1	0	0	0	0	0	0	0	0	1	0	0	2
Coniferous Woodland	0	0	1	0	0	0	0	0	0	0	0	0	1
Deciduous Woodland	1	3	0	0	0	0	0	1	0	0	1	0	6
Mixed Woodland	1	1	0	0	0	0	0	0	0	0	0	0	2
Shrubland	0	0	0	0	0	0	0	0	0	0	1	0	1
Grass	0	1	0	0	0	0	0	3	0	0	0	2	6
Sparsely Vegetated / Barren	0	0	0	0	0	0	0	0	0	1	0	0	1
Artificial (roads, urban)	0	0	1	0	0	0	0	3	9	1	1	0	15
Cropland	0	6	1	0	0	0	17	14	0	6	16	0	60
Open Water	0	0	0	0	0	0	0	0	0	1	0	6	7
Total	14	30	14	0	0	0	17	26	9	10	20	13	153

[a] Examining the table across rows shows how a land cover category observed in the field is categorized on the map (related to Producer's Accuracy). Examining the table by columns shows how map land cover categories are categorized in the field (related to User's Accuracy).

Discussion

A primary goal of this pilot study was to explore the use of the sample survey approach in accuracy assessment, including sample design, owner contact, field data collection, and analysis. A sample design was developed to balance operational and statistical considerations and to cover the entire study area, regardless of accessibility. The stratified two-stage cluster sample design worked well to control sample sizes for map land cover categories and to encourage geographic spread across and within PSUs. The design proved sufficiently flexible that it was easily adapted for two neighboring states (Nusser and Klaas 2002).

Early in the project design phase, we discussed alternative definitions for the first-stage sampling unit, or PSU. A quad sheet (or quarter quad) has been used in the past as a sampling unit at this stage for other GAP accuracy assessment studies. Quad sheets provide an operational advantage in reducing travel time and workload relative to a systematic or simple random sample, but are sufficiently large to avoid overly clustered second-stage samples that reduce the statistical efficiency of the design. A second alternative is to define the PSU as a county or a portion of a county, which has similar properties but would provide significant operational efficiencies when identifying landowners.

The choice of a pixel as the second-stage sampling unit was simple to work with in the sampling process. The stratum identification provided the control needed to address sample size requirements for strata, and the allocation strategy allowed us to balance estimation goals for land cover classes. The gain in precision of accuracy estimates obtained from the nine-pixel design and the increased ability to gather data for rare land covers were deemed well worth the extra effort required to observe land cover for each of the pixels in the 3 x 3 pixel clusters.

The pilot study demonstrated the need to accurately locate the pixel. Without precise positioning, field staff may visit a pixel with a map land cover category different from the category associated with the true location of the selected pixel and destroy the control provided by stratification for land cover categories.

Protocols for contacting landowners had a large effect on the response rates in the study. Several attempts were made to contact landowners and different contact modes (e.g., telephone, mail) were used to improve response rates. Key strategies included using Iowa State University letterhead (rather than federal agency letterhead), explaining the study and its significance to Iowa and the landowner, offering a printed map of the area as a gift, and calling the landowner before the visit to remind him/her of the project to seek permission if needed. These protocols are derived from proven sample survey methodologies that are known to maximize response rates (Salant and Dillman 1994).

One of the advantages of the design used is that all land was eligible to be assessed for accuracy, and thus the results apply to the entire target area. Although few areas are physically inaccessible in the Midwest, there is still a need to develop ground-truthing methods for inaccessible or otherwise unobservable sample units. For example, aerial photography may provide a surrogate material for unobservable units.

A major concern with the current pilot study was the use of 1999 field data to assess the accuracy of a land cover map derived from 1992 imagery. Large changes in land cover can occur in this time span that confound assessments of the digital map.

Literature Cited

Cochran, W.G. 1977. Sampling techniques. Wiley, New York. 428 pp.

Congalton, R. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37:35-46.

Lohr, S.L. 1999. Sampling: Design and analysis. Brooks/Cole Publishing Company, Pacific Grove, California. 494 pp.

Nusser, S.M., and E.E. Klaas. 2002. Final performance report to EPA Region 7, Part II: GAP accuracy assessment pilot study. Environmental Protection Agency Contract X997387-01 Final Report. Iowa Cooperative Fish and Wildlife Research Unit, Iowa State University, Ames, Iowa. 77 pp.

Salant, P., and D.A. Dillman. 1994. How to conduct your own survey. Wiley, New York. 232 pp.

Stehman, S.V. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* 62:77-89.

Thompson, S.K. 1992. Sampling. Wiley, New York. 343 pp.

An Evaluation of Helicopter Use for Collecting Land Cover Data for Southwest ReGAP in Colorado

DONALD L. SCHRUPP[1], DIANNE D. OSBORNE[2], AND LEE E. O'BRIEN[3]

[1]Colorado Division of Wildlife, Denver

[2]Bureau of Land Management, National Science & Technology Center, Denver

[3]Natural Resource Ecology Laboratory, Colorado State University, Fort Collins

As a part of the Southwest Regional Gap Analysis Project, the Colorado Division of Wildlife (CDOW) and U.S. Bureau of Land Management (BLM) conducted an evaluation of helicopter-based methods for collecting ground-truth reference information and compared this methodology to collecting data via automobile and on foot. These data are used for classifying Landsat-7 Enhanced Thematic Mapper satellite imagery in developing a land cover map of a five-state region in the Southwest. It was found that although more expensive than traditional ground-based collection of field data, the helicopter method had some advantages.