# Pixels, blocks of pixels, and polygons: choosing a spatial unit for thematic accuracy assessment

Stephen V. Stehman[1] and James D. Wickham[2]

[1] State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210 USA

[2] U.S. Environmental Protection Agency, Environmental Sciences Division, Research Triangle Park, NC 27711 USA

[*] **Corresponding author:**

Stephen Stehman
SUNY College of Environmental Science and Forestry
1 Forestry Drive
Syracuse, NY 13210 USA
Email: svstehma@syr.edu
Voice: 1 315 470 6692
Fax: 1 315 470 6535

**Abstract**

Pixels, blocks of pixels, and polygons are all potentially viable spatial assessment units for conducting an accuracy assessment. We develop a population-based statistical framework to examine how the spatial unit chosen affects the outcome of an accuracy assessment. The population is conceptualized as a difference map created by overlaying a complete coverage reference classification and the target map being evaluated. The per-class areas of agreement and disagreement derived from this population are summarized by a population error matrix and accuracy parameters (e.g., overall, user's and producer's accuracies). The population and values of the accuracy parameters are strongly affected by the protocols implemented for the response design which include the choice of spatial unit, how within-unit homogeneity is addressed when assigning class labels, and the definition of agreement between the reference and map classification. Several complete coverage populations are used to illustrate how accuracy results are affected by the spatial unit chosen for the assessment and also to evaluate how spatial misregistration of the map and reference locations impacts accuracy results for different spatial units. The sampling design implemented for accuracy assessment does not change the population or values of the accuracy parameters, but the choice of spatial unit will influence decisions regarding use of strata and clusters in the design. A universally best spatial assessment unit does not exist, so it is critical to recognize how the population, values of the accuracy parameters, and sampling design are impacted by the choice of spatial unit.

Key Words: response design; stratified sampling; cluster sampling; land cover; change; location error;

## 1. Introduction

Accuracy assessment is an established component of the process of creating and distributing thematic maps. The fundamental basis of an accuracy assessment is a location-specific comparison between the map classification and the ground condition or "reference" classification. Reporting thematic map accuracy in the form of an error matrix is a standard practice when the map and reference classifications are based on categories such as land-cover classes (Story and Congalton 1986; Congalton and Green 1999, 2008). However, opinions vary on the appropriate spatial unit for comparing the map and reference classifications to obtain the data summarized by an error matrix. Pixels, blocks of pixels (i.e., square arrays of pixels), and polygons are the spatial units commonly used. The lack of consensus regarding choice of assessment unit is evident from the 33 map accuracy assessments reviewed by Stehman and Czaplewski (1998), who reported that 14 assessments used a pixel as the spatial unit, 10 assessments used a square block of pixels (e.g., 2x2, 3x3), and 9 assessments used a polygon. Strahler et al. (2006), Janssen and van der Wel (1994) and Richards (1996) support a pixel-based assessment, whereas Congalton and Green (1999) recommend using a block of pixels or a polygon.

Because it is impractical to obtain a census of the reference classification, a sample of units is selected from the region of interest (ROI) and the reference classification is obtained for each sample unit. The response design is the protocol for determining the reference classification of a sampled assessment unit (Stehman and Czaplewski 1998). Key decisions for the response design include choosing the spatial assessment unit, specifying how the reference information will be obtained (e.g., from field observation or high resolution imagery) and stating how agreement between the map and reference classification will be defined. The results of an accuracy assessment are strongly influenced by the choice of spatial unit and definition of agreement. The choice of sampling design does not affect the values of the accuracy parameters, but different sampling designs will differ in terms of the precision of the estimates of the accuracy parameter values.

The objective of this article is to elucidate how the choice of the spatial unit for accuracy assessment affects the implementation and results of the assessment. We develop a population-based conceptual framework to demonstrate how the population and the parameter values targeted by an accuracy assessment are determined by the choice of spatial unit and response design (Section 2). Several example populations provide the basis for illustrating numerically how different choices for the spatial unit and response design translate to different values of the accuracy parameters and therefore different results. The same population framework is then applied to illustrate how the spatial assessment unit impacts the outcome of an accuracy assessment when location error is present (i.e., when the spatial units of the map classification and the spatial units of the reference classifications are not aligned). The practical ramifications of the choice of the spatial unit on the sampling design are discussed in Section 3. The key concepts and results developed in the article are previewed in Table 1.

## 2. A Population Framework for Accuracy Assessment

A statistical population is defined as the collection of all elements of interest and one or more quantities ("variables of study") associated with each element (Särndal et al. 1992, p.5). A parameter is a function of the quantities assigned to each element where this function incorporates all elements of the population. For example, population means, totals, and ratios (e.g., a ratio of means of two variables) are common parameters. For accuracy assessment, a population can be defined as all spatial units forming a partition of the region of interest (ROI) where the spatial unit could be a pixel, a square block of pixels, or a polygon, and the observations or variables of study associated with each spatial unit are obtained from a complete coverage reference classification (i.e., the reference map) and the map to be evaluated (i.e., the target map). The reference map is itself a population in which the observation on each spatial unit is the reference class associated with that unit. Similarly, the map being evaluated may be viewed as a population with the map class assigned to each unit being the observation of interest. The reference population represents the true condition on the ground and the map population represents the classified or predicted condition. The difference between these two populations is the population of interest for accuracy assessment, where an observation of this population could be an indicator variable representing whether a spatial unit is classified correctly (i.e., the variable is assigned the value of 1 if the reference label and map label agree and the value of 0 if the labels disagree) or a quantity representing the degree of agreement between the map and reference classification for that spatial unit. In the case of a binary representation of "agree" or "disagree", the population can be created by overlaying the reference and target maps and determining the per-class area of agreement and the area of disagreement by type of misclassification. It is these class-specific areas of agreement and disagreement that are summarized by the population error matrix commonly used to describe accuracy. Our evaluation will focus on the parameters overall, user's, and producer's accuracies computed from a population error matrix.

The perspective underlying this population framework is developed from the design-based inference framework (Särndal et al. 1992, sections 2.5 and 2.7). In this framework, the observation or measurement taken on each unit of the population is regarded as a fixed constant, not a random variable, and uncertainty is attributable to the randomization distribution resulting from the sampling design (i.e., uncertainty is represented by the variation of the estimate of the parameter of interest over the set of all possible samples). The value of a parameter is computed from a census of the population.

In practice, a reference map does not exist. Consequently, the reference classification will be available for only a sample of the ROI, and the error matrix and accuracy parameters must be estimated from this sample. The estimation objective still targets the error matrix and associated parameters of the population created by overlaying complete coverage reference and target maps. Other parameters that quantitatively compare complete coverage reference and target maps may also be of interest (e.g., Power et al. 2001, Dungan 2006, Hargrove et al. 2006, and White 2006) but will not be discussed further.

The population described above is the foundation of an area-based accuracy assessment in which the objective is to characterize accuracy in terms of area or proportion of area correctly classified and area or proportion of area misclassified. An alternate approach leading to a different population is a per-polygon or per-object accuracy assessment in which the focus is on whether the polygons or objects mapped are classified correctly as individual entities. For example, Smith et al. (2002) investigated the distribution of small water bodies because of their potential significance for altering sediment delivery, and an important feature of this distribution is the number of small water bodies by geographic region. In this example, the accuracy assessment could focus on the small water bodies as countable objects (i.e., distinct individual entities) as opposed to quantifying the area of small water bodies. In such a per-object assessment (in which the objects are defined by the map classification), an object may be considered correctly labeled if the majority of the object's area, as determined from the reference classification, corresponds to the map label. A per-object assessment thus may employ a fundamentally different definition of agreement than an area-based assessment and the different approaches may produce different accuracy outcomes. The focus of this article will be area-based assessment.

The three spatial units we discuss are a pixel, a block of pixels, and a polygon. We assume that each pixel is comprised entirely of a single class, but a block is not similarly assumed to be internally homogeneous and could be comprised of a mix of classes (the "mixed pixel" case would be treated in the same manner as a heterogeneous block). Of the three choices of spatial assessment unit, polygons are commonly viewed as more natural on the basis that they represent real features of the landscape. This viewpoint is appropriate if the polygons are defined by the reference classification, but not if polygons are defined by the map classification. The polygons often used in practice for accuracy assessment are defined by the map classification, and these map polygons are not always features of interest. For example, map polygons may be added, eliminated, or changed as the map is revised prior to completion of the final map product. Congalton and Green (2008, figure 5.5) present an illuminating example of this phenomenon in which a sample map polygon selected on the basis of a preliminary version of a map turns out to include portions of the area of three different polygons of the final map. Consequently, the polygon assessment unit in Congalton and Green's example is not a real feature of the final map and such a polygon would not be relevant to the objective of assessing the accuracy of the final map. Among the options for spatial assessment unit, polygons defined by the reference classification are closest to representing actual earth surface features. However, it is difficult in practice to design and implement an accuracy assessment using reference polygons as the spatial assessment units (see Section 3).

The minimum mapping unit (mmu) specified for the target map may be taken into consideration when choosing the spatial unit. It is not sufficient to state that the mmu is the spatial unit for the assessment because simply specifying the mmu does not unambiguously partition the ROI. One option for taking into account the mmu would be to partition the ROI into square blocks where the area of each block is equivalent to the area defined by the mmu (e.g., a 9-pixel mmu would translate to a 3x3 pixel block unit). However, as is generally true of any partition formed by blocks, some blocks will contain a mix of classes and this within-block heterogeneity will affect the response design, sampling design, and analysis. The mmu obviously influences the partition of the ROI into map polygons because the mmu determines the smallest area that is possible for a map polygon.

Pixel, block, and polygon assessment units are all arbitrary partitions of the ROI. Congalton and Green (2008, p. 70) criticize using a single pixel as the spatial unit for an accuracy assessment stating that a pixel is an arbitrary unit that does not have a meaningful relationship to most of the earth surface features that are mapped. Although Congalton and Green (2008, p. 71) advocate using a block of pixels as the spatial assessment unit, they acknowledge that a block of pixels is also arbitrary. The validity of the results of an accuracy assessment does not depend on whether the assessment unit is a meaningful entity, but instead depends on whether the area representation portrayed by the population error matrix is meaningful. Regular size and shape spatial units (e.g. pixels) partition the landscape into convenient units for analysis. Although surface characteristics such as land cover do not in reality typically conform to such a spatial partition of the landscape, pixels still provide useful information regarding accuracy if the partition formed by these units preserves the areas of agreement and disagreement of the population defined by overlaying the reference map on the target map. For an area-based accuracy assessment, it is the preservation of these areas of agreement and disagreement that is the critical trait required of the spatial partition, and failure to preserve these areas is exacerbated by larger spatial units. Part of the criticism of pixel-based assessments appears to be attributable to the failure to distinguish between area-based and per-object assessments. When conducting a per-object assessment in which a decision is made whether each object is classified correctly, it would generally not be reasonable to examine a single pixel to assess whether the majority of the area of a multiple-pixel object is correctly classified. But for an area-based accuracy assessment in which accuracy is defined in terms of the location-specific area representation of agreement and disagreement, a pixel assessment unit is a legitimate and practical option.

### *2.1 Heterogeneity within spatial assessment units*

For any choice of assessment unit, the decision of whether to treat these units as homogeneous or heterogeneous (according to the reference and map classifications) will exert a strong influence on the population. If the units are treated as homogeneous, a single class label can be assigned to each unit. In contrast, if the units are regarded as heterogeneous, this heterogeneity must be accounted for in the class labeling protocol. For example, the class information provided for a heterogeneous block may be a vector of area proportions such as (0.8, 0.2) indicating that the area of the block is 80% class A and 20% class B. Note that in this case the values 0.8 and 0.2 would still be regarded as constants in the design-based framework, but the response is now a vector of observations instead of a single observation. Alternatively, the class label information could be a fuzzy membership vector where membership is defined as the degree to which the unit belongs to each of the classes (e.g., the vector (0.15, 0.85) indicates membership of 0.15 in class A and membership of 0.85 in class B). When within-unit heterogeneity exists it is still an option to impose an approach that essentially treats the assessment units as homogeneous. For example, a unit with area proportions of 0.8 and 0.2 for classes A and B could be assigned a label of class A and the unit subsequently treated as homogeneous in the analysis.

The definition of agreement and the data analysis will depend on this decision regarding homogeneity of the assessment unit. If the assessment unit is treated as homogeneous in terms of both the map and reference classification (i.e., the simplest case of a single map class and a single reference class) then agreement exists if the map and reference labels match, and if the reference and map classes differ, it is straightforward to specify the type of disagreement. A traditional error matrix analysis is readily applied to the case of homogeneous assessment units. If the assessment unit is treated as heterogeneous, analyses taking into account the mixed character of the unit are typically more complex and less familiar to most practitioners than the error matrix analyses. Binaghi et al. (1999), Lewis and Brown (2001), Pontius and Cheuk (2006) and Kuzera and Pontius (2008) are examples of analyses applicable to heterogeneous assessment units in which the results are summarized by an error matrix and associated accuracy measures. Another approach to quantify agreement between the target map and reference map appropriate for heterogeneous assessment units is to estimate measures such as mean deviation, mean absolute deviation, root mean square error, and correlation. Willmott (1982), Willmott and Matsuura (2006), Ji and Gallo (2006), Pontius et al. (2008), and Riemann et al. (2010) critique these approaches and suggest still other alternatives. It is not our intent to recommend particular measures or to review all analyses potentially applicable to heterogeneous assessment units. Instead, our purpose is to emphasize that an analysis different from the traditional error matrix approach will need to be applied in these cases.

*2.2 Empirical demonstration of the effect of assessment unit on accuracy parameter values*

To quantitatively illustrate the impact of the choices made for the spatial assessment unit and response design, we constructed hypothetical populations for three regions. For the populations labeled "Florida", the target map is the adjusted NLCD 1992 land cover of the United States (Fry et al. 2008) and the reference map is the NLCD 2001 land cover (Homer et al. 2007). The other two regions, both located in North Carolina (USA), are labeled "Fayetteville" and "Dare". For these two regions, the target map is the NLCD 2001 land cover and the reference map is the NOAA C-CAP 2001 land cover. In reality, true complete coverage reference maps do not exist for these regions, so available complete coverage maps such as NLCD and C-CAP are used as hypothetical reference maps for the purpose of creating test populations. The NLCD and C-CAP products have a 30-m x 30-m pixel resolution and each pixel has a single map class label. The land-cover classes used in this analysis are Level I NLCD and C-CAP classes water, urban, barren, forest, shrub, agriculture, and wetland. Detailed descriptions of the land-cover data sources are available at www.mrlc.gov (NLCD) and www.csc.noaa.gov/digitalcoast/data/ccapregional/ (C-CAP). The three target maps are shown in Figures S1-S3 of the supplemental online material.

Three assessment units are investigated in this analysis, a 30-m x 30-m pixel, a 3x3 pixel block, and a map polygon, where the polygons are defined by the target map being assessed (i.e., a polygon is a contiguous area of homogeneous mapped land cover, and contiguity is defined using the four neighboring pixels). The pixel units are homogeneous in terms of both the map classification and reference classification, but block units may be heterogeneous according to either the map or reference classifications, and the map polygon assessment units may be heterogeneous according to the reference classification. When heterogeneity within pixel blocks and polygons is present the mode class is used to label the unit and agreement is defined based on comparing the mode map class and the mode reference class. When more than one class qualifies as the mode for either the reference or map classification, a label is assigned according to the following randomly ordered sequence: forest, wetland, shrub, urban, barren, agriculture, and water. For example, if the map classification for a 3x3 block results in four pixels of forest, four pixels of water, and one pixel of wetland, the class label assigned would be forest because forest precedes water in the list used to decide the mode when ties occur. Other rules for resolving ties could be constructed and would result in different populations. The rule we apply was chosen for simplicity and so that similar blocks agree.

The population areas of disagreement and agreement obtained by overlaying the reference and target maps provide visual evidence of the change in accuracy resulting from the choice of spatial unit (Figures 1-3). The differences in accuracy resulting from using different assessment units are summarized by overall, user's, and producer's accuracies, with all accuracy parameter values (Table 2) computed from the complete coverage map information. The three regions vary in how accuracy results change for different spatial units. For Florida, accuracy does not change substantially for different spatial units, whereas the Fayetteville region shows the largest differences in accuracy with the differences being smaller between the pixel and block assessments relative to the differences between the pixel and polygon assessments. The results for the Dare region are intermediate to the other two regions in that differences in accuracy among the three spatial units exist but these differences are not as large as those observed for the Fayetteville region. With the exception of some erratic differences observed for the very rare classes (e.g., urban in the Dare population and water in the Fayetteville population) the largest difference in accuracy between the pixel and block units occurs for user's accuracy of shrub in the Dare population (11% difference). Changes in accuracy of 5% or more between the pixel and block results for user's accuracy and producer's accuracy occur for several land-cover classes.

The accuracy results for the polygon assessment unit show greater variation from the pixel and block unit results. The differences are most dramatic for the Fayetteville population where overall accuracy is 12% higher for the polygon assessment compared to the pixel assessment. The class-specific accuracies of the polygon assessment can be much different from the pixel and block results. For example in the

Dare population, user's accuracy of forest is 85% for the polygon unit compared to 69% and 68% for the pixel and block units, and producer's accuracy of forest is 95% for the polygon unit compared to 79% and 77% for the pixel and block units. Producer's accuracy of shrub is another case in which class-specific accuracy is much higher for the polygon assessment than for the pixel or block assessment. For the Florida and Dare populations, the polygon accuracy results are generally slightly higher than the pixel and block results.

The results in Table 2 are intended to provide case study examples demonstrating the dependence of accuracy results on the choice of spatial unit. Additional comparisons of accuracy parameter values based on different spatial units for several additional populations are presented in the Appendix. We have not attempted to discover or model how accuracy results for a particular ROI will vary depending on the spatial assessment unit used. The size and shape of land-cover patches and accuracy of the classification likely interact with other factors in a complicated manner to determine the sensitivity of the accuracy results to the choice of assessment unit. A useful direction for future research would be to investigate if the sensitivity of accuracy results to choice of spatial assessment unit can be modeled as a function of landscape structure, classification accuracy, and other factors.

### *2.3 Spatial Registration (Location) Error*

Congalton and Green (2008) criticize the use of a pixel as an assessment unit, asserting that it is too sensitive to positional errors that arise from fitting remotely sensed data to a map surface. Registration errors affect all spatial units, and no spatial unit is entirely free of location error (see, for example, McRoberts 2010). Conceptually, spatial misregistration is a "halo" around the assessment unit. For example, a ± 1 pixel root mean square error (RMSE) means that the single pixel, the entire block of pixels, or the entire polygon could be shifted ± 1 pixel along the x-dimension, y-dimension, or both dimensions. Larger spatial units might be expected to be less sensitive to registration error because the proportion of the total area of the spatial unit that is affected by location error decreases as the size of the spatial unit increases. However, this decrease in sensitivity develops slowly as a function of increasing size. For example, decreased sensitivity to spatial location error is modest for small pixel blocks as illustrated by Figure 4. Panel A of Figure 4 depicts a 3x3 pixel block overlaid on a Landsat composite. An interpreter collecting reference data would use the information in Panel A to locate the sample unit on the reference medium. Once the sample unit had been located, the interpreter would use the reference information (Panel C) to assign a reference label. If spatial location error exists, however, the sampling unit could be shifted on the reference medium (Panel D). This misalignment of the map and reference data is shown as a shift of the sample block location between Panel C and Panel D, with a change in the mode reference class possibly resulting from the shift.

The ultimate impact of location error is determined by the change in the population and values of the accuracy parameters resulting from location error relative to the population and parameter values when location error is absent. To evaluate the effect of location error, we shifted all three reference maps by one pixel down and one pixel to the left. The target map is unchanged by this shift in the reference map, but the shift introduces location error and changes the population and values of the accuracy parameters from the parameter values of the spatially aligned population (Table 2). The accuracy parameter values resulting from the location-error impacted difference populations are shown in Table 3. In general, the effect of location error is greatest for the pixel assessment unit followed by the block and polygon units. For all three regions, location error produces a decrease in overall accuracy. The Florida population showed the largest changes in accuracy attributable to location error with the decrease in overall accuracy of 28% for the pixel unit, 20% for the block unit, and 12% for the polygon unit. Large decreases in class-specific accuracy were also observed for the Florida population. For example, user's accuracy for shrub decreased by 38% for the pixel unit, 28% for the block unit, and 27% for the polygon unit, whereas producer's accuracy for shrub decreased roughly 40%, 30%, and 20% for the pixel, block, and polygon units. Relative to the Florida population, the Fayetteville and Dare populations generally had smaller decreases in accuracy caused by shifting the reference map. More common land-cover classes generally show smaller decreases in accuracy between the aligned and shifted reference maps. For example, urban and agriculture are the two most common classes in the Fayetteville population and the decreases in user's and producer's accuracies for these classes caused by location error are generally among the smaller decreases observed (Table 3b). Similarly, forest and wetland are the two most common classes in the Dare population and the decreases in accuracy for these classes are generally small, although the Dare population is the least affected population in terms of results changing because of location error.

Use of an assessment unit larger than a pixel to mitigate the impact of location error is supported to some degree by these results (Table 3). However, the decreases in overall accuracy for both the block and polygon units provides evidence that these units are also susceptible to substantial errors in accuracy attributable to spatial misregistration between the map and reference locations. The effect of location error when using blocks or pixels is also evident for user's and producer's accuracies since most land-cover classes have lower values for these parameters for the spatially misaligned population.

## 3. The impact of spatial assessment unit on sampling design and estimation

### *3.1 Defining a sampling universe and frame*

Constructing a sampling design requires specifying a sampling universe, defined as the set of spatial units that form a partition of the ROI. In practice, a universe of pixels, blocks of pixels, or map polygons is straightforward to construct. Because the universe must be spatially exhaustive, it is not valid to exclude heterogeneous areas as is sometimes done to avoid location error issues. The pixels, blocks, or polygons forming the universe can be represented by a "list frame", defined as a list of all such units in the ROI along with a spatial address for each unit (e.g., spatial coordinates or an identification number unique to each unit). Because pixels and blocks provide a partition of the ROI independent of the map or reference classification, it is possible to construct a list frame of pixels or blocks before the map is finalized. A list frame of map polygons could be readily produced from either a final or preliminary map, although using a preliminary map is a questionable practice because not all of these map polygons will exist in the final map. If polygons are defined by the reference classification, it is not feasible to construct a complete list frame of the universe of reference polygons because a census of the reference classification would be required. Protocols for implementing basic sampling designs such as simple random, stratified random, systematic, and cluster sampling from a list frame are described in Cochran (1977), Lohr (1999), and Särndal et al. (1992). Stehman (1999, 2009) provides a general overview of sampling designs applicable to accuracy assessment.

*3.2 Stratified sampling and cluster sampling*

Two important considerations when choosing the sampling design are whether to group the spatial assessment units into strata to control the sample size allocated per stratum (for the purpose of decreasing standard errors of class-specific accuracy estimates) and whether to group the units into clusters to spatially constrain the sample within these clusters (for the purpose of decreasing costs associated with travel time to field sites or costs associated with the number of aerial photographs or very high resolution images required to obtain the reference data). The choice of a pixel, block, or polygon assessment unit has ramifications on how such stratified and cluster sampling designs would be implemented. We will address the stratification and clustering considerations separately.

Typically in accuracy assessment sampling designs the strata correspond to the mapped area of each class (e.g., all area mapped as forest is one stratum) and the sample sizes allocated to rare class strata are chosen to achieve specified standard errors of the user's accuracy estimates. If each assessment unit has a single map class, as is the case for a pixel or a map polygon, it is a simple matter to assign each unit to a single stratum corresponding to the unit's map class. Stratification is more complicated for a block assessment unit because not all blocks are comprised of a single map class, and a protocol stating how to assign each heterogeneous block to a stratum must be specified. A variety of assignment rules could be envisioned to create the stratification of blocks (e.g., a block could be assigned to a stratum based on the most common map class within the block or based on the class associated with the center of the block), but the effectiveness of different stratum assignment rules has not been investigated.

A cluster is a group of pixels, blocks, or polygons that is treated as a single entity in the sampling protocol. Pixel and block assessment units can be easily grouped into regularly-shaped clusters so pixel and block units conveniently fit the nested structure of cluster sampling. Forming clusters of polygons is more cumbersome than forming clusters of pixels or clusters of blocks. Polygons vary in size and shape, so there is no intuitive way to group polygons to form clusters so that the clusters are approximately uniform in size and shape. Polygons lack the regular size and adjacency features possessed by pixels and blocks that allow easy nesting of pixels or blocks within same size clusters. The advantage of cluster sampling to constrain the sample is diminished for polygons because of the variation in the size of the clusters formed. For example, a cluster of four very large polygons covers a much larger area than a cluster of four small polygons. Analysis of cluster sampling designs is considerably simpler when the clusters are equal in size so analyzing a cluster sample of polygons will be more complex relative to analyzing a cluster sample of pixel or block assessment units.

Note that a cluster of pixels appears at first glance to be the same as a block assessment unit but a cluster of pixels is treated differently from a block unit. The response design associated with a block assessment unit does not produce a reference label for each pixel within the block but rather assigns the reference classification to the block as a single entity. In contrast, a cluster comprised of pixels would include a reference classification for each pixel in the cluster, so a pixel remains the assessment unit within the cluster, and the cluster is the primary sampling unit (Stehman 1997).

*3.3 Point sampling to select pixel, block, or polygon units*

An alternative to selecting the sample from a list frame is spatial point sampling. In this protocol, a spatial sample of points is first selected within the ROI and the pixel, block, or polygon within which a sample point falls is selected into the sample. Point sampling is effectively the same as sampling from a list frame if the spatial units partitioning the ROI are all equal in area and have the same shape, as would be the case for pixel or block assessment units, and the decision of whether to use a list frame or point sampling approach with pixel or block assessment units would be made on the basis of convenience of implementation. In contrast, for a polygon assessment unit, the point sampling protocol will select polygons into the sample with probability proportional to polygon area, so larger polygons will have a higher probability of being selected (Figure 5). These unequal inclusion probabilities must be accounted for in the analysis.

The point sampling approach provides a way to select a sample when it is impractical to construct a list frame. For example, if the spatial unit is a polygon defined by the reference classification (i.e., a "reference polygon"), a list frame of all such polygons is not available. To obtain a sample of reference polygons by point sampling, the sample point locations are selected, and the reference polygon that contains each sample point is delineated based on the reference classification (Figure 5). Only those reference polygons intercepted by sample points would need to be identified, so it is no longer necessary to create a list of all reference polygons comprising the population.

Stratified and cluster sampling would not be viable options when point sampling is used to select reference polygons. Stratification requires assigning all reference polygons in the entire ROI to strata based on each polygon's reference class label, and obtaining this information would be impractical because it is tantamount to a census of reference polygons. Two-phase sampling (Cochran 1977) in which the stratification assignment is required only for a large first-phase sample instead of a census ameliorates this practical disadvantage associated with a reference polygon assessment unit. Cluster sampling of reference polygons is not viable because all reference polygons within a cluster would need to be delineated, and this places an impractical burden on reference data collection.

### *3.4 Estimation*

The values of the population parameters resulting from the choice of response design and assessment unit are estimated from the sample. The sampling design has no effect on these parameter values because in the design-based inference framework the value of a population parameter remains fixed regardless of the particular sample selected. It is the <u>estimate</u> of that fixed parameter value that changes depending on the sample and sampling design. It is critical to recognize that the value of a target parameter (e.g., overall, user's, or producer's accuracy) is determined by the choice of spatial assessment unit and response design (see Section 2), not by the sampling design. The specific formula used to estimate a parameter will differ depending on the sampling design and the precision of an estimate will vary for different sampling designs.

The theory of probability sampling ensures the availability of an estimator that is either unbiased or consistent for the parameter of interest. If the accuracy estimator is a Horvitz-Thompson estimator (Stehman 2001, p. 728), it is an unbiased estimator (Särndal et al. 1992, p. 43). Sometimes an unbiased estimator is not available, as may occur when estimating a ratio of two parameters (e.g., producer's accuracy is the estimated area of agreement for a specific class divided by the estimated area of that class according to the reference classification). In such cases, a consistent estimator can be constructed (Särndal et al. 1992, Sec. 5.3; Overton and Stehman 1995), where a consistent estimator is one in which "… the sampling distribution of the estimator can be considered tightly concentrated around $\tau$ [the parameter], when $n$ [the sample size] is large enough" (Särndal et al. 1992, p. 166). Thus an unbiased estimator guarantees that the parameter of interest is estimated correctly on average (averaging over all possible samples that could be selected), and a consistent estimator (although not necessarily unbiased) ensures that the estimate for a particular sample will not stray too far from the true value of the parameter.

The values of the parameters targeted by the sample-based estimators depend on the population created by the overlay of the target map and reference map and the definition of agreement specified by the response design (see Section 2). Although sampling theory exists to support unbiased or consistent estimation of any population parameter, the specific estimator formulas for some combinations of parameter and sampling design may need to be derived. Typically estimators used in accuracy assessment are presented only for simple random sampling (Congalton and Green 2008) and different estimators are needed for other sampling designs (e.g., Card 1982, Stehman 1996, Stehman and Foody 2009). Deriving variance estimators may also be necessary, again most likely when the design is not simple random sampling.

To reiterate the key message of this section, once the population and values of the parameters are determined by the choice of response design and spatial unit, the sampling design and data analysis protocol can provide statistically defensible estimates of these parameters. Any probability sampling design will permit either an unbiased or a consistent estimator of a parameter, so there is no distinction among sampling designs relevant to this feature of estimation. Different sampling designs will yield different values for the variance of an estimator of a particular accuracy parameter, so the comparison of sampling designs should be on the basis of the variance of the estimator and not bias (or consistency) of the estimator.

## 4. Additional Considerations

### *4.1 Block assessment units*

Choosing a 3x3 pixel block as the assessment unit and defining agreement based on the mode class of the block results in the map population assessed being different from the map provided to users. Defining agreement based on the mode class of a 3x3 pixel block implies assessment of a map that has been "filtered" to produce a classification at the support of a block (e.g., an area of 900 $m^2$ for a 30-m x 30-m pixel). For example, suppose a 3x3 pixel block has five of the nine pixels labeled as forest according to the reference classification, and five of the nine labeled as forest according to the map, but only one pixel in common is forest according to both the map and reference classification. In terms of total forest area of the block, agreement is 55% (5 out of 9), yet the overlay of the target and reference maps for the block would show only 11% agreement of common overlapping area of forest (1 out of 9 pixels). The mode class for both the map and reference classification is forest, so if agreement is defined by comparing the mode class, the block would be classified correctly even though only a single pixel has forest for both the map and reference classifications.

Czaplewski (2003) provides a pointed criticism of accuracy assessments using data aggregated to a block level when the map provided to users is not similarly aggregated. The mismatch between the spatial support at which the accuracy assessment is conducted and the spatial support of the map brings into question the utility of the accuracy results obtained from such an approach. In contrast, response designs that use a single pixel as the spatial unit assess the map exactly as it is distributed to users. These considerations reinforce the importance of clearly specifying the population that is the targeted objective of the accuracy assessment.

*4.2 Polygon assessment units*

Use of a map polygon assessment unit introduces concerns not present for pixel or block assessment units because the utility of a map polygon for accuracy assessment is inseparably dependent on the target map. A map polygon may become obsolete for accuracy assessment if the map undergoes a revision that changes the map polygons forming the original partition of the ROI that existed when the sample of map polygons was selected. The most intuitive example of revision would be aggregation of classes to form a simplified legend (e.g., Anderson et al. (1976) Level II to Level I) and a re-analysis of the data based on the aggregated classification (e.g., Pontius and Malizia 2004). Class aggregation re-draws the map polygons, so a sample of map polygons based on a Level II classification would not necessarily provide a sample of map polygons that existed for the Level I classification. Because the sampling and response designs are directly linked to the specific map polygons partitioning the ROI, it would generally not be possible to use a partition based on Anderson Level II polygons to estimate accuracy for the map of Level I polygons. Pixels and blocks of pixels retain their utility even if the map classes are aggregated or the map is otherwise revised because their spatial boundaries remain defined and fixed despite label changes.

Because polygon assessment units vary in size, stratifying the population by polygon size merits consideration. For example, for the Florida population, the map polygons (obtained from the NLCD 1992) range in size from 0.09 ha (a single pixel) to 139.95 ha, and the largest 2.3% of the polygons account for 50% of the total area. For the reference map (obtained from the NLCD 2001), the range in polygon size is 0.09 ha to 231.93 ha, with the largest 2.0% of the polygons accounting for 50% of the total area. An equal probability sampling design (e.g., simple random, systematic, or stratified random with proportional allocation) will result in a sample with the same size distribution of polygons as the population, so a large proportion of the sample will consist of small polygons. Stratification by polygon size could be used to increase the proportional representation of larger polygons in the sample, but it is unclear whether there is an advantage to increasing the sample size of large polygons as no studies have examined the impact of different sampling design choices on the standard errors of accuracy estimates obtained from a polygon-based assessment. Further, stratification by polygon size would likely be combined with stratification by map class and this two-way stratification increases the overall complexity of the sampling design and analysis.

Very large polygons also introduce challenging problems for the response design. Obtaining the reference classification for the entire area of a very large polygon may be too expensive or impractical, so a portion of the polygon may be sampled and the reference characteristics of the polygon estimated from the sample. There is little research on how these factors affect accuracy assessment results.

## 5. Summary

The three most commonly used spatial units for accuracy assessment are a pixel, a block of pixels, and a map polygon. While a universally best spatial assessment unit does not exist, the choice of spatial unit has broad implications on the conduct and outcome of the assessment. The results of a map accuracy assessment depend on the spatial unit chosen to serve as the basis of the assessment because different spatial units lead to different populations and values of the accuracy parameters. The population perspective (see Section 2) provides a rigorous conceptual framework for evaluating the impact of the choice of spatial unit. Specifically, the population of interest in accuracy assessment may be viewed as the result of overlaying a complete coverage reference map with the map to be evaluated and quantifying the class-specific areas of agreement and disagreement between the reference map and the target map. The values of the accuracy parameters computed from this population are the quantities estimated from the sample of reference data. Greater awareness of this population framework will clarify the ramifications of the choice of spatial unit on the outcome of an accuracy assessment (Table 1).

The focus of this article is area-based accuracy assessments in which the area of each land-cover type correctly classified and the area incorrectly classified by type of misclassification are the primary data for describing accuracy. These areas are summarized by an error matrix and the accuracy parameters derived from the error matrix. The difference map created by overlaying a complete coverage reference classification on the target map generates the population (i.e., the per-class areas of agreement and the areas of each class-specific type of disagreement). The ROI is then partitioned by the spatial unit chosen for the accuracy assessment.

As the smallest spatial assessment unit, a pixel best preserves the population areas of agreement and disagreement when the ROI is partitioned. Many accuracy assessments have been conducted using a pixel as the assessment unit. Therefore, specific details of the sampling design, response design, and analysis protocols associated with implementing a pixel-based assessment have been extensively developed and applied. A pixel-based assessment easily accommodates sampling designs employing strata or clusters, whereas blocks are less amenable to stratification and polygons are less practical to use in cluster sampling. The traditional error matrix analyses are readily implemented for a pixel-based assessment.

Blocks and polygons are less likely than pixels to be homogeneous so the response design and analysis protocols must be more complex to account for within-unit heterogeneity. A common practice when using a block or polygon assessment unit is to revert to the protocols developed for a pixel-based assessment and to assume (although rarely explicitly stated) that the block or polygon units are homogeneous. In the likely case that within-unit heterogeneity is present, the mode class is often assigned as the class label and this labeling protocol changes the areas of agreement and disagreement defining the population and correspondingly changes the values of the accuracy parameters. This is a critical feature of accuracy assessment that must be recognized. Assessments based on block or polygon units generally do not preserve the areas of agreement and disagreement that would be obtained by overlaying the unpartitioned target map and the unpartitioned reference map.

The accuracy results for the example populations (Table 2) illustrated a range of outcomes from little change in accuracy with different spatial units (Florida population) to substantial differences in accuracy with different spatial units (Fayetteville and Dare populations). The impact of choice of spatial unit on the results of an accuracy assessment is not only important for single map assessments, but also for comparative studies of accuracy. For example, a researcher evaluating the performance of a new classifier by comparison against an existing classifier should consider the spatial unit used in the assessment of the two classifiers (and perhaps use the same unit to avoid confounding sources of uncertainty).

The choice of spatial unit to serve as the basis of an accuracy assessment is a critical decision. The spatial unit must be chosen with the understanding that the population and therefore the values of the accuracy parameters describing the population are determined by the spatial unit in combination with the response design. Further, the sampling design must be appropriate for the spatial unit chosen. Sampling designs using strata and clusters that are commonly and easily implemented for a pixel unit are more cumbersome to implement when using block or polygon units. Better recognition of the impacts of the choice of spatial unit and the advantages and disadvantages of each unit will lead to better accuracy assessment methodology and improve the validity of the results.

**Acknowledgments**

We thank four anonymous reviewers for their helpful and constructive comments. SVS acknowledges funding support through an Intergovernmental Personnel Agreement (IPA) from the U.S. Geological Survey Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. This article has not been subject to review by the U.S. Environmental Protection Agency or U.S. Geological Survey and does not necessarily reflect the views of either agency.

**References**

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data, U.S. Geological Survey Professional Paper 964, Washington, DC: U.S. Geological Survey, 28 p.

Binaghi, E., Brivio, P. A., Ghezzi, P., & Rampini, A. (1999). A fuzzy set-based accuracy assessment of soft classification, *Pattern Recognition Letters*, *20*, 935-948.

Card, D. H. (1982). Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering and Remote Sensing*, 48, 431-439.

Cochran, W. G. (1977). *Sampling Techniques* (3rd edn.). New York: John Wiley & Sons.

Congalton, R. G., & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL: CRC Press.

Congalton, R. G., & Green, K. (2008). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (2nd edn.). Boca Raton, FL: CRC Press.

Czaplewski, R. L. (2003). Accuracy assessment of maps of forest condition: Statistical design and methodological considerations. In *Remote Sensing of Forest Environments: Concepts and Case Studies* (Wulder, M. A., & Franklin, S. E., eds.). Boston: Kluuwer Academic Publishers, pp. 115-140.

Dungan, J. L. (2006). Focusing on feature-based differences in map comparison, *Journal of Geographical Systems*, *8*, 131-143.

Fry, J. A., Coan, M. J., Homer, C. G., Meyer, D. K., & Wickham, J. D. (2009). Completion of the National Land Cover Database (NLCD) 1992-2001 Land Cover Change Retrofit Product. U. S. Geological Survey Open-File Report 2088-1379 (available from http://pubs.usgs.gov/of/2008/1379, accessed April 2011).

Hargrove, W. W., Hoffman, F. M., & Hessburg, P. F. (2006). Mapcurves: a quantitative method for comparing categorical maps. *Journal of Geographical Systems*, *8*, 187-208.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C. et al. (2007). Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, *73*, 337-341.

Janssen, L. L. F., & van der Wel, F. J. M. (1994). Accuracy assessment of satellite derived land-cover data: A review. *Photogrammetric Engineering and Remote Sensing*, *60*, 419-426.

Ji, L., & Gallo, K. (2006). An agreement coefficient for image comparison, *Photogrammetric Engineering and Remote Sensing,* 72:823-833.

Kuzera, K., & Pontius, R. G., Jr. (2004). Categorical coefficients for assessing soft-classified maps at multiple resolutions. *In* Proceedings of the joint meeting of the 15th annual conference of The International Environmetrics Society and the 6th annual symposium on Spatial Accuracy Assessment in the Natural Resources and Environmental Sciences, 28 June – 1 July 2004, Portland, ME.

Lewis, H. G., & Brown, M. (2001). A generalized confusion matrix for assessing area estimates from remotely sensed data. *International Journal of Remote Sensing*, 22, 3223-3235.

Lohr, S. L. (1999). *Sampling: Design and Analysis*. New York: Duxbury Press.

McRoberts, R. E. (2010). The effects of rectification and Global Positioning System errors on satellite image-based estimates of forest area, *Remote Sensing of Environment*, *114*, 1710−1717.

Overton, W. S., & Stehman, S. V. (1995). The Horvitz-Thompson theorem as a unifying perspective for probability sampling: with examples from natural resource sampling. *American Statistician*, *49*, 261-268.

Pontius, R. G., Jr., & Cheuk, M. L. (2006). A generalized cross-tabulation matrix to compare soft-classified maps at multiple spatial resolutions. *International Journal of Geographical Information Science*, *20*, 1-30.

Pontius, R. G., Jr., & Malizia, N. R. (2004). Effect of category aggregation on map comparison. In *Lecture Notes in Computer Science 3234: Proceedings of the Third International Conference, GIScience 2004* (Egenhofer, M. J., Freksa, C., & Miller, H. J., eds). Adelphi, Maryland (USA): Springer, pp. 251-268.

Pontius, R. G., Jr., Thontteh, O., & Chen, H. (2008). Components of information for multiple resolution comparison between maps that share a real variable. *Environmental and Ecological Statistics*, *15*, 111-142.

Power, C., Simms, A., & White, R. (2001). Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, *15*, 77-100.

Richards, J. A. (1996). Classifier performance and map accuracy, *Remote Sensing of Environment*, *57*, 161-166.

Riemann, R., Wilson, B. T., Lister, A., & Parks, S. (2010). An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data. *Remote Sensing of Environment*, *114*, 2337-2352.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model-Assisted Survey Sampling*. New York: Springer-Verlag.

Smith, S. V., Renwick, W. H., Bartly, J. D., & Buddemeier, R. W. (2002). Distribution and significance of small, artificial water bodies across the United States landscape. *The Science of the Total Environment*, *299*, 21-36.

Stehman, S. V. (1996). Estimating the Kappa coefficient and its variance under stratified random sampling. *Photogrammetric Engineering and Remote Sensing*, *62*, 401-407.

Stehman, S. V. (1997). Estimating standard errors of accuracy assessment statistics under cluster sampling. *Remote Sensing of Environment*, *60*, 258-269.

Stehman, S. V. (1999). Basic probability sampling designs for thematic map accuracy assessment, *International Journal of Remote Sensing*, *20*, 2423-2441.

Stehman, S. V. (2001). Statistical rigor and practical utility in thematic map accuracy assessment, *Photogrammetric Engineering and Remote Sensing*, *67*, 727-734.

Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, *30*, 5243-5272.

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

Stehman, S.V., & Foody, G. M. (2009). "Accuracy Assessment" (Chapter 21). In *The SAGE Handbook of Remote Sensing* (Warner, T. A., Nellis, M. D., & Foody, G. M.. eds.). London: Sage Publications Ltd., pp. 297-309.

Story, M., & Congalton, R.G. (1986). Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, *52*, 397-399.

Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., et al. (2006). Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps, EUR 22156 EN – DG, Office for Official Publications of the European Communities, Luxembourg, 48 pp.

White, R. (2006). Pattern based map comparisons. *Journal of Geographical Systems*, *8*, 145-164.

Willmott, C. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, *63*, 1309-1313.

Willmott, C. J., & Matsuura, K. (2006). On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science*, 20, 89-102.

**Table 1.** Useful concepts and results to guide selection of an accuracy assessment spatial unit.

**General**: The three primary components of an accuracy assessment are the response design, sampling design, and analysis (Stehman and Czaplewski 1998). The choice of spatial assessment unit (e.g., pixel, block of pixels, or polygon) must be considered in terms of the ramifications on all three components.

- **Defining the population and accuracy parameters**

    - The population that determines the values of the accuracy parameters targeted by the assessment is conceptualized as resulting from overlaying a complete coverage reference classification (i.e., a reference map) and the target map to be evaluated. The population may be viewed as a difference map resulting from this overlay showing where the map and reference classifications agree and where they disagree. The corresponding areas of class-specific agreement and class-specific disagreement may be obtained from the difference map. The difference map will depend on the partition of the region of interest (ROI) created from the spatial unit chosen for the assessment. Different populations will yield different values of the accuracy parameters (e.g., different overall accuracy values).

    - For an area-based accuracy assessment, the per-class area of agreement and area of disagreement are summarized by a population error matrix where the cells of the error matrix represent the proportion of area of agreement for each class and the proportion of area misclassified for each type of error. The population and accuracy parameter values obtained from the population error matrix will differ depending on the choice of spatial unit. Changing the spatial unit changes the population and consequently the results of the accuracy assessment.

    - Pixels, blocks of pixels and polygons are all arbitrary spatial units. The validity of an accuracy assessment does not depend on whether the spatial assessment unit is a real surface feature of the earth but instead depends on the area representation of agreement and disagreement resulting from use of the spatial unit to partition the ROI.

    - The population and values of the accuracy parameters are not affected by the choice of sampling design.

- **Response design – determining the reference classification and definition of agreement**

    - Heterogeneity within the spatial unit requires specification of how heterogeneity will be accommodated in the labeling protocol and definition of agreement. These decisions will have a substantial impact on the population and values of the accuracy parameters. As a general rule, the likelihood of heterogeneity within a spatial unit will increase with size of the unit.

    - Map polygons do not necessarily represent real earth surface features and use of map polygons does not support reporting of accuracy results for different levels of class aggregations. The potentially large variation in polygon size creates practical challenges to the response design protocol and also motivates consideration of stratifying the sampling design by polygon size. Little research has been conducted exploring how variation in polygon size affects reference data collection, sampling, and analysis.

    - Polygons defined by the reference classification ("reference polygons") are appealing because they represent real objects of interest but present practical challenges when constructing the response and sampling designs. Methodological developments are needed before reference polygons can be considered a viable option.

- **Sampling design and estimation**

    - The sampling design requires specifying the universe of all spatial units forming a partition of the ROI. The spatial units making up the universe must be non-overlapping and spatially exhaustive.

    - For a stratified design, each spatial unit must be assigned to one and only one stratum. Block assessment units may be internally heterogeneous and this complicates the protocol for assigning each block to a stratum.

    - Polygons are less amenable than pixel or block assessment units for the purpose of cluster sampling because polygons have no natural grouping or nesting structure to form clusters.

- It is possible to construct an unbiased or a consistent estimator for any accuracy parameter of interest when a probability sampling design is implemented and working within the design-based inference framework. Consequently, it is pointless to consider comparing sampling designs on the basis of bias or consistency of estimators derived for different designs.

- Different sampling designs result in different precision (i.e.,variance) for the estimator of a given accuracy parameter, so it is meaningful to compare sampling designs on the basis of variance of the accuracy estimators.

- **Location or registration error (map and reference data are not spatially aligned)**

  - Registration errors affect all spatial units. Spatial misregistration can be conceptualized as a "halo" around the assessment unit.

  - The ultimate impact of location error is observed in the change in the population and values of the accuracy parameters relative to the parameter values of a population determined by overlaying perfectly spatially aligned reference and target maps (i.e., location error is absent).

  - Pixel-based assessments are generally more sensitive to location error than are block- and polygon-based assessments as evidenced by larger changes in the values of the accuracy parameters between spatially aligned and spatially unaligned reference and target maps. However, location error can still have a considerable effect on accuracy results when block or polygon units are used

Table 2. Parameter values for user's and producer's accuracies resulting from different spatial assessment units, pixel, 3x3 pixel block, and map polygon. If a block or polygon is heterogeneous in terms of the map or reference classification, the mode class is used and agreement is defined as a match between the mode map class and the mode reference class. Parameter values for block and polygon units are reported as deviations from the pixel-based parameter value. Positive deviations indicate higher accuracy for the block- or polygon-based assessment, and negative deviations indicate lower accuracy for the block- or polygon-based assessment. Classes are ordered by decreasing percent of area mapped.

a) Florida

| | | Mean | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Map | Patch | User's Accuracy (%) | | | Producer's Accuracy (%) | | |
| Class | Area (%) | Size(ha) | Pixel | Block | Polygon | Pixel | Block | Polygon |
| Wetland | 20 | 16.6 | 100 | 0 | 0 | 55 | -6 | -6 |
| Agriculture | 20 | 4.8 | 99 | 0 | 0 | 98 | 0 | +1 |
| Forest | 19 | 3.8 | 60 | -5 | -3 | 86 | -2 | -1 |
| Urban | 18 | 20.1 | 100 | 0 | 0 | 95 | -2 | +5 |
| Shrub | 16 | 2.3 | 85 | -3 | 0 | 92 | -1 | +5 |
| Water | 7 | 12.4 | 92 | 0 | +6 | 100 | 0 | 0 |
| Barren | <1 | 0.9 | 100 | 0 | 0 | 100 | 0 | 0 |
| Overall accuracy | | | 86 | -3 | -1 | | | |

b) Fayetteville

| | | Mean | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Map | Patch | User's Accuracy (%) | | | Producer's Accuracy (%) | | |
| Class | Area (%) | Size(ha) | Pixel | Block | Polygon | Pixel | Block | Polygon |
| Agriculture | 32 | 6.8 | 86 | +2 | +5 | 83 | +2 | +6 |
| Urban | 31 | 29.2 | 72 | +4 | +18 | 85 | +2 | +5 |
| Forest | 17 | 2.6 | 69 | -1 | +16 | 79 | -2 | +16 |
| Wetland | 10 | 3.3 | 85 | 0 | +12 | 65 | 0 | +14 |
| Shrub | 9 | 0.9 | 68 | -7 | +6 | 54 | -5 | +29 |
| Water | <1 | 1.7 | 89 | +2 | +5 | 61 | +16 | +20 |
| Overall accuracy | | | 77 | +1 | +12 | | | |

c) Dare

| Class | Map Area (%) | Mean Patch Size(ha) | User's Accuracy (%) Pixel | Block | Polygon | Producer's Accuracy (%) Pixel | Block | Polygon |
|---|---|---|---|---|---|---|---|---|
| Forest | 45 | 24.1 | 90 | +1 | +8 | 91 | -1 | -1 |
| Wetland | 34 | 20.1 | 87 | 0 | 0 | 87 | +1 | +8 |
| Shrub | 9 | 2.3 | 71 | +11 | +11 | 77 | 0 | +1 |
| Agriculture | 9 | 5.3 | 77 | +6 | +7 | 81 | +1 | -4 |
| Urban | 3 | 2.5 | 56 | -51 | -51 | 16 | +6 | +21 |
| Overall accuracy | | | 85 | +2 | +4 | | | |

Table 3. Effect of location error on parameter values for user’s and producer’s accuracies. Column “A” is accuracy (expressed as a %) for the spatially aligned reference and map data and column “UN” is the deviation in accuracy of the spatially unaligned reference and map data. Negative deviations in column UN indicate that accuracy is lower in the spatially unaligned population. Classes are listed in order of decreasing percent area based on the map classification.

a) Florida

| | User’s Accuracy | | | | | | Producer’s Accuracy | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pixel | | Block | | Polygon | | Pixel | | Block | | Polygon | |
| Class | A | UN | A | UN | A | UN | A | UN | A | UN | A | UN |
| Forest | 60 | -22 | 55 | -15 | 57 | -10 | 86 | -31 | 84 | -24 | 85 | -16 |
| Agriculture | 99 | -31 | 99 | -25 | 99 | -9 | 98 | -30 | 98 | -24 | 99 | -14 |
| Urban | 100 | -35 | 100 | -25 | 100 | -14 | 95 | -34 | 93 | -24 | 100 | -18 |
| Shrub | 85 | -38 | 82 | -28 | 85 | -27 | 92 | -41 | 91 | -31 | 97 | -20 |
| Wetland | 100 | -20 | 100 | -15 | 100 | -8 | 55 | -12 | 49 | -7 | 49 | -4 |
| Water | 92 | -10 | 92 | -9 | 98 | -3 | 100 | -11 | 100 | -12 | 100 | -2 |
| Barren | 100 | -47 | 100 | 0 | 100 | -58 | 100 | -41 | 100 | -50 | 100 | -11 |
| Overall | 86 | -28 | 83 | -20 | 85 | -12 | | | | | | |

b) Fayetteville

| | User's Accuracy | | | | | | Producer's Accuracy | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pixel | | Block | | Polygon | | Pixel | | Block | | Polygon | |
| Class | A | UN | A | UN | A | UN | A | UN | A | UN | A | UN |
| Agriculture | 83 | -6 | 88 | -5 | 91 | -2 | 86 | -11 | 85 | -5 | 89 | -7 |
| Urban | 85 | -23 | 84 | -14 | 90 | -3 | 72 | 0 | 87 | -5 | 90 | -6 |
| Forest | 79 | -30 | 68 | -11 | 85 | -14 | 69 | -13 | 77 | -15 | 95 | -12 |
| Wetland | 65 | +3 | 85 | -12 | 97 | -8 | 85 | -34 | 65 | -9 | 79 | -8 |
| Shrub | 54 | -21 | 61 | -17 | 74 | -38 | 68 | -42 | 49 | -13 | 83 | -23 |
| Water | 61 | -3 | 91 | -9 | 94 | -20 | 89 | -49 | 77 | -13 | 81 | -17 |
| Overall | 77 | -15 | 78 | -8 | 89 | -9 | | | | | | |

c) Dare

| | User's Accuracy | | | | | | Producer's Accuracy | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pixel | | Block | | Polygon | | Pixel | | Block | | Polygon | |
| Class | A | UN | A | UN | A | UN | A | UN | A | UN | A | UN |
| Forest | 90 | -7 | 91 | -4 | 98 | -1 | 91 | -9 | 90 | -5 | 90 | -3 |
| Wetland | 87 | -7 | 88 | -5 | 87 | -1 | 87 | -7 | 88 | -4 | 95 | -3 |
| Shrub | 71 | -13 | 82 | -12 | 82 | -12 | 77 | -24 | 77 | -11 | 78 | -1 |
| Agriculture | 77 | -1 | 83 | -5 | 84 | +1 | 81 | -9 | 82 | +13 | 77 | +3 |
| Urban | 56 | -45 | 5 | +2 | 5 | -4 | 16 | +23 | 22 | +11 | 37 | -23 |
| Overall | 85 | -8 | 87 | -5 | 89 | -2 | | | | | | |

**Figure Captions**

Figure 1. Florida populations represented as difference maps for the three spatial assessment units, pixel, block, and polygon. Agreement between the map and reference classifications is shown in white, disagreement in black. Overall accuracy is 86% for the pixel unit, 83% for the block unit, and 85% for the polygon unit.

Figure 2. Fayetteville populations represented as difference maps for the three spatial assessment units, pixel, block, and polygon. Agreement between the map and reference classifications is shown in white, disagreement in black. Overall accuracy is 77% for the pixel unit, 78% for the block unit, and 89% for the polygon unit.

Figure 3. Dare populations represented as difference maps for the three spatial assessment units, pixel, block, and polygon. Agreement between the map and reference classifications is shown in white, disagreement in black. Overall accuracy is 85% for the pixel unit, 87% for the block unit, and 89% for the polygon unit.

Figure 4. Effect of location error for a 3x3 pixel block assessment unit. A) Location of a 3x3 pixel block assessment unit delineated on a Landsat image; this is what a photointerpreter would first look at to locate an assessment unit selected by the sampling design. B) Block assessment unit of panel A classified by the map being evaluated; this information would not be provided to the interpreters. C) Sample block unit overlaid on a high resolution image used for determining the reference classification; the block is nearly homogeneous forest. D) Same image as in C but with the block shifted one pixel down and one pixel to the left to create a spatial misalignment of the map and reference locations; the reference classification for the block must now address the mixed nature of the block as forest and pasture.

Figure 5. Systematic sample of points for selecting a sample of polygon assessment units. A polygon would be included in the sample if a systematic sample point falls within the polygon. The probability that a polygon is selected by the point sampling protocol is proportional to the area of the polygon.

Figure 1

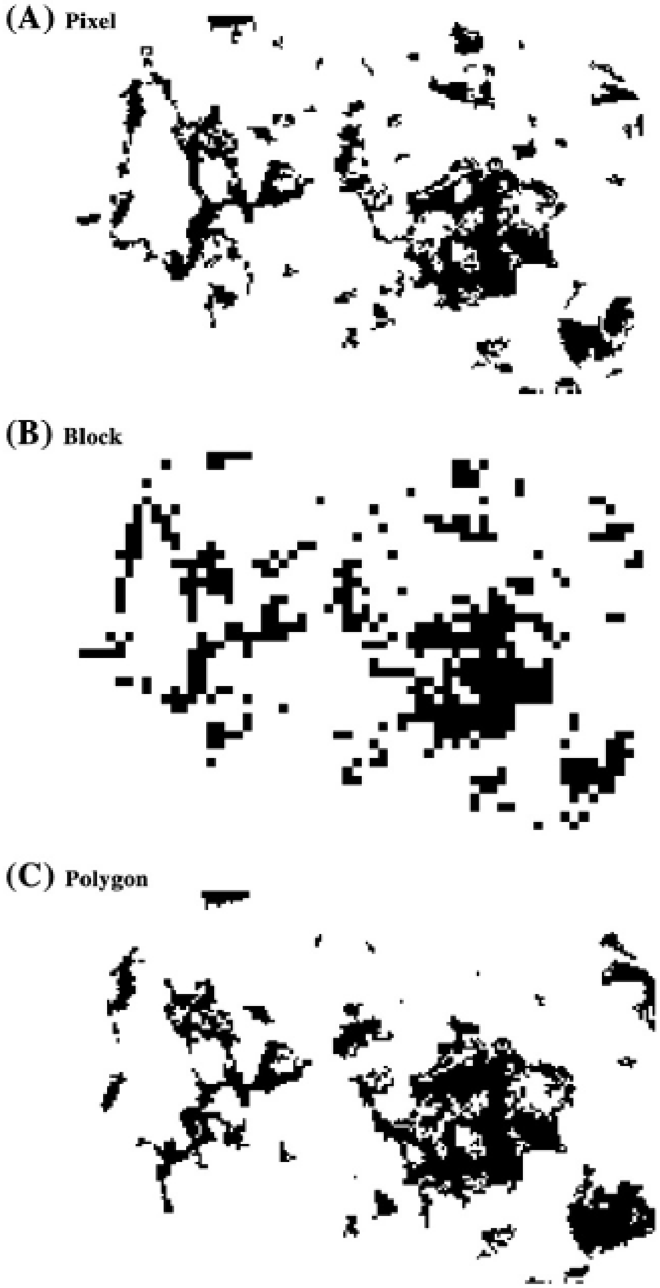


Figure 2

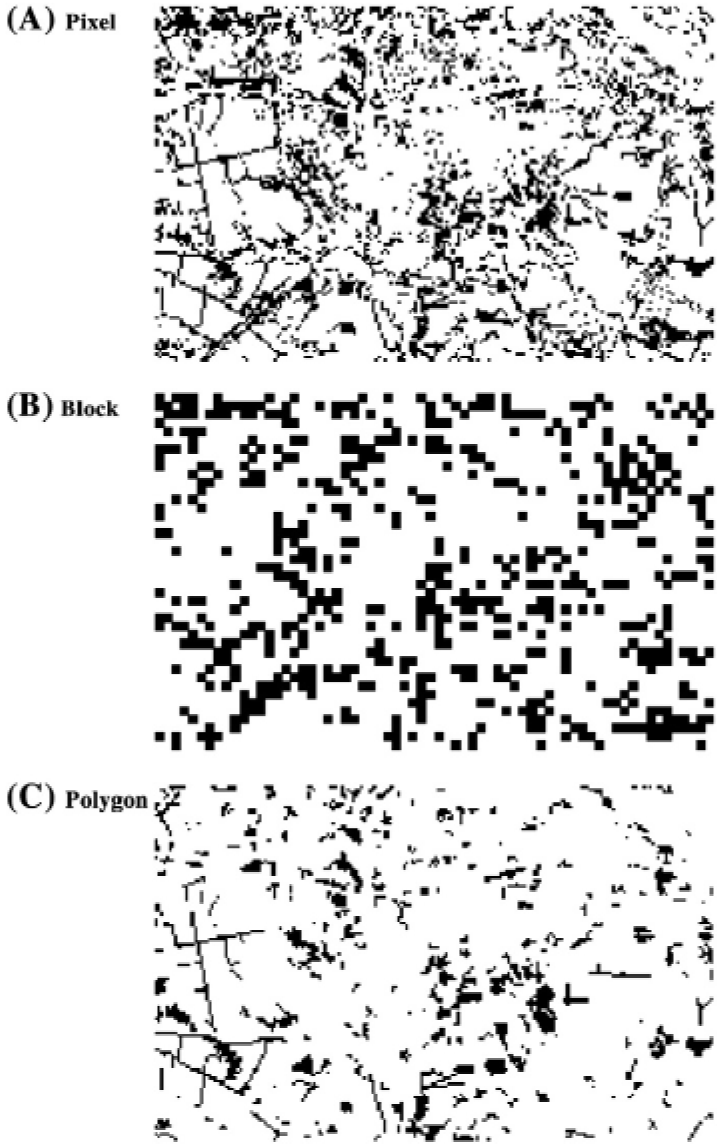


Figure 3

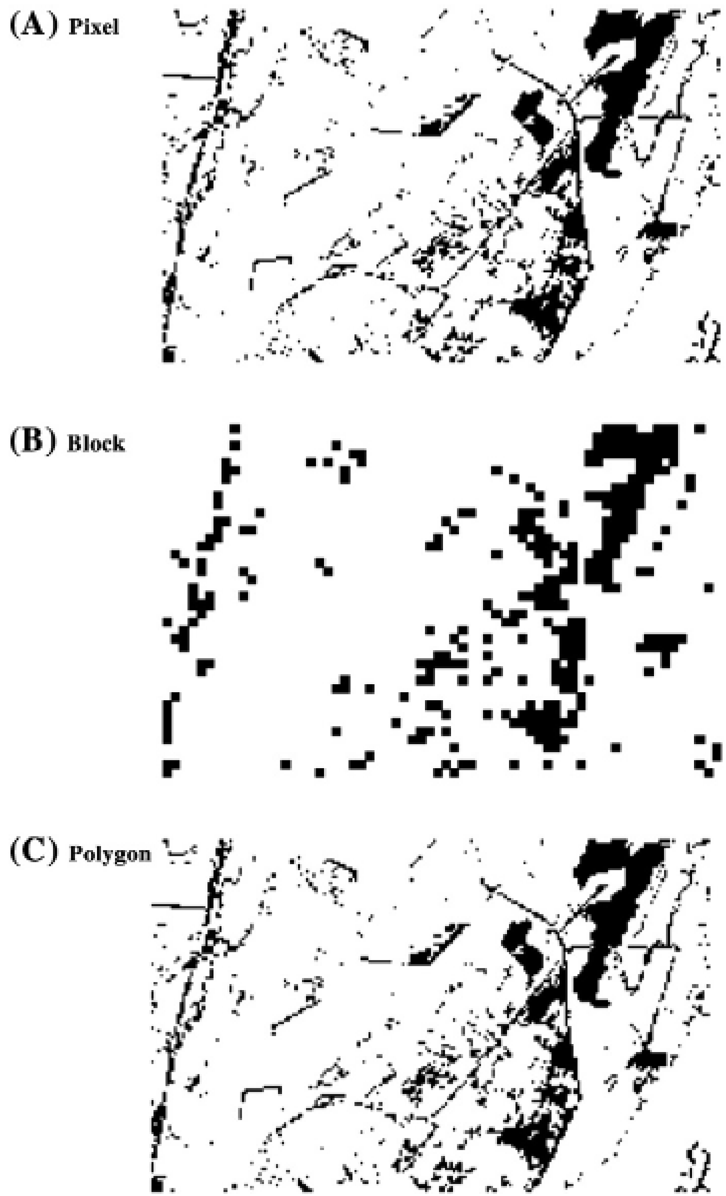


Figure 4

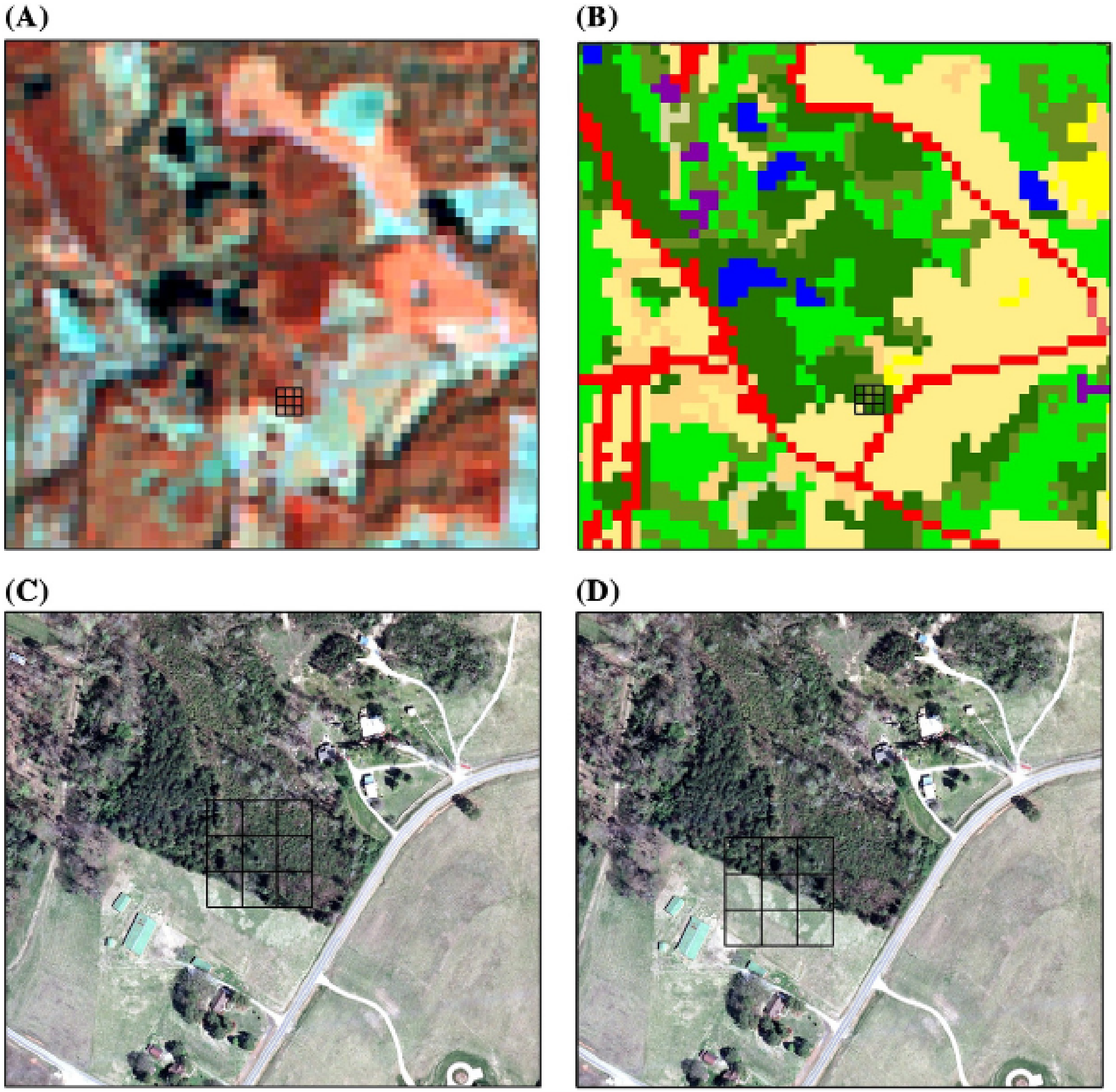
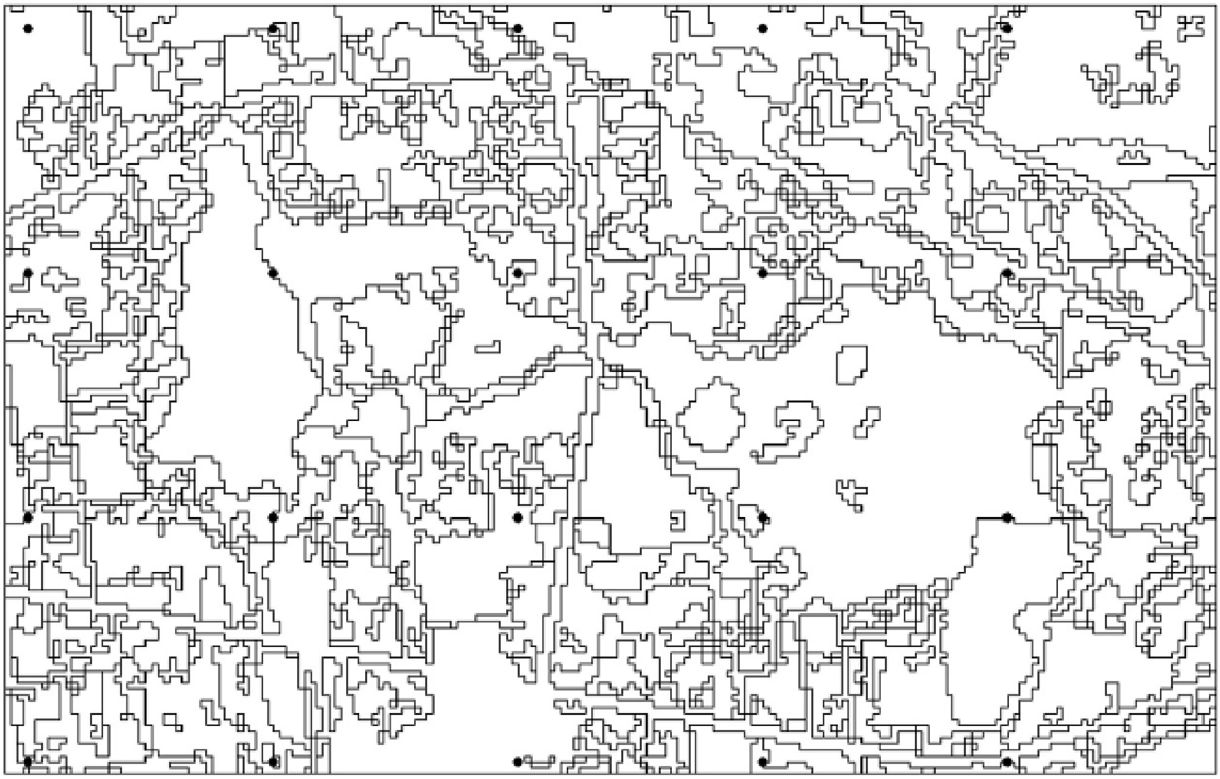


Figure 5



**Supplemental Online Material – Figure Captions**

Figure S1. Florida region target map – National Land Cover Data (NLCD) 1992.

Figure S2. Fayetteville region target map – National Land Cover Data (NLCD) 2001.

Figure S3. Dare region target map – National Land Cover Data (NLCD) 2001.

**Appendix: Effect of spatial assessment unit on accuracy parameter values for additional (low accuracy) populations.**

The populations created by shifting the Florida, Fayetteville, and Dare reference maps (Section 2.3) represent additional example populations for examining how the values of the accuracy parameters change as a function of spatial assessment unit. Although the primary purpose for creating the location-error impacted populations discussed in Section 2.3 was to examine how accuracy results changed for different spatial units when location error was present, these location-error impacted populations provide additional examples for evaluating changes in accuracy parameters resulting from use of different spatial assessment units (related to Section 2.2). Table A1 provides a comparison of accuracy parameters resulting from the different spatial units for a set of location-error affected populations in which accuracy is lower than the populations used in Table 2. The general trends in the results observed from Table A1 are similar to those seen in Table 2 but the magnitude of the differences in accuracy obtained from different spatial units is greater for the lower accuracy populations of Table A1. Specifically, the general trend observed in Table 2 that pixel- and block-based accuracies are more similar to each other than they are to polygon-based accuracy is observed in Table A1. However, the magnitude of the increase in block-based accuracy relative to pixel-based accuracy is much greater in Table A1 than in Table 2. The polygon-based accuracies are similarly much higher than the pixel-based and block-based accuracies for the low accuracy populations of Table A1. For example, overall accuracy for the polygon assessment is 15%, 18%, and 10% higher than the pixel assessment for the Florida, Fayetteville, and Dare populations in Table A1, whereas the corresponding changes in overall accuracy are -1%, +12%, and +4% for the Table 2 examples. In Table A1, block- and polygon-based user's and producer's accuracies are almost always higher than the corresponding pixel-based accuracies, with increases in accuracy of 10-25% not unusual. Generally the increase in accuracy of the block- and polygon-based assessments relative to the pixel-based accuracy is smaller if the land-cover class comprises a relatively large percentage of the area (e.g., forest in Florida and Dare, and agriculture in Fayetteville). The results of Appendix Table A1 suggest that differences in accuracy resulting from different spatial units are magnified when accuracy is lower.

**Table A1**. Parameter values for user's and producer's accuracies resulting from different spatial assessment units for three additional populations. If a block or polygon is not homogeneous in terms of the map or reference classification, the mode class is used and agreement is defined as a match between the mode map class and the mode reference class. Block and polygon accuracies are reported as deviations from the pixel-based accuracy parameter values. Positive deviations indicate higher accuracy for the block- or polygon-based assessment, and negative deviations indicate lower accuracy for the block- or polygon-based assessment. Land-cover classes are listed in order of decreasing percent of map area.

a) Florida (location-error impacted population)

| | Map | User's Accuracy (%) | | | Producer's Accuracy (%) | | |
|---|---|---|---|---|---|---|---|
| Class | Area (%) | Pixel | Block | Polygon | Pixel | Block | Polygon |
| Ag | 20 | 68 | +6 | +22 | 68 | +6 | +17 |
| Wetland | 20 | 80 | +5 | +12 | 43 | -1 | +2 |
| Forest | 19 | 38 | +2 | +9 | 55 | +5 | +14 |
| Urban | 18 | 65 | +20 | +21 | 61 | +8 | +21 |
| Shrub | 16 | 47 | +7 | +11 | 51 | +9 | +26 |
| Water | 7 | 82 | +1 | +13 | 89 | -1 | +9 |
| Barren | <1 | 53 | +47 | -11 | 59 | -9 | +30 |
| Overall | | 58 | +5 | +15 | | | |

b) Fayetteville (location-error impacted population)

| | Map | User's Accuracy (%) | | | Producer's Accuracy (%) | | |
|---|---|---|---|---|---|---|---|
| Class | Area (%) | Pixel | Block | Polygon | Pixel | Block | Polygon |
| Ag | 32 | 77 | +6 | +12 | 75 | +5 | +7 |
| Urban | 31 | 62 | +8 | +25 | 72 | +10 | +12 |
| Forest | 17 | 49 | +8 | +22 | 56 | +6 | +27 |
| Wetland | 10 | 68 | +5 | +21 | 51 | +5 | +20 |
| Shrub | 9 | 33 | +11 | +3 | 26 | +10 | +34 |
| Water | <1 | 58 | +24 | +16 | 40 | +24 | +24 |
| Overall | | 62 | +8 | +18 | | | |

c) Dare (location-error impacted population)

| | Map | User's Accuracy (%) | | | Producer's Accuracy (%) | | |
|---|---|---|---|---|---|---|---|
| Class | Area (%) | Pixel | Block | Polygon | Pixel | Block | Polygon |
| Forest | 45 | 83 | +4 | +14 | 82 | +3 | +5 |
| Wetland | 34 | 80 | +3 | +6 | 80 | +4 | +12 |
| Ag | 9 | 76 | +2 | +9 | 72 | +3 | +8 |
| Shrub | 9 | 58 | +12 | +12 | 53 | +13 | +24 |
| Urban | 3 | 11 | -4 | -10 | 39 | -6 | -25 |
| Overall | | 77 | +5 | +10 | | | |