# SURVEY METHODS FOR ASSESSING LAND COVER MAP ACCURACY

**S. M. Nusser and E. E. Klaas**

**September 2002**

## 1. Introduction

The increasing availability of digital photographic materials has spurred efforts by agencies and organizations to generate digital land cover maps for states, regions, and the US as a whole. For example, the national Gap Analysis Program (GAP) was developed by the U.S. Geological Survey (USGS) Biological Resources Division to address the need for information on wildlife population distributions in relation to habitat availability and current management practices. GAP is a cooperative program under which partial funding is provided to states for identifying the current distribution and management status of land cover types and wildlife habitats (Crist and Deitner, 1998). As part of this effort, states are creating digital maps of land cover and species distributions for identifying mismatches between species habitat requirements and the land management status of the habitat (Jennings, 2000). Most of the land cover maps are being created from a combination of satellite imagery, classification tools to process the imagery, and/or auxiliary information to augment the information generated by the imagery.

Regardless of the method used by a state, these maps are subject to numerous sources of error that arise from source materials and methods used to generate thematic information from these materials (Congalton and Green, 1993). In order to understand the information content of these maps, it is desirable to generate precise estimates of accuracy rates describing misclassification errors for the entire map area. This usually involves comparing an alternative measure of "truth" or reference land cover, with the map land cover at a subset of points or areas on the map. A variety of methods have been used to implement such

assessments, but not all of these approaches will generate estimates of accuracy rates that have good statistical properties.

For example, quantifying the accuracy of a GAP land cover map involves comparing the thematic content of the digital map with corresponding thematic reference data obtained from the state (or other target area). Typically, assessment locations are selected from the target area, and reference data are gathered, for example, via field visits or photo-interpretation (Congalton and Green, 1993). Methods of selecting assessment locations range 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). A variety of sampling units may be used in selecting a sample, including land areas or points on the land.

Ideally, accuracy estimates are based on probability 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 natural design and estimation framework that balances statistical and operational considerations in relation to study objectives (Cochran, 1977; Salant and Dillman, 1994; Lohr, 1999). Probability sampling designs can be created to target areas requiring more intensive study, reduce the effort in areas that are difficult to access, and/or rely on clusters of observation units in order to reduce study costs. Data collection methods used in survey sampling provide effective approaches for contacting land owners and gaining their cooperation to access private land, thereby minimizing bias from nonresponse. In addition, just as scripted interviews with well-defined and simple questions provide 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. Finally, estimation approaches are readily available from this framework that take into account the survey design used in the study, nonresponse due to lack of access to a sample location, and known information about the target area, such as land cover surface areas from the final map. The sample survey approach focuses on minimizing total survey error (Groves, 1989), which includes sampling error arising from the sampling process and controlled via the sample design and estimation strategies, as well as nonsampling errors (Lessler and Kalsbeek, 1992) such as sampling frame errors and selection bias (e.g., omitting part of the target population) and measurement errors.

As part of an effort to explore ideas for designing an integrated accuracy assessment plan in EPA Region 7, we worked with GAP representatives from Iowa, Kansas, Missouri and Nebraska to design and conduct pilot studies using a sample survey approach to assess the accuracy of GAP land cover maps in the region. 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 appropriate sample design and estimation strategies as well as effective protocols for gaining permission to access private land and observation units that would minimize measurement error and time required to collect data relative to resource constraints. The Iowa pilot study was the most comprehensive of three small studies designed to explore the use of a survey sampling framework for GAP accuracy assessment (Nusser and Klaas, 2002), and is the focus of this paper.

Based on study objectives, we selected a stratified two-stage sample of pixels that relied on clustering to reduce travel costs and stratification to encourage geographic spread and to control sample sizes across land cover classes. We developed rigorous but practical contact methods to obtain a high response rate for field reference data with relatively low

effort. Part of the study was devoted to exploring the trade-off between data collection effort and statistical gains by collecting data for a single sample pixel and for a cluster of pixels centered at the sample pixel. We used the data to calculate weighted estimates of accuracy parameters as well as standard errors of the estimates that accounted for the survey design to explore the statistical aspects of the choice for observational unit.

In this paper, we describe the design, implementation and results of the Iowa pilot study. We begin by focusing on methodological considerations in selecting observation units, the sample design, and field assessment methods. The study design and results are summarized, and we discuss the benefits and challenges of the methods implemented.

## 2. Methodological Considerations

### 2.1 Observation Units

In designing the study, we began by reviewing various options for units that could be used in field measurement and sampling. Two types of units considered for assessment of land cover maps were polygons and pixels (Stehman and Czaplewski, 1998). Polygons may take the form of regularly shaped observation units defined without regard to land cover category (e.g., a rectangle defined by the cell of a grid covering the target area). More commonly in accuracy assessment studies, a polygon corresponds to an area of uniform thematic composition (i.e., one land cover category) on the map being assessed. Although defining the observation unit as a homogeneous land-cover polygon on a map may be intuitively appealing, working with the land cover polygon in the reference domain (e.g., the field, a high quality photograph) may be quite difficult in practice. For example, the ground assessor or photo-interpreter may find it difficult to identify the border of a sample polygon due to its irregular shape and size. In addition, accurately recording the composition of the polygon

can be challenging when the reference material reveals that the sample polygon is, in fact, not homogenous at all.

An alternative observation unit is a pixel on the land cover map, a square region representing the smallest identifiable unit on the map. By definition, a pixel corresponds to the smallest unit of spectral data on the satellite image, and thus a pixel corresponds to exactly one land cover category on the map. A pixel from the images used in this study represents an area on the ground of 30 m x 30 m. If multiple land cover categories exist within the pixel boundaries for the reference source, difficulties may still be encountered (Crist and Deitner, 1998; Congalton and Green, 1993), and pre-defined rules are needed to determine the land cover category for the pixel.

One idea that was discussed at this stage was obtaining data on more than just a single pixel. Considerable effort is expended to field-visit a pixel on the land, and we thought that relatively little effort may be needed to gather additional data to improve the precision of the estimates. Thus, we considered collecting field data for the sample pixel plus the eight adjacent pixels, forming a 3 x 3 pixel grid centered on the sample pixel. A 5 x 5 pixel grid was also explored since some states in the region were working with minimum mapping units of 2 ha; the area of a 5 x 5 pixel grid is 2.25 ha.

### 2.3 Sample Designs

Our goal was to design and implement a probability sampling design that applied to the full target population as a key element in developing a foundation for precise and approximately unbiased assessments of map accuracy for the entire area of interest. Many designs are possible using a sample survey framework (Cochran, 1977; Lohr 1999), and here we focus on strategies that apply to obtaining field measurements for accuracy assessment. For the pilot study, we were interested in using cluster sampling so that travel resources would be used as

efficiently as possible. For this purpose, we considered 7.5' quadrangle sheets (quads) as the basis for forming primary sampling units since this had worked well for many states. Stratification can be used at this stage to ensure geographic spread of the first-stage sample. We proposed using pixels as the second-stage sample unit, to be selected from the list of pixels that fell within the first-stage sample units. In this second stage of sampling, stratification can be used to ensure that the sample is spread across analysis domains (map land cover categories) and to allocate sample sizes to strata in relation to the importance ascribed to land cover categories. This strategy results in an unequal probability design that balances statistical and operational considerations.

### 2.4 Assessment Methods

The response methodology describes how the reference data are collected and recorded. In accuracy assessment studies, this typically involves applying a classification scheme to the reference source material. The reference classification scheme should be mutually exclusive and exhaustive, and include a direct correspondence with the map land cover classification scheme. It has been recommended that the reference data be classified on a hierarchical scheme that provides more detail than is discernable from the map land cover data (Congalton, 1991; Crist and Deitner, 1998). When a hierarchical classification is used, reference land cover classes can be collapsed into broader land cover categories which correspond to categories on the map being assessed.

Another consideration in the response methodology is the source of the reference data. Reference data are often collected using aerial photography. The use of such photography may lead to questionable results, however, since the interpretation and accuracy of these photos vary. As a result, Congalton (1991) notes that ground visits are thus preferred to aerial photography when they are financially and practically feasible. There are problems

with ground visits as well, however. For example, to obtain target sample sizes, the initial sample size needs to be inflated to account for nonresponding units (e.g., inaccessible or permission denied). This is a strategy used in sampling human populations where the goal of obtaining responses from all sampling units selected is unrealistic. The field assessor must also correctly locate the selected points in order for the design to have the desired result. The availability of precise positioning for GPS receivers has reduced this concern. However, if observation units cannot be precisely located, then strategies used to obtain adequate sample sizes for subpopulations are thwarted. Such issues can drastically influence both the initial sample size and the sampling scheme used in the study (Crist and Deitner, 1998). Finally, the choice of response methodology can be affected by the terrain of the land and the map being assessed. In areas that are inaccessible, high resolution photography may be explored as an alternative method of gauging the accuracy of the land cover map. If available for the year associated with the land cover map, such photography may also be better than pursuing a field assessment years after the map's nominal reference year.

In Iowa, high resolution photography was not available, and a field visit strategy was selected to obtain reference data. Even though the study was conducted prior to the removal of selective availability of GPS signals, we had access to a GPS receiver that would receive precise positioning signals, making it possible to accurately find sample pixels in combination with maps and photographic materials.

To summarize accuracy assessment data, we followed the standard contingency table method, which relies on an error matrix (Congalton, 1991; Stehman, 1997; Crist and Deitner, 1998). An error matrix is a square array of numbers that presents summary information on units classified as map land cover category *s* and reference land cover category *t*. Cell values may be the number (or percentage) of sample units or the estimated land area (or percentage) corresponding to the map and reference land cover. One summary statistic calculated from

the error matrix is the overall accuracy of the map, which estimates the proportion of area within the target region for which the map and the reference data are in agreement. Two other measures of accuracy include the producer's accuracy and the user's accuracy. The producer's accuracy is an estimate of the percentage of field area associated with a land cover category for which the map and the reference data are in agreement. The user's accuracy estimates the percentage of map area associated with a land cover category for which the map and reference data are in agreement. It should be noted that the interpretation of error matrix summaries is a function of the materials and processes used to perform the accuracy assessment (Congalton and Green, 1993). In our study, an unavoidable nuisance factor was the difference between the date of the reference data and the date of the satellite image and auxiliary information used to construct the map.

If unequal probability designs are used and/or differential nonresponse occurs across strata, weighted estimates of these accuracy measures should be calculated to account for the sampling design and nonresponse (Stehman and Czaplewski, 1998). When unequal probability designs are used, sampling weights must be calculated and included in the analyses to account for varying selection probabilities for sampling units (Congalton, 1988; Stehman, 1999). The sampling weight of a unit reflects how many elements in the population are represented by that single unit. The simplest form of a sampling weight is an inverse selection probability. The higher the weight (e.g., in hectares) assigned to a sampling unit, the more observation units (e.g., surface area) within the population it represents. Weights can also be used to account for nonresponse that occurs when access to land is denied (Lohr, 1999). In calculating weights, ratio adjustments can be implemented so that weights accurately reflect the surface area of the state, individual map land cover categories, and/or other geographic subdivisions.

## 3. Iowa GAP Accuracy Assessment Pilot Study Methods

### 3.1 Study area

A pilot study was initiated in 1999 to explore appropriate sample design, field data collection, and analysis methodologies for accuracy assessment of GAP land cover maps for Iowa. The land cover maps were developed by integrating a computer-assisted analysis of 1992-94 Landsat Thematic Mapper (TM) data with field observations and National Wetlands Inventory data. The target population was defined by four counties in northeastern Iowa: Allamakee, Clayton, Fayette and Winneshiek counties. This region was selected because the land cover mapping process was nearly complete in this area.

### 3.2 Sample Design

A stratified two-stage cluster sample design (Lohr, 1999) was used to select sample pixels for field visits from the four-county study area. The first stage involved selecting area segments roughly the size of a 7.5' quadrangle. In the second stage, individual pixels were selected.

The study area included 70 whole or partial USGS 7.5' quadrangles (quads). For the most part, the primary sampling unit (PSU) was defined to be a single quad. However, to ensure that all of the PSUs covered roughly the same amount of land area, some PSUs were defined to be a combination of partial quads or a partial with a whole quad (Figure 1). First stage strata were created to ensure geographic spread of the PSUs and to ensure coverage of all land cover categories. The study area was divided into five strata, each consisting of eight to 12 PSUs. The borders of the strata were defined so that within-stratum variation in land cover was relatively low and among-strata variability in land cover was relatively high. Two PSUs were randomly selected from each stratum, for a total of ten PSUs, using systematic sampling from a list that reflected serpentine geographic ordering of the PSUs.

Individual pixels were selected from PSUs in a second stage of sampling. Resource constraints dictated sample size. Iowa staff had a goal of field visiting 200 points within the study area. We expected that access would be denied for approximately 15% of the sample points, indicating 236 sample points would be needed to achieve 200 responses. The pixel sample was selected from the ten PSUs using a stratified design. The strata were defined to be nine relatively homogeneous land cover categories, collapsed from the original 29 vegetation classes in Iowa (Table 1). Land cover categories were defined 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* is 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* is a single vegetation class. The three woodland land cover categories were not present on the land cover map, but were observed in the field during the study. Thus, nine land cover strata were used to stratify the pixel frame.

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 (which usually calls for stratum sample sizes proportional to stratum area) with the desire to

obtain estimates for the land cover categories defined as strata (which usually involves equal allocation across strata). In addition, we incorporated an adjustment factor to increase the sample size for land covers that were difficult to classify and reduce the sample size for land covers that were easier to classify, and then applied minimum and maximum sample sizes per stratum.

More specifically, the initial allocation of sample pixels of land category $s$ was proportional to the square root of the total area of land cover category $s$ in the study region, $A_s$, multiplied by an adjustment coefficient for the land cover category $s$, $K_s$. Thus, the allocation rule used was

$$n_s^0 \propto K_s \sqrt{A_s} \ ,$$

where $n_s^0$ is the number of pixels in land cover category $s$ in the sample. The adjustment coefficient, $K_s$, reflects the priority of land cover category $s$ relative to study objectives. For a land cover category that was thought likely to be less accurately classified or that had a small land area, $K_s = 2$; for a land cover category that is relatively easy to classify or had a large area, $K_s = 0.5$; and for all other land cover categories, $K_s = 1$. To create the final allocation across strata, a minimum and maximum sample size was determined (16 and 44 pixels, respectively). Thus, the initial sample allocation for each land cover category, $n_s^0$, was further adjusted to obtain the final allocation $\{n_s: s = 1, 2, ..., 9\}$ such that

$$16 \le n_s \le 44$$

and

$$\sum_{s=1}^{9} n_s \le 236 .$$

The adjustment factors, initial sample size allocation, and final sample size allocation, $n_s$, are presented in Table 1.

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. A random starting point was selected and the pixels were sampled systematically from the list within each land cover category stratum. Figure 2 presents the distribution of the sample pixels in relation to the PSUs. Each selected pixel identifies a point on the land that was to be field-visited if possible.

Because the time required to collect field data was not well known, the sample was divided into three subsamples, corresponding to 50%, 25%, and 25% of the full sample, respectively, so that a balanced fraction of the sample could be completed and a decision made about resources availability for completing the next subsample. A systematic procedure was used to divide the sample so that the subsample were balanced across land cover categories and dispersed geographically. 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.

### 3.3 Field Assessment

#### 3.3.1 Determination of Land Ownership and Obtaining Permission to Access Land

The 236 sample pixels from Allamakee, Clayton, Fayette, and Winneshiek counties were plotted as points on a topographic map using ArcView, and printed on a color printer. A spreadsheet was prepared with the following data columns with the pixel ID, pixel coordinate, and public land survey information (township, range, and section). Maps and spreadsheet information were taken to offices of the County Auditor or Assessor in each county and used to look up property owners on large scale plat map in the county office. County offices that assess property taxes are known to have the most recent information on

land ownership because land sales must be recorded with these offices soon after the sale is final. Plat directories (Farm and Home Plat and Directory. 1999. Farm and Home Publishers, Ltd, Box 305, Belmond, Iowa 50421) 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 were prepared using Iowa State University letterhead and mailed to each of the 198 private land owners along with a color land cover map (8.5 x 11 in) of their county as a gift. Two copies of the letter were enclosed in the mailing. Landowners were requested to sign and return one copy in a postage paid envelope; the other copy was to be kept for their files.

A total of 90 letters (45.4%) were returned, and 87 of these granted permission to enter their property. Most of the responses were received in the first two weeks after mailing. Field assessments began about two months after the letters were mailed. The day or evening 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. Phone calls resulted in an additional 58 landowners who granted permission to visit their land and 8 who denied access. Due to insufficient time and resources, no follow-up calls were made to 42 landowners, and these sites were not visited or were assessed from nearby roads. These sample sites corresponded to subsamples 2 and 3 in Fayette County, and subsample 3 in Clayton County.

### 3.3.2 Field Observations

Selected target pixels were located in the field using topographic maps and GPS receivers. Land cover was assessed for the target pixel (30 x 30 m) and the eight adjoining pixels.

Vegetation classes were recorded in the field using a list of codes for each of the 29 mapped vegetation classes, which were later collapsed to reflect the 12 classes defined for this study (Section 3.2). At first, an attempt was made to assess a 5 x 5 grid of 25 pixels, but this proved to be too time consuming. In forested areas it was usually necessary to navigate and walk to each pixel in order to make an accurate assessment. Nine pixels could be reached in a reasonable amount of time (less than 30 minutes), whereas 25 pixels required an average of one-hour or more in rough terrain.

## 3.4 Estimation

### 3.4.1 Overview

Two sets of analyses were performed to consider trade-offs in data collection effort and precision, one using all nine pixels from each cluster (nine-pixel data) and a second based only on center pixels (center-pixel data). In what follows, there are $n = 153$ pixel clusters (indexed by $j$) with $m = 9$ pixels each (indexed by $h$), which were selected from nine strata (indexed by $k$). There are 12 possible land cover categories for the map (indexed by $s$) and the field (indexed by $t$) data. Sample weights are in units of hectares.

### 3.4.2 Weighting

Sample weights were calculated to account for the unequal probability sample design and the presence of nonresponse in the study. Two sets of weights were calculated, one for use in estimating accuracy rates with center-pixel data and the other for use with the nine-pixel cluster data. A ratio adjustment was applied to incorporate known information on the surface areas for each land cover category on the map. This corresponds to using a ratio estimator for accuracy rates (see Section 3.4.3). Ratio estimators are biased, but are design-consistent and generally have a smaller mean square error than the corresponding Horvitz-Thompson

estimator (Särndal, Swensson, and Wretman, 1992). An additional benefit is that the sum of weights for sample points with a given map land cover category is equal to the known surface area for the land cover category on the map.

Weights for the center pixel of each cluster were calculated as follows. The initial weight for center pixel $j$, belonging to first-stage stratum $k$, and classified as having map land category $s$ was defined to be

$$w_{kj} = \frac{1}{\pi_{kj}} \frac{A_s}{\hat{A}_{c,s}},$$

where $\pi_{kj} = n_k m_s / N_k M_s$ is the inclusion probability for center pixel $j$ in the first stage stratum $k$ having map land cover category $s$, $N_k$ is the total number of PSUs in the first-stage stratum $k$, $n_k$ is the number of PSUs selected in the sample for stratum $k$, $M_s$ is the total number of pixels of map land category $s$ in the first-stage sample, $m_s$ is the number of pixels of map land category s in the second-stage sample, $A_s$ is the area in hectares of map land category $s$ for the entire study region, $\hat{A}_{c,s} = \sum_k \sum_j \pi_{kj}^{-1} G_{kj}(s)$ is the Horvitz-Thompson estimator of the surface area for map land cover $s$ using the center pixel data, and

$$G_{kj}(s) = \begin{cases} 1, & \text{if the center pixel in cluster } j \text{ in stratum } k \text{ has map land cover category } s \\ 0, & \text{otherwise} \end{cases}. \quad (1)$$

The ratio adjustment ensures that estimates are consistent with known information about land cover areas; i.e.,

$$\sum_k \sum_j w_{kj} G_{kj}(s) = A_s .$$

To calculate weights for each pixel within a nine-pixel cluster, a similar approach was used. For pixel $h$ with map land category $s$ associated with center pixel $j$ in first-stage stratum $k$,

$$w_{kjh} = \frac{1}{9\pi_{kj}} \frac{A_s}{\hat{A}_{9,s}},$$

where $\hat{A}_{9,s} = \sum_k \sum_j \sum_h (9\pi_{kj})^{-1} G_{kjh}(s)$ is an estimator of the surface area for map land cover $s$ using the nine-pixel cluster data, and

$$G_{kjh}(s) = \begin{cases} 1 & \text{if pixel } h \text{ in cluster } j \text{ in stratum } k \text{ has map land cover category } s \\ 0 & \text{otherwise} \end{cases} . \quad (2)$$

As with the center pixel weight, the nine-pixel weight includes a ratio adjustment so that the sum of the weights for pixels classified as having map land cover category $s$ equals the map area associated with land cover $s$, or

$$\sum_k \sum_j \sum_h w_{kjh} G_{kjh}(s) = A_s .$$

Determining the exact inclusion probability for pixel $h$ in cluster $j$ in stratum $k$ requires an intensive calculation process to identify the center-pixel inclusion probability for all possible nine-pixel clusters that contain pixel $h$. (We assume the chances of selecting two pixels that generate overlapping clusters is negligible.) Note that $9\pi_{kj}$ is an approximation to this inclusion probability, which is exact if the nine possible ways in which pixel $h$ can be included in the sample are all equally likely. The approximate inclusion probability for pixel $h$ is closer to the true inclusion probability if the area around the pixel is homogeneous, which is reasonably likely in a Midwestern landscape. The approximation could be improved by using known inclusion probabilities for the pixels adjacent to the pixel being considered. For example, the inclusion probability for the center pixel is equal to the sum of inclusion probabilities for all nine pixels; a possible estimator for a non-center pixel is nine times the average of the four or five available inclusion probabilities for the non-center pixel plus its neighbors.

### 3.4.3 Accuracy Rate Estimators

To compare field-observed and map-determined land cover categories, standard accuracy measures were considered, including the overall accuracy rate and producer's and user's accuracy rates for each of twelve land cover categories (Congalton, 1991). Estimators for accuracy rates are expressed below as weighted means, and represent design-consistent ratio estimators that account for the unequal probability sample design and an adjustment for known information about map land cover areas. Estimators are derived separately for the nine-pixel cluster data and the center pixel data.

The estimator for the overall accuracy rate, or the percentage of the study area for which the field and map land cover categories were consistent, using the data from all pixels in the cluster is defined by

$$\hat{OA} = 100 \frac{\sum_k \sum_j \sum_h w_{kjh} I_{kjh}(s = t)}{\sum_k \sum_j \sum_h w_{kjh}},$$

where for map land cover category $s$ and field land cover category $t$,

$$I_{kjh}(s = t) = \begin{cases} 1, & \text{if } s = t \text{ for pixel } h \text{ in cluster } j \text{ in stratum } k \\ 0, & \text{otherwise} \end{cases}.$$

Estimates for user's and producer's accuracies were generated for each of the 12 possible land cover categories. The ratio estimator for the producer's accuracy rate for field land cover category $t$, $PA(t)$, is

$$\hat{PA}(t) = 100 \frac{\sum_k \sum_j \sum_h w_{kjh} I_{kjh}(s = t)}{\sum_k \sum_j \sum_h w_{kjh} F_{kjh}(t)}$$

where

$$F_{kjh}(t) = \begin{cases} 1 & \text{if pixel } h \text{ in cluster } j \text{ in stratum } k \text{ has field land cover class } t \\ 0 & \text{otherwise} \end{cases}.$$

The ratio estimator for the user's accuracy rate for GAP map land cover category $s$, *UA(s)* is defined by

$$\hat{UA}(s) \;=\; 100 \;\frac{\sum_k \sum_j \sum_h w_{kjh} I_{kjh}(s = t)}{\sum_k \sum_j \sum_h w_{kjh} G_{kjh}(s)},$$

where $G_{kjh}(s)$ is defined in equation (2).

For center-pixel data, the estimator for the overall accuracy rate, *OA*, is

$$\hat{OA} \;=\; 100 \;\frac{\sum_k \sum_j w_{kj} I_{kj}(s = t)}{\sum_k \sum_j w_{kj}},$$

where for map land cover category s and field land cover category t,

$$I_{kj}(s = t) \;=\; \begin{cases} 1, & \text{if } s = t \text{ for the center pixel in cluster } j \text{ in stratum } k \\ 0, & \text{otherwise} \end{cases}.$$

The center-pixel estimator for the producer's accuracy rate for field land cover category $t$, *PA(t)*, is

$$\hat{PA}(t) \;=\; 100 \;\frac{\sum_k \sum_j w_{kj} I_{kj}(s = t)}{\sum_k \sum_j w_{kj} F_{kj}(t)}, \qquad (3)$$

where

$$F_{kj}(t) \;=\; \begin{cases} 1 & \text{if the center pixel in cluster } j \text{ in stratum } k \text{ has field land cover class } t \\ 0 & \text{otherwise} \end{cases}.$$

The center-pixel estimator for the user's accuracy rate for map land cover category s, *UA(s)* is defined by

$$\hat{UA}(s) \;=\; 100 \;\frac{\sum_k \sum_j w_{kj} I_{kj}(s = t)}{\sum_k \sum_j w_{kj} G_{kj}(s)},$$

where $G_{kj}(s)$ is defined in equation (1).

### 3.4.5 Variance Estimation

A Taylor series expansion of the ratio estimator can be used to derive an approximate expression for the variance of the ratio estimator (Cochran, 1977). For the pilot study design, the variance estimator is based on cluster-to-cluster variation within strata. For example, for the nine-pixel cluster estimator for producer's accuracy in (3), the estimator for the variance is

$$\hat{V}[\hat{PA}(t)/100] = \sum_{k} \frac{m}{m-1} \sum_{j} (g_{kj.} - \bar{g}_{k..})^2 ,$$

where $g_{kj.} = [\sum_{k} \sum_{j} \sum_{h} w_{kjh} F_{kjh}(t)]^{-1} \sum_{h} w_{kjh} [I_{kjh}(s = t) - F_{kjh}(t) \hat{PA}(t)/100]$ and $\bar{g}_{k..} = \frac{1}{m} \sum_{j} g_{kj.}$.

We used the SURVEYMEANS procedure in SAS to calculate accuracy parameter estimates and corresponding variance estimates (http://www.sas.com/rnd/app/da/new/802ce/stat/chap14/sect3.htm). For the full nine-pixel cluster data, the estimation procedures assume that the sample design was a stratified random sample of clusters, where strata were defined as the center-pixel map land cover category ($k$), clusters are defined to be nine-pixel units ($j$), and clusters assumed to be selected within strata using simple random sampling. For the center-pixel data, no clustering was assumed (i.e., cluster size is one pixel). The domain estimation option was used to estimate user's and producer's accuracy rates. This option is a ratio estimator in which the numerator random variable is the indicator variable for a match in the map and field determination for the land cover category $s$, $I(s = t)$, and the denominator random variable is the indicator for the land cover category determined for the map [$G(s)$ for user's accuracy] or the field [ $F(t)$ for producer's accuracy].

## 4. Results

Overall accuracy was estimated to be 69.5% (s.e. = 2.0) using the nine-pixel cluster data. However, the estimated accuracy rates using nine-pixel data varied greatly across land cover categories (Table 2). 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, with water having a high accuracy rate, and smaller land cover classes having 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 3). 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 4) 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 would have a relative large impact. The center-pixel producer' accuracy estimate for mixed forest was 0% because map and field-determined mixed forest pixels were never in agreement at a center pixel, in contrast to the nine pixel data for which field and map matches for mixed forest were observed.

Using the full nine-pixel cluster data clearly provides additional information for rarer settings, as evidenced by the fact that a greater number of nonzero cells in the nine-pixel map by field land cover matrix (Table 3) relative to the center-pixel matrix (Table 5). 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. Standard errors for two producer's accuracy estimates deviated from this pattern, with the standard error for coniferous forest center-pixel estimate over ten times higher than the nine-pixel standard error, and the open water center-pixel standard error about half of the nine-pixel standard error. Overall, these results indicate that substantial gains in precision were made by observing additional data surrounding the center pixel.

## 5. Discussion

A primary goal of this pilot study was to explore the use of the full sample survey framework in accuracy assessment, including sample design, owner contact, field data collection, and analysis. Overall, the study methodology was operationally feasible and provided the basis for statistical estimates that minimized sampling and nonsampling errors given resources. The sampling frame covered the entire study area, regardless of accessibility, avoiding frame bias and providing the foundation to make inferences about the entire study area. 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. Clustering was used at the first stage of sampling and in the observational unit. The PSU clusters minimized travel costs by ensuring a subset of sample pixels were proximal to one another, and the expanded observational unit of a 3 x 3 pixel cluster lowered the cost per pixel for data collection. 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. Historically, a quad sheet (or quarter quad) has been used as a sampling unit at this stage for other GAP accuracy assessment studies. Quad sheets are sufficiently large to avoid overly clustered second stage samples that may reduce the statistical efficiency of the design. At the same time, they are small enough to provide an operational advantage in reducing travel time and workload relative to a systematic or simple random sample. A second alternative that may have worked better is to define the PSU to be a portion of a county, such as a township or set of townships (or equivalent political units in other areas of they US). The size of the PSU should be related to workload units, so that it contains a pixel sample that corresponds, for example, to a dayor week's workload. Using sub-county political borders rather than quadrangles to define a PSU offers two important advantages. First, it would avoid problems that occur when a sample quad intersects with two or more counties, requiring visits to multiple counties to obtain owner information. Second, counties are consistent with state boundaries, and complications associated with combining partial quads at state boundaries would be eliminated.

The choice of a pixel as the second stage sampling unit was simple to work with in the sampling process. The map land cover category stratum identification provided the control needed to address sample size requirements for strata. To balance estimation goals for land cover classes, we used a simple square root allocation rule, with bounds for minimum and maximum sample sizes that prevented too much effort being devoted to large land cover categories and too little effort being devoted to rarer categories. The extra adjustment for land cover classes that were especially easy or difficult to classify allocated sample size resources to land cover categories that were more troublesome and needed further investigation. Although this approach worked well, caution should be used in making allocation rules too complex. It is possible to over-design a sample and generate highly

variable sampling weights, which can lead to reduced precision for the estimated accuracy parameters. Also, when operational resources are limited, it may be difficult to assign adequate sample sizes to each land cover category. In general, it is preferable to combine related land cover categories into one category for the purposes of the study, rather than omitting the land cover category.

When availability of resources and time for completing a field study are in question, dividing the full sample into balanced portions can be very useful. The subsamples provide decision points at which the project team can evaluate resources and choose to stop or to complete an additional subsample. Field staff should be sure to complete entire replicates to retain the full properties of the design. This fact was not made sufficiently clear to our field staff and thus the stopping rule was executed for subsamples within counties rather than for the entire subsample.

Incorporating land cover category strata and a nonresponse adjustment for the sample size (due to inaccessibility or denied access) ensured that adequate sample sizes were obtained when assumptions were consistent with the actual access rates. It is possible to use more complicated nonresponse assumptions that vary in relation to differences in accessibility rates (e.g., physical barriers, denied permission), which may be useful in states with more challenging terrain.

The approach of using a pixel cluster as an observation unit worked reasonably well in the field. Early in the study, a cluster was defined to be a 5 x 5 collection of pixels because of its similarity to the minimum mapping unit (2 ha). However, mapping such a large region proved to be cumbersome and time consuming. The gain in precision of accuracy estimates and the increased ability to gather data for rarer 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. Costs associated with obtaining permission from land owners and travel to sample sites for

each center pixel are high relative to the per-pixel cost and information content for pixels adjacent to the center pixel.

Protocols for contacting landowners had a large impact 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 land owner, offering a printed map of the area as a gift, and calling the land owner before the visit to remind him/her of the project and 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). Effective contacting strategies typically require multiple contacts and multiple contact modes, and usually involve several weeks of effort to obtain high response rates.

In the future, it would be useful to develop written definitions for field-identification of land cover categories to avoid inconsistent application of the land cover classification scheme across field observers. A specific protocol is also needed to address field conditions where more than one land cover category is contained in the 30 m x 30 m pixel area. Examples of possible rules are to record the land cover with the most surface area, record the land cover at the center of the pixel, and so on. The rule set must account for the variety of conditions that exist in the field (e.g., one land cover class, a dominant land cover class, two land cover classes with roughly equal areas, more than two land cover classes with one dominant or with none dominant, etc.) in a manner that promotes unbiased observations.

One of the advantages of the design used is that all land was eligible to be assessed for accuracy. 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. Alternative ground-truthing methods should be approached with the same rigor as the primary observation protocols. In addition, it would be wise to select a sample of pixels on which both the primary and surrogate protocols can be implemented, enabling the impact of alternative ground-truthing methodology to be estimated and possibly adjusted for.

A major concern with the current pilot study is the use of 1999 field data to assess the accuracy of a land cover map derived from 1992 imagery. Temporal differences in land cover can become quite large in this time span, even in a relatively stable environment like the Midwest, confounding assessments of the digital map reflecting 1992 conditions generated from 1999 field observations with temporal effects. Future accuracy assessments are needed during the land cover map update process in a subsequent round of GAP mapping. Presumably, such an activity could be planned in advance, making it possible to collect accuracy assessment reference data in the same calendar year as the year of the satellite imagery used as source materials. The information on the map being updated would also useful in designing an efficient assessment sample.

**Acknowledgements**

The authors wish to thank Steve Stehman for his insightful comments on an earlier draft of this manuscript. This research was funded in part by Assistance Number X997387 01 between Iowa State University and the Environmental Protection Agency. GAP coordinators Jim Merchant (Nebraska), Chris Lauver (Kansas) and David Diamond (Missouri) participated in the design discussions leading up to EPA Region 7 studies for Iowa, Nebraska and Kansas. We thank Courtney Kies, Carsten Botts, and Soledad Fernandez for their assistance with sampling and statistical analyses; and Tom Rosberg and Craig Male for their field data collection efforts.

**References**

Cochran, W. G. (1977) *Sampling techniques.* John Wiley & Sons, New York.

Congalton, R. G. (1988) A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54(5) 593-600.

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

Congalton, R. and Green, K. (1993) A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, 59(5) 641-644.

Crist, P. and Deitner, R. (1998) Assessing land cover map accuracy. *National GAP Analysis Handbook*. USGS/BRD, Idaho Coop. Fish and Wildlife Unit, Univ. of Idaho, Moscow.

Groves, R. M. (1989) *Survey errors and survey costs*, John Wiley & Sons, New York.

Jennings, M. D. (2000) Gap analysis: concepts, methods, and recent results. *Landscape Ecology*, 15, 5-20.

Lessler, J. T. and Kalsbeek, W. D. (1992) *Nonsampling errors in surveys*. John Wiley & Sons, New York.

Lohr, S. L. (1999) *Sampling: design and analysis*. Brooks/Cole Publishing Company, Pacific Grove, CA. 494 pp.

Nusser, S. M. and Klaas, E. E. (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 Dillman, D. A. (1994) *How to conduct your own survey*. Wiley, New York, NY. 232 pp.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992) *Model assisted survey sampling*. Springer-Verlag, New York. 694 pp.

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

Stehman, S. V. (1999) Basic probability sampling designs for thematic map accuracy assessment. *International Journal of Remote Sensing*, 20, 2347-2366.

Stehman, S. V. and Czaplewski, R. L. (1998) Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, 64(3) 331-344.

Figure 1. Accuracy assessment study area in northeastern Iowa, partitioned into quads and primary sampling units (PSUs), with sampled PSUs shaded.

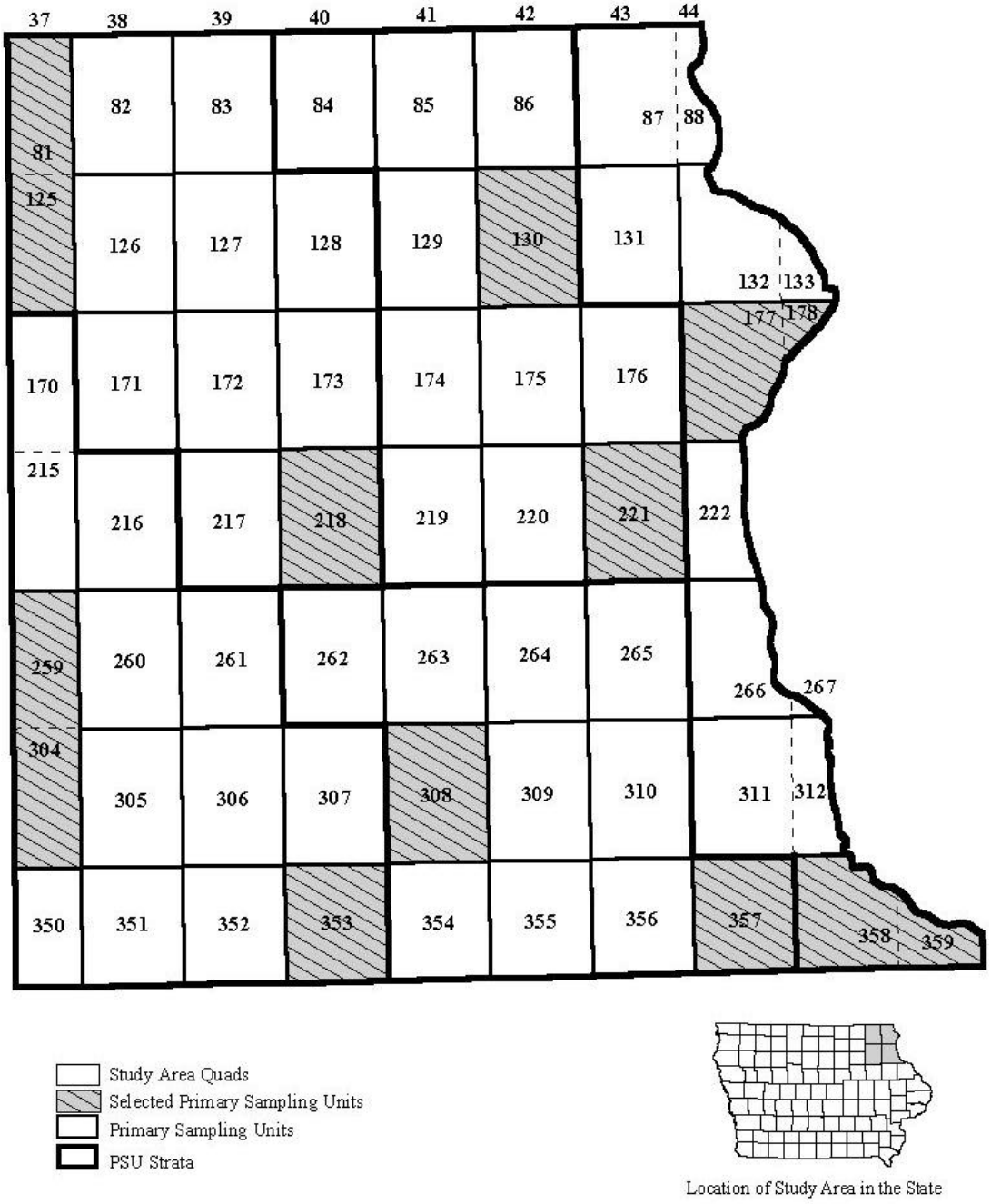

Figure 2. Sampled PSUs and sampled pixels by land cover category. Numeric labels denote quad identification. Subsamples are denotedby symbols, as described in the legend below.

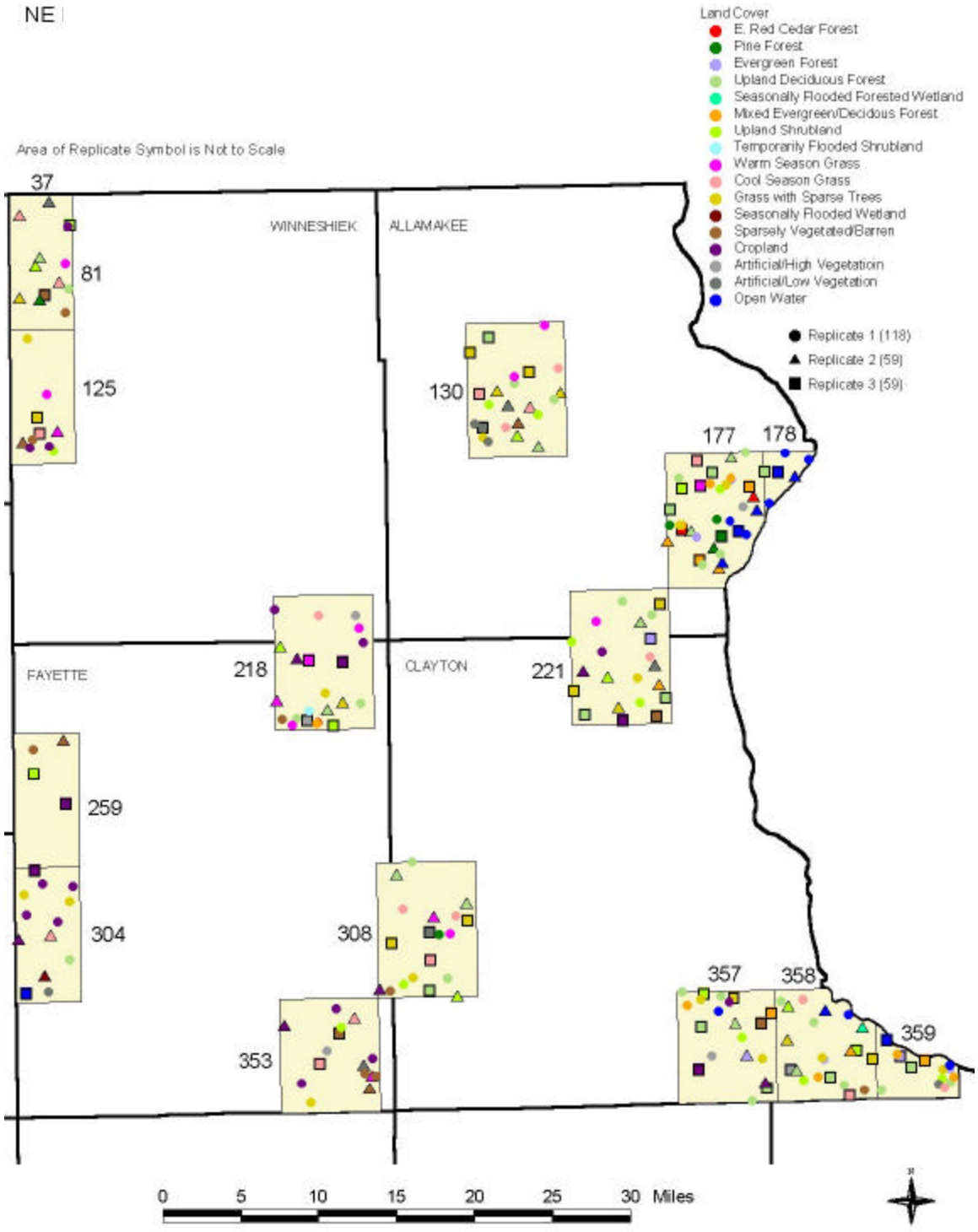

Table 1. Information used to create pixel sample allocation across land cover categories.

| Land Cover Category ($s$) | Total Area In Hectares ($A_s$) | Adjustment Coefficient ($K_s$) | Allocation Weight ($K_s A_s^{1/2}$) | Initial Allocation ($n_s^0$) | Final Allocation ($n_s$) |
|---|---|---|---|---|---|
| Coniferous Forest | 1,362 | 2 | 74 | 12 | 16 |
| Deciduous Forest | 146,846 | 1 | 383 | 61 | 44 |
| Mixed Forest | 2,635 | 1 | 51 | 8 | 16 |
| Shrubland | 5,202 | 2 | 144 | 23 | 24 |
| Grass | 112,282 | 1 | 335 | 53 | 44 |
| Sparsely Vegetated/Barren | 1,723 | 1 | 42 | 7 | 16 |
| Artificial (roads, urban) | 3,678 | 1 | 61 | 10 | 16 |
| Cropland | 451,658 | .5 | 336 | 53 | 44 |
| Open Water | 17,270 | .5 | 66 | 10 | 16 |
| Total | 742,656 | | 1,492 | 236 | 236 |

Table 2. Estimated accuracy rates, standard errors, and sample sizes by land cover category using nine-pixel cluster data.

| Land Cover Category ($s$) | Total Area with Consistent Field and Map Classifications (ha) | Estimated Field Area (ha) | Producer's Accuracy (%) | | | Map Area (ha) | User's Accuracy (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | $\hat{PA}(s)$ | s.e. | $n$ | | $\hat{UA}(s)$ | s.e. | $n$ |
| Coniferous Forest | 326 | 5,464 | 5.9 | (1.9) | 83 | 1,362 | 23.9 | (9.5) | 72 |
| Deciduous Forest | 91,902 | 128,660 | 71.4 | (3.7) | 381 | 146,846 | 62.5 | (3.4) | 371 |
| Mixed Forest | 153 | 1,204 | 12.7 | (8.7) | 23 | 2,635 | 5.8 | (2.9) | 69 |
| 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 | 0.0 | 0.0 | 75 |
| Grass | 7,795 | 13,659 | 57.1 | (7.4) | 55 | 112,282 | 6.9 | (1.5) | 247 |
| Sparsely Vegetated/Barren | 0 | 1,381 | 0.0 | 0.0 | 13 | 1,723 | 0.0 | 0.0 | 36 |
| Artificial (roads, urban) | 3,456 | 32,432 | 10.7 | (3.5) | 136 | 3,678 | 93.9 | (3.3) | 45 |
| Cropland | 402,789 | 499,237 | 80.6 | (2.1) | 536 | 451,658 | 89.2 | (2.1) | 347 |
| Open Water | 9,700 | 10,700 | 90.7 | (4.6) | 73 | 17,270 | 56.2 | (5.1) | 115 |
| Total | 516,121 | 742,656 | | | 1,377 | 742,656 | | | 1,377 |

Table 3. Observed number of pixels in the nine-pixel cluster data for each field and map land cover category combination. [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 | 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 on 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 on the field (related to User's Accuracy).

Table 4. Estimated accuracy rates, standard errors, and sample sizes by land cover category using center pixel data.

| Land Cover Category ($s$) | Total Area with Consistent Field and Map Classifications (ha) | Estimated Field Area (ha) | Producer's Accuracy (%) | | | Map Area (ha) | User's Accuracy (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | $\hat{PA}(s)$ | s.e. | $n$ | | $\hat{UA}(s)$ | s.e. | $n$ |
| 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 |

Table 5. Observed number of center pixels for each field and map land cover category combination.[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 on 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 on the field (related to User's Accuracy).