Factors Affecting Spatial Variation of Classification Uncertainty in an Image Object-based Vegetation Mapping

Qian Yu, Peng Gong, Yong Q. Tian, Ruiliang Pu, and Jun Yang

Abstract

Much effort has been spent on examining the spatial variation of classification accuracy and associated factors on a per-pixel basis. In the past few years, object-based classification has attracted growing interest. This paper examines factors affecting the spatial variation of classification uncertainty in an object-based vegetation mapping. We studied six categories of factors in an object-based classification: general membership, topography, sample object density, spatial composition, sample object reliability, and object features. First, classification uncertainty (classification accuracy on a per-case basis) is derived with a bootstrap method. Then, six categories of factors are quantified by categorical or continuous variables. In this step, the appropriate radius for calculating the spatial composition metrics of sample objects is also discussed. Finally, classification uncertainty is modeled as a function of those factors using a mixed linear model. The significant factors are identified and their parameters are estimated from restricted maximum likelihood fit. The modeling results show that elevation, sample object size, sample object reliability, sample object density, and sample spatial composition significantly influence the object-based classification uncertainty. Many of these factors are closely related to the object-based approach. The result of this study helps in understanding classification errors and suggests further improvement of the classification.

Qian Yu is with the Department of Geosciences, University of Massachusetts-Amherst, 611 N. Pleasant St., Amherst, MA 01003 (qyu@geo.umass.edu).

Peng Gong is with the Department of Environmental Science, Policy and Management, University of California at Berkeley, 137 Mulford Hall, Berkeley, CA 94720.

Yong Q. Tian is with the Department of Environmental, Earth & Ocean Sciences, University of Mass achusetts-Boston, 100 Morrissey Blvd, Boston, MA 02125.

Ruiliang Pu is with the Department of Geography, University of South Florida, 4202 E. Fowler Avenue, NES 107, Tampa, FL 33620.

Jun Yang is with the Department of Landscape Architecture and Horticulture, Temple University-Ambler, 580 Meetinghouse Road, Ambler, PA 19002.

Introduction

Remotely sensed data are used extensively in land-cover mapping through a variety of classification approaches. The technique of image classification groups image units to map units or cartographic regions based on spectral features and/or ancillary features according to certain mapping objectives. Accuracy is calculated to assess the fitness of using remotely sensed data and the quality of a particular classification. A confusion matrix (or error matrix) is currently the core of quantitative measures of accuracy (Congalton and Green, 1999). Many measures of classification accuracy may be derived from a confusion matrix, such as the Kappa coefficient for overall evaluation, and the user's accuracy or producer's accuracy for individual classes considering the error of commission or omission. These statistical measures of accuracy report the percentage of correctly classified pixels as a general quantitative measure of the classification quality over the entire image area.

The classification error is normally neither uniformly nor randomly distributed over a region; instead, the spatial pattern of accuracy is often very distinct (Campbell, 1981; Lunetta *et al.*, 1991). It is often necessary to study accuracy on a per-case basis (per-pixel or per-object), in order to understand the magnitude and variation of accuracy over a region and to effectively use the classification results. Therefore, classification uncertainty is adopted to supplement the general measure of accuracy provided by the confusion matrix and its derivatives (Foody, 2002). Uncertainty is a quantitative measure of doubt and distrust on a per-case basis when a classification decision is made in crisp or hard way, i.e., per-case accuracy or local accuracy (Unwin, 1995). Some aliases of classification uncertainty have been used, such as misclassification probabilities, classification probabilities, reliability, or case-based accuracy. Classification uncertainty can be defined in a number of ways (Gong *et al.*, 1996; Pu and Gong, 2004). Most calculations of classification uncertainty are closely related to specific classifiers and are often derived as a by-product of the classification likelihood. For example, in the supervised maximum likelihood classification, uncertainty is calculated from posterior probability that a classification unit (pixel) actually belongs to a particular class (Foody *et al.*, 1992; Gong *et al.*, 1996; Canters, 1997). A measure of uncertainty in a minimum distance

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 8, August 2008, pp. 1007–1018.

0099-1112/08/7408–0000/$3.00/0

classification can be conveniently derived as the inverse distance between a classification unit to the cluster mean. In addition, a few other measures of uncertainty are based on classification variability over multiple simulated test datasets. These datasets are usually simulated from nonparametric statistical procedures, which are independent of the classifier (Steele *et al.*, 1998). Much of the recent effort in studying the spatial distribution of classification quality has been focused on the visualization of classification uncertainty (Fisher, 1994; Van der Wel *et al.*, 1998; Foody, 2002). Visualization of classification uncertainty provides not only an understanding of the distribution of classification quality and help in identifying problematic areas, but it also reveals the factors driving this variation and suggests possible improvements.

Some researchers have reported that terrain, sample size, sample distribution, landscape complexity, or even reliability of sample data are important factors in classification accu-racy and its spatial variation in a pixel-based classification (Congalton, 1988a; Dicks and Lo, 1990; Gong and Howarth, 1990; Stehman, 1992; Friedl et al., 2000; Foody, 2002). For example, through the comparison of forest, agriculture and range, Congalton (1988a and 1988b) concluded that the patterns of classification error are influenced by landscape features such as sampling distribution, topography, and spatial complexity. Steele et al. (1998) used two sites to illustrate the strong spatial correspondence between the patterns of misclassification probabilities and terrain. In a review paper, Foody (2002) pointed out that spatial distribution of samples has important implications on classification accuracy. The statistical requirements of sampling have to be balanced against realistic constraints due to physical accessibility, budget, or other practical restrictions. Relatively less attention has been paid to the effect of unreliable training samples in classification accuracy. Unreliable training data can result from the subjective nature of field and photo verification, image registration and so on (Lunetta *et al.*, 1991). Under some circumstances, the problem of sample reliability is very implicit and therefore has often been ignored, such as the impure mapping unit or unconfident interpretation, especially for automatic sampling over a large dataset in an object-based classification context.

Rather than visually comparing pattern similarity or reporting the summarized accuracy affected by the above factors, several recent researchers have used statistical analysis to study the relationship between classification error and potential explanatory variables related to landscape and sample distribution (Moisen *et al.*, 2000; Smith *et al.*, 2002; Smith *et al.*, 2003; Van Oort *et al.*, 2004). Most of them used a dichotomous response variable to indicate whether or not each sample pixel was correctly classified, and developed a logistic regression to model this binary response. They assessed the effects of various characteristics of sample pixels on classification accuracy, including land-cover class, focal neighborhood heterogeneity, patch size, and landscape indices. They come to the conclusion that "per-pixel classification accuracy was significantly higher for pixels with more heterogeneous focal neighborhoods, pixels located in larger patches, pixels located in regions with a less heterogeneous landscape and pixels located in regions with a more coarsely textured landscape" (Van Oort *et al.*, 2004).

Objected-based classification has been adopted more and more to extract information from high-resolution imagery. In contrast to conventional approaches, the processing unit in object-based classification is the image object rather than the pixel. The classification decision is based on features describing the characteristics of each image object. The "sample" data used in classification has to be a subset of image objects containing ground truth information, and we call it "sample object." This change of processing unit has new impacts on classification accuracy. First, the generation of sample objects may introduce clustered and unreliable samples. In pixel-based vegetation classification, it is well-accepted to use pixels overlapping with the field survey plot as sample pixels. While in object-based classification, a field survey plot rarely perfectly overlaps with image object(s). It is often partially superimposed on more than one image object. Taking image objects partly cov-ered by a field survey plot as sample objects, object-based classification unavoidably uses some sample objects of the same class in clusters. In addition, some sample objects included are not reliable since they are only partially within the survey plot. Second, if we define "sample" as the training or test data in the unit directly handled by an classifier (i.e., object), sample size and sample distribution are not only determined by field sampling strategy, but also by how the ground truth is projected onto image objects in order to select sample objects. This process is relatively subjective. Third, in contrast to pixel-based classification, the dimension of image objects varies. The characteristics of the image objects might play a role in the classification accuracy to a certain degree. Therefore, it is necessary and meaningful to examine the uncertainty of the object-based classification in association with the properties of sample objects.

This paper evaluates the impact of these potential factors on object-based classification uncertainty. It is well recog-nized that classifiers, accuracy assessment methods, and data preprocessing (e.g., atmospheric correction) affect classification accuracy (Congalton, 1988; Gong and Howarth, 1990; Lunetta *et al.*, 1991). However, these factors are not specific to individual classification units (pixels or objects) and do not vary over spatial location. In other words, they are treated as "global factors" influencing the global or general accuracy of a classification. These global factors are unlikely to lead to the spatial pattern of classification uncertainty or the variation of per-case accuracy for a classification procedure. Instead of analyzing the effects of global factors, the objective of this study is to investigate the potential explanatory variables affecting spatial variation of classification uncertainty in an object-based classification. Along the lines of previous work by Smith *et al.* (2002 and 2003) and van Oort *et al.* (2004), in this study we consider the factors related to sample distribution and landscape complexity. Our objective is distinct from theirs because of its object-based context. In addition, for several reasons, we include three categories of explanatory variables: terrain, sample reliability, and object features. First, our classification has a focus on detailed vegetation classes over a relatively large area using high-resolution imagery. Topography as ancillary features played an important role in the classification. Second, because of the aforementioned process of sample object generation, the affects of unreliable samples become more significant in object-based classification than in pixel-based classification. Third, due to the irregular dimension of image objects, the category of object features is included to better characterize image objects. Size, brightness, shape, and orientation of image objects in this category are tested as potential factors in classification uncertainty.

This study examines the factors describing each sample object with general membership, landscape characteristics (topography, object density, spatial composition), sample object reliability, and object features. Our hypothesis is that these factors have significant influence on classification uncertainty. We tested the hypothesis based on an object-based vegetation mapping with airborne high-resolution images. A mixed linear model was used to examine the effect of spatially related factors on classification uncertainty. Classification uncertainty is first derived from the bootstrap method. The possible factors affecting spatial variation are quantified by either continuous or categorical variables. Then, classification uncertainty is fitted by those explanatory variables with the mixed linear model. Based on the model, we identify significant factors and discuss the relationships between them.

Methods

Data and Earlier Results

In this study, object-based vegetation classification is based on airborne images at 1 m resolution in Point Reyes National Seashore, California (Yu *et al.*, 2006). The study area is about 72,800 ha (180,000 acres) and covered by 26 frames of the Digital Airborne Imagery System (DAIS) images. The sensor system has four bands: Blue (450 to 530 nm), Green (520 to 610 nm), Red (640 to 720 nm), and Near-infrared (770 to 880 nm). Tonal balance was conducted among all the images. No further atmospheric correction was performed in this study, considering the clear and dry atmospheric conditions when the image data were collected and a relatively low flight altitude of 2,500 m above ground. First, pixels were grouped into image objects through the Fractal Net Evolution Approach (FNEA) segmentation using eCognition® 4.0 from Definiens AG (Baatz *et al.*, 2004). Then, features were generated for each object. After conducting feature selection, a non-parametric classifier, nearest-neighbor was employed to classify the objects into 48 classes. The classification scheme was designed at the alliance level according to the vegetation classification system of the California National Plant Society (CNPS) (The California Natural Diversity Database, 2003). The alliance is a physiognomically uniform group of plant associations sharing one or more dominant species. Among 48 classes, 43 are vegetation alliances, which could be categorized into forest, shrub, and herb.

Our sample data were acquired mainly from the 1,329 field survey plots (0.5 ha each) provided by the National Park Service, supplemented with our field reconnaissance. Those field survey plots were superimposed on the image objects. Any image objects overlapping with the field survey plots by more than 10 percent of its own area were taken as a sample object (Baatz *et al.*, 2004; Yu *et al.*, 2006). As a result, the locations of the 6,824 sample objects are presented in Figure 1.

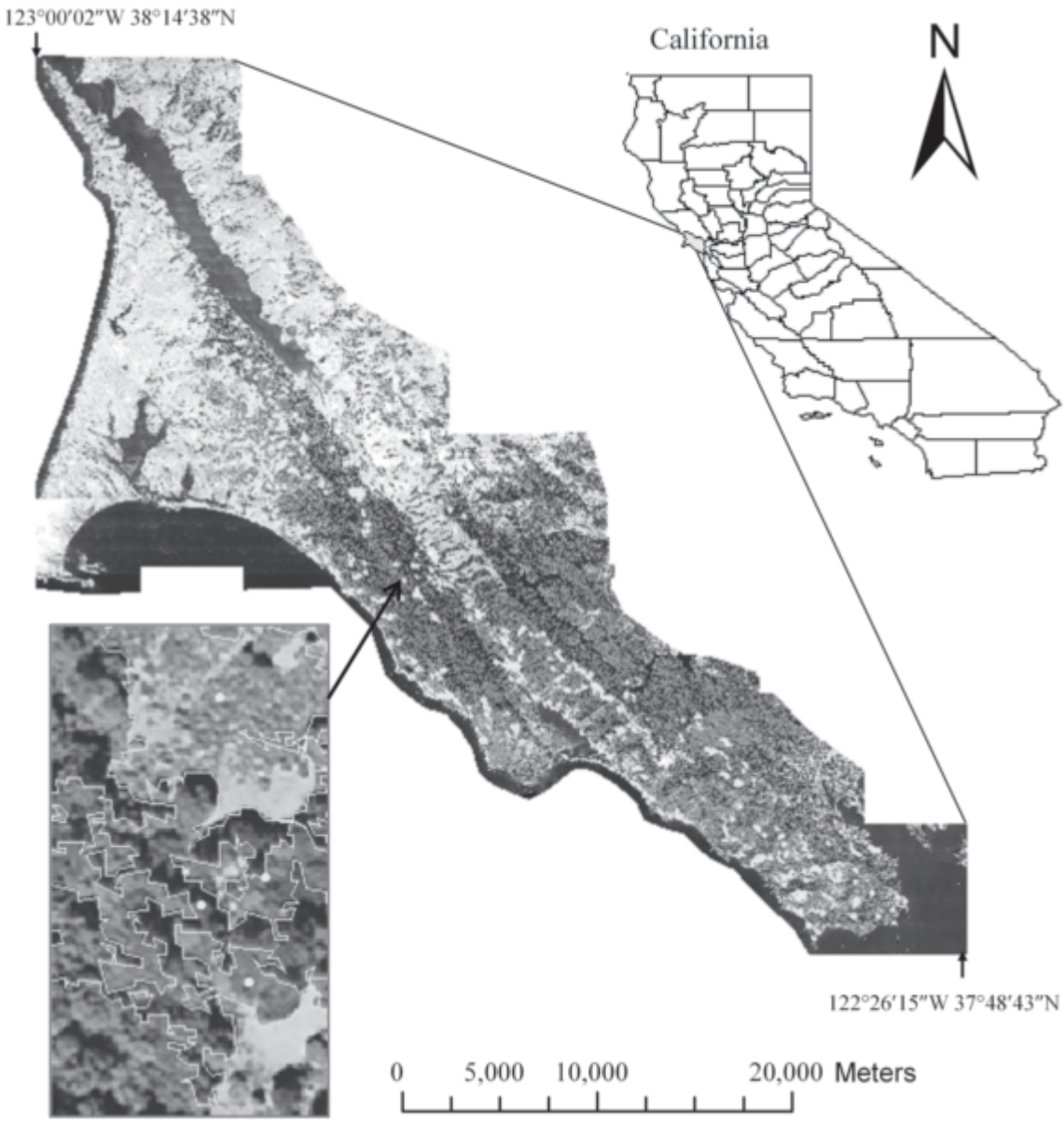

Figure 1. Study site, sample objects in highlighted (yellow) polygons and their centroids indicated by dots. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

The reason for choosing 10 percent as overlapping criterion will be discussed in a following section. The object-based classification achieved an overall accuracy of 58 percent over the entire study area. The classification was reported and results were discussed by Yu *et al.* (2006). Due to the complexity and high heterogeneity of the vegetation types, the global statistical accuracy might not be optimal to evaluate the classification over the entire study area. We want to further study how the accuracy varies spatially. This understanding would improve our knowledge in three areas: (a) the map quality derived from the classification, (b) identification of the potential factors to improve the classification according to error spatial distribution, and (c) accuracy variations caused by sample objects generation and object dimension associated with an object-based classification.

Estimation of Classification Uncertainty with Bootstrap Methods

The bootstrap is a general nonparametric procedure for assessing statistical accuracy and quantifying expected prediction error or uncertainty (Efron and Tibshirani, 1993; Hastie *et al.*, 2001). The advantage of bootstrap estimation is the independence of classifier and its broad applicability. The basic idea of bootstrap is to randomly draw datasets from the sample data with replacement by a Monte Carlo simulation (Hilborn and Mangel, 1997). Each bootstrap dataset has the same size as the original sample data. This is done B times, producing B bootstrap datasets. The classification is then conducted on each of the bootstrap datasets, and the accuracy is assessed by examining the behavior of the classification over the B replications. Since Bootstrap methods involve generating "replicate" sample datasets from the original samples, they improve confidence in parameter estimates and uncertainty calculation with better data quality and increased sample size. Each hypothetical sample dataset, called a Bootstrap sample (set), represents a potential set of observations that could be obtained when resampling from the population.

It is necessary to modify the bootstrap algorithm when it is applied to nearest neighbor classifiers (Efron and Tibshirani, 1997; Steele *et al.*, 1998). In nearest neighbor classification, the observation to be classified is assigned to the membership of the closest sample in feature space. If the observation is in the bootstrap sample, the nearest observation is itself and the observation is certain to be correctly classified. Consequently, the classification error will be underestimated. To avoid this problem, leave-one-out bootstrap is adopted to classify only the observations that are not in the bootstrap sample (Steele *et al.*, 1998). Figure 2 illustrates how to derive case-based accuracy with the Bootstrap method. In the case of 1-nearest neighbor classification, the original sample set is denoted as $Z = (z_1, z_2, \ldots, z_N)$ with N observations. B bootstrap sample set Z^{*b} ($b = 1, 2, \ldots, B$) are drawn from the original sample set. Each bootstrap sample set Z^{*b} also has N observations from N times draw with replacement, certainly some of which are duplicates. Taking a bootstrap sample set as training samples, observations that do not appear in the bootstrap sample set b are classified. The duplicated observations are treated as one observation. $C(Z^{*b})$ represents the classification result with Z^{*b} as training samples. This process is repeated with the next bootstrap sample until every observation has been classified at least some predefined number of times. It was set 500 times in this study. For each observation (sample object), classification uncertainty is defined as the ratio of the number of times being correctly classified to the total number of times being classified. Some studies calculated incorrect classification ratio, which is complementary and comparable to this definition (Steele *et al.*, 1998). It is worth mentioning that this mean estimate of the accuracy tends to be slightly optimistic (Hastie *et al.*, 2001). From this estimation, a larger classification uncertainty value actually indicates two facts: first, the observation has high probability of being correctly classified; and second, the classification result is *certain* and less dependent on a specific sample set since the uncertainty is the average accuracy from different sample sets. This classification uncertainty is also called per-case accuracy because of this implication.

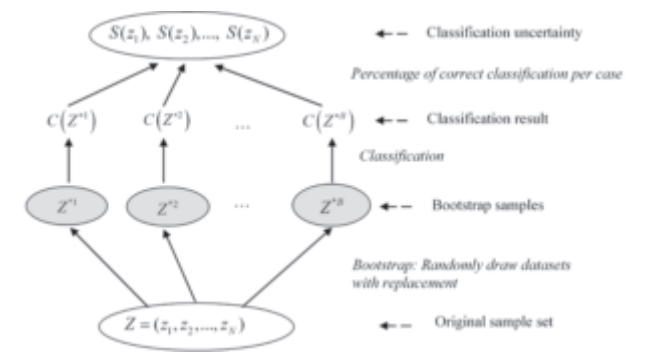

Figure 2. Derivation of classification uncertainty with Bootstrap algorithm.

Quantifying Factors Affecting the Classification Quality

Six categories of factors are quantified as explanatory variables to model classification uncertainty: topographic variables, object membership, sample object density, spatial composition, sample object reliability, and object features. These variables are discussed in detail in this section.

Topographic Variables

Five topographic variables are derived using 10 m DEM from the USGS Bay Area Regional Database, including elevation (ELEV) in meters, slope (SLOPE) and aspect (ASPECT). To better characterize solar radiation, two radiation indices, SI (Solar Index) and TRASP (Transformed Aspect) are transformed from slope and/or aspect using Equations 1 and 2 (Lewis and Hutchinson, 2000; Moisen *et al.*, 2000):

$$SI = \cos(aspect) * \tan(slope) \quad (1)$$

$$TRASP = \frac{1 - \cos((\pi/180)(aspect - 30))}{2} \quad (2)$$

where *aspect* is in degrees from north, and slope is in degrees. TRASP has a range of 0 to 1. It is assigned zero to land oriented in a north-northeast direction (30 degrees east of north), the coolest and wettest orientation and one to the hotter, drier south-southwest slope.

Object Membership

The forty-eight classes in the classification belong to four general groups: forest, shrub, herb, and non-vegetation, according to a vegetation key created specifically for the study site (Keeler-Wolf, 1999). The major confusion or misclassification occurred between classes belonging to the same general group. To compare the classification strength for each group due to their spectral separability, we adopted the membership of the four general groups (GROUP) as an explanatory factor. We assume the classification accuracies of classes in the same general group are more closely correlated than the classes between groups. GROUP is a categorical variable with four levels, which represents the four distinct groups.

Sample Object Density

Sample object density is included in the analysis to test the effect of two factors: sampling scheme and sample object generation. First, although a large number of survey plots were collected in this study, sampling was constrained by accessibility. Most field survey plots are located close to a road. A non-random sampling scheme is likely to introduce bias to the accuracy assessment (Foody, 2002). The accuracy assessment may be less credible for regions with sparse samples. Therefore, the spatial distribution of sample objects may lead to spatial variation of classification uncertainty. Second, the method of sample object generation also determines the sample object density and thus affects uncertainty. Polygons from image segmentation, i.e., image objects, are the minimum classification unit in object-based classification. The most straightforward method of generating sample objects is to overlay survey plots onto image objects and calculate the ratio of the area covered by a survey plot to the entire area of the corresponding image object. An image object will be selected as a sample object if this ratio is greater than a predefined minimum percentage, as illustrated in Figure 3. This widely accepted method has been adopted by the commercial software eCognition® (Baatz *et al.*, 2004). The lower this percentage, the more objects are taken as sample objects. Therefore, the spatial pattern of sample density is determined by both the field sampling scheme and the criteria used to define sample objects.

We choose 10 percent as the minimum overlapping criterion in order to include more sample objects. In this study, the size of image objects is equivalent to the size of the survey plot (the average area of image objects is 4,300 m^2 and that of the survey plots is 5,000 m^2) and the dimension is irregular. Consequently, the portion of an image object covered by a field survey plot is often a small fraction of the entire image object. The number of sample objects we can acquire is quickly reduced by increasing the "overlapping" criterion. Meanwhile, more and more survey plots are not covered by any image object with the fraction greater than the increased overlapping criterion. Consequently, fewer survey plots are taken into account for sample object generation. Figure 4 illustrates that the number of sample objects generated and the number of survey plots used decrease rapidly with increasing the overlapping criterion. Therefore, a relatively small ratio is adopted to guarantee more field survey plots are considered as ground truth. On the other hand, using a lower ratio will include some unreliable sample objects. We addressed this concern by adding Sample Reliability (SamRel) as an explanatory factor. The effect of sample reliability on classification accuracy is one of our interests to study.

The sample object density (SamDen) is estimated using a kernel function illustrated in Figure 5 (Bailey and Gatrell, 1995). First, the centroid of each sample object is extracted as point features. Then, the quadratic kernel function is used to estimate the density of the point pattern (Silverman, 1998). Conceptually, the bivariate probability density function is symmetric curved surface about each point, determining the weight of the neighbor points as a function of distance to the point. The value of the weight is the highest at the location of the point, and diminishes with increasing distance from the point, reaching 0 at the bandwidth. Bandwidth is the radius of the range within which a point's neighbor points significantly contribute to the density of the point, which is designated according to the desired amount of smoothing. The density at each spatial location is calculated by adding the values of all the kernel surfaces superimposed on this location. In Figure 5, the grey level represents the estimated sample object density. Larger bandwidth includes more neighbor samples in the density estimation and the density changes in a slower and smoother way.

Figure 3. Sample object generation. The three grey-filled polygons are field survey plots and the empty polygons are sample objects, which satisfy the overlapping criterion. The round dots indicate the centroids of sample objects and the triangle symbols indicate the centroids of the field survey plots.

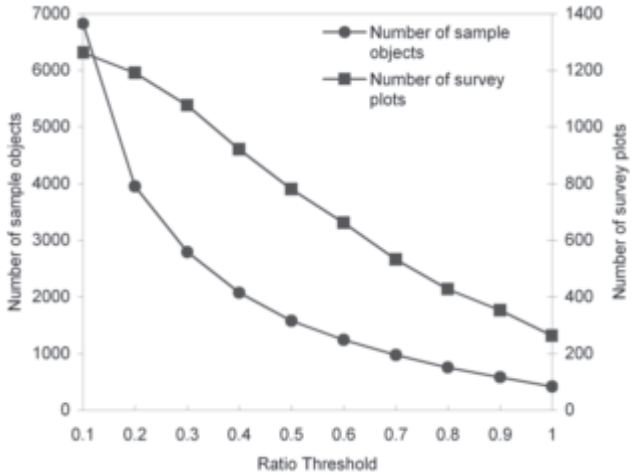

Figure 4. The relationship between the overlapping criterion (the proportion of an image object overlapping with the field survey plot to the whole area) and the number of sample objects generated and survey plots taken into account.

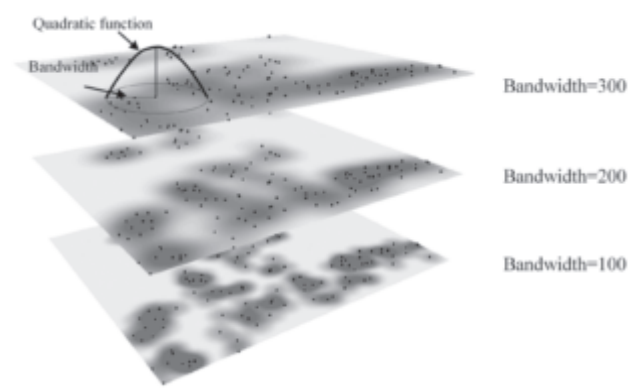

Figure 5. Calculation of sampling object density by Kernel estimation. The black dots indicate the centroids of the sample objects; the gray scale on image indicates the value of the sample object density. The estimated sample object density looks smoother with larger bandwidth.

Spatial Composition of Sample Objects

Spatial composition describes the variety of the samples and their spatial pattern/layout at a certain scale (Turner *et al.*, 2001). Spatial composition indices in this context refer to the quantitative measures of the occurrence of sample object classes in a given neighborhood. Those indices have been used to quantify the spatial heterogeneity of the focal neighborhood (3×3 window) around each pixel in studying spatial variability of pixel-based classification accuracy (Moisen *et al.*, 2000; Smith *et al.*, 2003; Van Oort *et al.*, 2004). We adopted four spatial composition indices: richness, two diversity indices, and evenness. In the calculation of these indices, we ignore dimension and use the centroids to represent the sample objects (Dimension of image objects is considered individually in another category of explanatory variables, "object features"). The four metrics are calculated for circular areas centered at each centroid. Richness (RICH) is defined as the number of sample classes present within a given radius. Two diversity indices (D_1 and D_2) are calculated as

$$D_1 = \exp\left(-\sum_{i=1}^{S}\left[\left(\frac{n_i}{n}\right)\ln\left(\frac{n_i}{n}\right)\right]\right) \tag{3}$$

$$D_2 = \left[\sum_{i=1}^{S}\frac{n_i\,(n_i-1)}{n(n-1)}\right]^{-1} \tag{4}$$

where S is richness, n is the total number of sample objects within a given radius, and n_i is the number of sample objects belonging to the ith sample class. Then n is the sum of n_i. Figure 6 illustrates how to count S, n_i, and n at the location of a sample object.

Actually, D_1 is the exponential of Shannon's entropy, and D_2 is the reciprocal of Simpson's index (Hill, 1973). The higher the value of D_1 and D_2, the greater the diversity. Both D_1 and D_2 start with 1 as the lowest possible value, representing only one class within the radius; their maximum

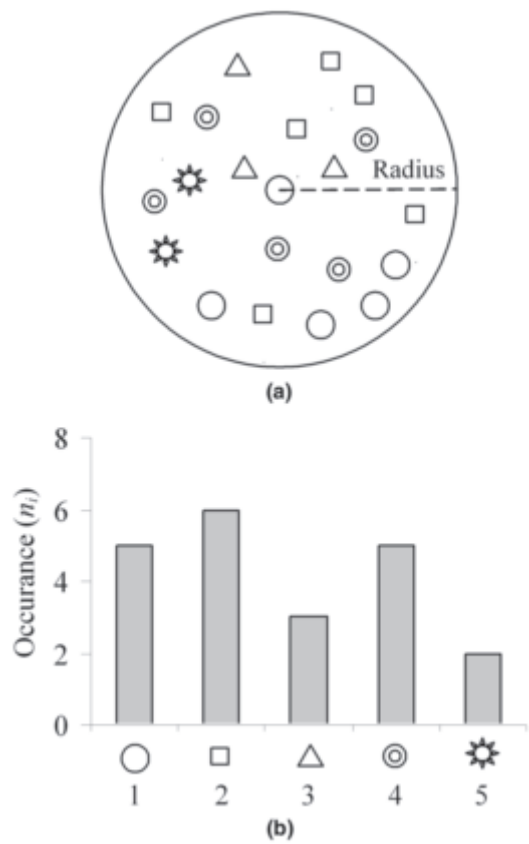

Figure 6. Illustration of calculating spatial composition metrics. Symbols distinguish different classes (five classes): (a) Sample objects within a certain radius, (b) Occurrence frequency of sample classes. S is the number of sample classes present within certain radius; n is the total number of sample objects within a certain radius, and n_i is the number of sample objects belonging to the ith sample class. In this case, $S = 5$, $n_1 = 5$, $n_2 = 6$, $n_3 = 3$, $n_4 = 5$, $n_5 = 2$, and $n = \Sigma n_i = 21$.

values are the number of classes. The two diversity indices are then used to calculate evenness (EVEN), referring to how evenly the proportions of vegetation classes are distributed:

$$EVEN = (D_2 - 1)/(D_1 - 1) \quad (5)$$

According to Equation 4, if there is only one sample object in the radius ($n = 1$), the denominator of D_2 will be zero. In this case, D_2 is assigned the value of D_1. According to Equations 3 and 5, if there is only one sample type within this radius ($s = 1$ and $n_i = n$), D_1 is equal to 1. Therefore, the denominator of EVEN will be zero. In this case, the value of zero is assigned to EVEN since the sampling is very uneven among classes.

Sample Object Reliability

Unreliable sample objects sometimes occur in the process of preparing the sample objects before an object-based classification. Although some unreliable samples could result from incorrect interpretation and geo-referencing as in pixel-based mapping, here we focus only on unreliable sample objects created due to the batch process of sample object generation. Two steps in this vegetation mapping may influence the reliability of sample objects: survey plot delineation and sample object generation. First, the survey plots are re-constructed from the GPS-measured center coordinates of the field survey sites. The size of field survey plots is approximately 0.5 hectare, without fixed dimension or orientation. A circular area with radius of 40 meters around the measured center coordinates was created, and those circular polygons were delineated as survey plots. This step constituted an approximation to the actual field measurement. Second, sample objects are obtained through batch processing of overlaying survey plots and image objects. Therefore, some sample objects are not fully covered or dominated by a single alliance or by the same alliance indicated from the survey plot, although they satisfy the minimum overlapping criterion.

To determine the effect of sample reliability on classification accuracy, we re-examined each sample object. A step of image interpretation was conducted to verify the sample objects one by one, and to mark the mis-assigned sample objects. The incorrect class labels include forest objects assigned as shrub or vice versa near the forest boundaries, roads assigned to a vegetation class when they pass by a vegetation survey plot, mowed ground assigned as herb, and so on. Figure 7 shows four examples of unreliable sample objects. A dichotomous variable, sample object reliability (SamRel) is defined to indicate the correctness of sample object membership. The value of zero represents questionable sample objects; the value of one represents reliable sample objects.

Object Features

In addition to the above indices characterizing the sample allocation, four indices that depict the image objects are derived: brightness, area, shape index, and main direction. This category of indices quantifies the spectral value, size, shape, and orientation of image objects, respectively (Baatz *et al.*, 2004). Brightness is the average digital number over five bands (Blue, Green, Red, NIR, and intensity) of the spectral mean values of pixels composing an image object. The shape index (SHAPE) describes the smoothness of the image object borders. Mathematically it is the border length of the image object (e) divided by four times the square root of its area (A), i.e., $\frac{e}{4\sqrt{A}}$. The denominator is equivalent to the perimeter of the square with the same area. The more fractal an image object appears, the higher its shape index. The main direction of an image object is the direction of the eigenvector belonging to the larger of the two eigenvalues derived from the covariance matrix of the spatial distribution of the image object (Baatz *et al.*, 2004).

Mixed Linear Model

Due to the constraints of field sampling and sample object generation discussed in the Introduction, the correlation between variables and the autocorrelation of individual variables are inevitable. A simple linear model cannot reveal the underlying process properly if the variances and covariances are in reference to spatial location and/or other factors. Therefore, a mixed linear model was applied to study the effect of the above factors on classification uncertainty. A mixed linear model is a generalization of the standard linear model, which allows the data to exhibit correlation and non-constant variability (Littell *et al.*, 2002). In contrast to simple linear models, a mixed linear model contains both fixed- and random-effects parameters and uses covariance structure for random effects. A mixed linear model can be written as:

$$Y = X\beta + Z\gamma + \varepsilon \quad (6)$$

where Y is the vector of observed dependent variable values, X is the known matrix of predictor-variable values, and β is

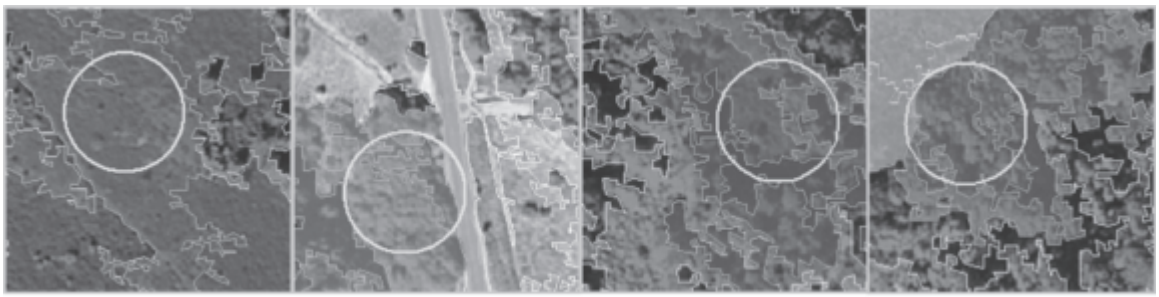

Figure 7. Examples of sample object reliability. Circles (yellow) represent field survey plots and green shaded polygons denote sample objects: (a) A reliable sample object of Coyote Brush; (b) through (d) Three unreliable sample objects: (b) an image object of road mistakenly assigned as willow, (c) an image object of California Bay mistakenly assigned as Coyote Brush, and (d) an image object of perennial grassland mistakenly assigned as California Bay. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

the fixed-effects parameter vector. The mixed model generalizes the simple linear model by adding the known design matrix Z and the unknown vector of random-effects parameters γ. Instead of assuming ε as an unknown Gaussian random error vector with zero-mean and constant variance, the residual error ε is an unknown random error vector whose elements are no longer required to be independent and homogeneous. A key assumption in the mixed model is that γ and ε are normally distributed with:

$$E\begin{bmatrix}\gamma\\ \varepsilon\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

$$\mathrm{Var}\begin{bmatrix}\gamma\\ \varepsilon\end{bmatrix} = \begin{bmatrix}G & 0\\ 0 & R\end{bmatrix}$$

where G and R are the variance-covariance matrices of γ and ε, respectively. The variance of Y can be modeled by specifying the structure of G and R.

The corrected Akaike Information Criterion (AIC_c) is adopted as our criterion to select the appropriate covariance model and to measure the model fitness. The definition of AIC is based on the modeling perspective that the underlying true model is essentially infinite dimensional and the goal of modeling is to find the best approximating finite dimensional model (Hoeting *et al.*, 2006). Therefore, based on the Kullback-Leibler information, AIC was developed as a measure of the loss of information incurred by fitting an incorrect model to the data:

$$\mathrm{AIC} = -2\log L + 2(p + k + 1)$$

where L is the model likelihood, p is the number of fixed effect terms, and k is the number of random effect terms. The first term is a measure of the quality of fit of the model, while the second term is a penalty factor for the introduction of additional parameters into the model. AIC may perform poorly if there are too many parameters in relation to the size of the sample. Further small-sample (second-order) bias adjustment led to the AIC_c statistic, in which the penalty term is multiplied by the additional bias correction factor $n/(n - p - k - 2)$, i.e.,

$$\mathrm{AIC_c} = -2\log L + [2(p + k + 1)(n/(n - p - k - 2))]$$

where n is sample size (Burnham and Anderson, 2002; Hoeting *et al.*, 2006). AIC_c examines the complexity of the model together with goodness of fit to the sample data, and produces a measure which balances between the two (Littell *et al.*, 2002). Since AIC_c is on a relative scale, the model with the smallest AIC_c values is the best approximation for the information in the data.

The model fitting is conducted with the MIXED procedure in SAS (SAS Institute Inc., 2002). The estimation method is residual maximum likelihood (REML). The mixed linear model includes both fixed-effects and random-effects, and accounts for the different variance for the random effect. Variables of fixed-effects include five topographic variables, sample object density, four spatial composition indices, sample object reliability and four object features. Sample membership GROUP was set as a random effect since we assume the classification uncertainty within a general group exhibits more correlation than across general groups.

Using the mixed linear model, classification uncertainty is modeled as a function of the six categories of variables (Table 1). Sample object density and spatial composition indices measure the spatial heterogeneity of sample objects at a location; their values are estimated based on the occurrence of sample objects within a predefined radius. Consequently, this raises the question of radius selection. By varying the radius from 100 m to 1,000 m with an interval of 100 m, the sample object density (SamDen), richness (RICH), diversity index 1 (D_1), diversity index 2 (D_2), and evenness (EVEN) are calculated for each sample object. For convenience, we will simply call this group of variables spatial metrics. The mixed model is fitted respectively for the spatial metrics based on the ten different radii.

Results and Discussion

Model

Several covariance structures of G and R are tested including variance components, spherical spatial and exponential spatial. Based on the rule that a smaller AIC_c indicates a more preferable model, the structure of variance components is adopted for both G and R, which models a different variance component for each random effect (GROUP). Figure 8 plots the AIC_c for two cases based on a series of radii: one model defining GROUP as a random effect, and the other only including fixed effects. Declaring GROUP as a random effect sets up a common correlation among all observations belonging to the same general group: forest, shrub, herb, and non-vegetation. The difference of AIC_c between the two models ranges from 86 to 111 over all radii, which is shown as the interval between the two curves. It demonstrates that the random effect included in the model significantly lowers the AIC_c and improves the correlation modeling in the data.

Table 2a lists the parameter estimates for random effect. From the p-value, the effect of the non-vegetation group is significant, but the effects of the three vegetation groups are not. Therefore this result confirms that it is necessary to model an additional correlation between all samples in non-vegetation level to the rest of the samples in vegetation level. In other words, the classification uncertainty between the samples within a general group (vegetation or non-vegetation) exhibits more correlation than those across general groups, but the correlation of samples within each vegetation group (forest, shrub, and herb) is not significantly different from samples across other vegetation groups. We could use one covariance component for all vegetation

TABLE 1. EXPLANATORY VARIABLES

Category	Variables	Type	Radius
Object membership	Group	Categorical	N/A
Topography	Elevation, Slope, Aspect, SI, TRASP	Continuous	N/A
Sample object density	SamDen	Continuous	100, 200,.., 1000
Spatial composition	RICH, D_1, D_2, EVEN	Continuous	100, 200,.., 1000
Sample object reliability	SamRel	Dichotomous	N/A
Object features	Brightness, Area, Shape Index, Main Direction	Continuous	N/A

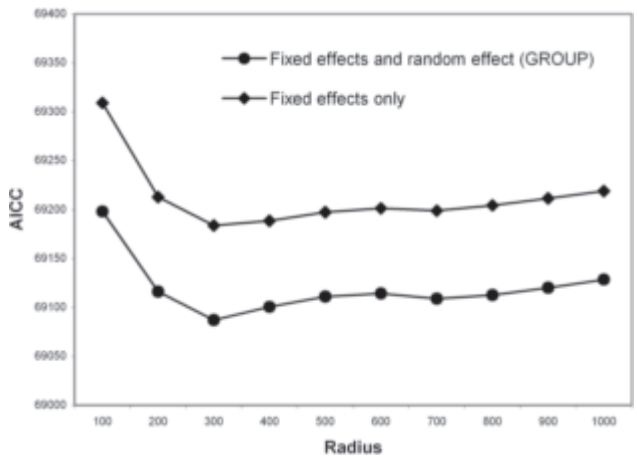

Figure 8. AICC of the fitted models based on different radii.

groups. Therefore, GROUP could alternatively be designed by two levels: vegetation and non-vegetation. This simplified GROUP will not reduce the model efficiency.

Figure 8 also illustrates the variation of AIC_c as a function of the radius at which the spatial metrics are calculated. The lowest AIC_c appears at a radius of 300 meters, which means spatial metrics estimated with a radius of 300 meters can best model classification uncertainty. To have a better idea of 300 meters compared with the average sampling distance, we calculated two statistics: the median of the distances from any survey plot to its closest neighbor plot is 141 meters, and the median of the distances from any sample object to its closest neighbor sample object is 40 meters. Referring back to Figure 3, the former depicts the median distance between two closest triangle symbols and the latter refers to the median distance between two closest round dots. These two statistics explain the significant variables of spatial composition (D_1, D_2, RICH, and EVEN) in the model listed in Table 3. The spatial composition indices at a radius of 100 meters are not significant in modeling classification uncertainty. The reason for this is that a radius of 100 meters confines the sample objects partly covered by the same survey lot, i.e., belonging to the same class. In this case, the variables of sample spatial composition are not really meaningful. At a radius of 200 to 500 meters, sample objects generated from other survey plots are involved, and these indices reasonably reflect the spatial composition of sample objects. Therefore, more variables are significant or have very low p-values (<0.10). At 300 meters, the significance reaches its maximum. It means the classification uncertainty of a sample object is influenced most by neighbor sample objects within 300 meters, which is about twice the median distance between any two survey plots. Spatial composition indices tend to be less significant when the radius becomes greater than 500 meters, since a large radius has a stronger normalization effect and reduces the spatial variation. When the radius is greater than 900 meters,

Table 2. Parameter Estimates and Their Standard Errors in the Mixed Model: (a) Estimates for Random Effect, and (b) Estimates for Fixed Effects

GROUP Level	Forest	Shrub	Herb	Non-veg
Estimate	−10.504	−4.5224	−4.7173	19.7437
Std Error	6.8448	6.8451	6.8491	7.1061
P > |t|	0.2239	0.5567	0.5411	0.0598

(a)

Factors	Intercept	SamRel (0/1)	ELEV	AREA	SamDen*	RICH*	D_1*	D_2*	EVEN*
$\hat{\beta}$	59.4802	−7.1443/0	0.0449	0.0006	0.0429	4.7126	−8.3173	−4.1340	2.5464
Std Err	7.1490	1.4308	0.0046	0.0001	0.0118	1.4124	3.0352	1.8857	0.8310
P > |t|	0.0020	<.0001	<.0001	<.0001	0.0003	0.0009	0.0062	0.0284	0.0022

Note: Spatial metrics SamDen, RICH, D_1, D_2, and EVEN are calculated based on a radius of 300 meters.

(b)

Table 3. Significant Variables from Mixed Linear Model

		Radius (m)									
		100	200	300	400	500	600	700	800	900	1000
Topography	ELEV	√*	√	√	√	√	√	√	√	√	√
	SLOPE	√	√	0.11	0.11	0.09	0.09	0.08	0.10	0.11	0.12
Object features	AREA	√	√	√	√	√	√	√	√	√	√
Sample object reliability	SamRel	√	√	√	√	√	√	√	√	√	√
Sample object density	SamDen	√	√	√	√	√	√	√	√	√	√
Spatial composition	RICH	0.07	√	√	√	0.06	√	0.34	0.40	0.18	0.06
	D_1	0.18	√	√	√	0.08	√	0.66	0.90	0.18	0.07
	D_2	0.31	0.99	√	0.33	0.40	0.55	√	√	0.36	0.65
	EVEN	0.45	√	√	0.86	0.07	0.53	0.10	0.98	0.86	0.60

Note: "√" indicates the variable is significant, i.e., $p < 0.05$.

none of the spatial composition indices are related to classification uncertainty.

Variation of Spatial Metrics with Radius

Spatial metrics denote the group of variables representing sample object density and spatial composition. To examine the variation of the spatial metrics with radius, we correlated each spatial metric for all sample objects at a radius of 300 meters to the same spatial metric at any other radius. The correlation coefficients for each spatial metric (SamDen, RICH, D_1, D_2, and EVEN) are presented in Figure 9. The values of the spatial metrics are dependent not only on the spatial distribution of sample objects, but also on the radius in the estimation. D_2 and EVEN appear less stable and more sensitive to radius than RICH and D_1. This confirms the fact proved by Hill (1973) that D_1 and D_2 in Equations 3 and 4 are the diversity numbers of first order and second order. D_2 depends more on common species/classes and less on rare ones. When the radius is small, both the number of sample objects and the number of classes are small. Therefore, most types look like "rare." As the radius increases, more sample objects are included and common species are explicit. Then, D_2 changes dramatically. SamDen is the least dependent on radius since its estimation takes into account the weight as a function of distance. The sample objects farther away are given less weight in the estimation of the sample object density than those closer.

Referring to the significant variables listed in Table 3, the more stable the spatial metrics are, the more chance the spatial metrics exhibit significance in the model. Sample object density is the most stable variable and it is significant in the model at any radius, while EVEN is the least stable variable and is significant only at radii of 200 and 300 meters. The radius at which spatial metrics are derived is critical in assessing their significance. Some previous research reported the heterogeneity measures are not significant in modeling the classification accuracy in a pixel-based classification (Moisen *et al.*, 2000). Our result provides a possible explanation that the extent of heterogeneity measures could be decisive. In an object-based approach, the extent is the radius, and in a pixel-based approach, the extent is the window size.

Significant Factors Affecting the Classification Uncertainty

In the mixed linear model, we identified the significant variables at the level of $\alpha = 0.05$ by F-test. Table 3 lists these variables for each radius related to the calculation of spatial metrics. In the case of a radius of 300 meters, all the spatial metrics are significant. At least one spatial variable from each of six categories is significant. These significant variables are elevation, object size, sample object density, spatial composition, and sample object reliability. In this case, the model reaches the lowest AIC_c in the fitting. $\hat{\beta}$ denotes the estimate of the parameter β in the Equation 6. The $\hat{\beta}$ estimated by restricted maximum likelihood (REML) is listed in Table 2b. Since the ranges of explanatory variables are different from each other, the absolute value of $\hat{\beta}$ is not directly interpretable. Instead, its sign discloses the positive or negative correlation to classification uncertainty, and p-value indicates the significance. We elaborate on the effects of these factors below:

1. Elevation is positively correlated with the classification uncertainty. It means that classification results are better at high elevation areas than at low elevation areas. From our understanding, the effect of elevation on classification uncertainty is driven by the plant phenologies and spatial distribution at low and high elevations. The images used in this study were collected in the summer, the dry season in California. In the study area, areas at low elevation are often covered with dry herbs, while areas at high elevation are mostly covered by forest. Because forest areas are still green and have a more distinct spectral signature in summer, the classification accuracy is higher. In contrast, dry herbs at low elevation areas loose chlorophyll in summer, making it difficult to distinguish herb species. Therefore, the correlation of topography and classification uncertainty may not necessarily be extended to other areas.
2. Diversity indices are negatively related to classification uncertainty, while sample object density, evenness and richness are positively related to classification uncertainty. The negative parameter estimate of diversity confirms that better classification tends to occur in the areas of more homogeneity and less diversity. Higher diversity implies a larger percentage of boundary objects. These boundary objects have a mix of landscape objects and contain mixed pixels from two or more land-cover classes, which always cause greater confusion between classes. This result is consistent with the conclusion in previous studies of pixel-based classification error. The areas with high sample object density provide more ground truth information so that they reasonably achieve better classification quality. High evenness guarantees that every class present has identically representative samples in the classification, so it rationally benefits classification.
3. The dummy variable, sample object reliability, significantly influences the classification uncertainty. The parameter estimate of sample object reliability is −7.1 for SamRel equal to zero. Since the dependent variable, classification uncertainty, varies from 0 to 100(%) and sample object reliability is a dichotomous variable, the reliable and unreliable sample groups would have 7.1 percent difference in the estimated per-case classification accuracy on average. In this study, about 12 percent of the sample objects were assigned to an incorrect class (Table 4), which mainly resulted from the procedure of automatically generating sample objects from survey plots. As we expected, those unreliable sample objects strongly influenced classification accuracy. Through individual checks, we found a large portion of the unreliable samples had a classification uncertainty equal to zero.

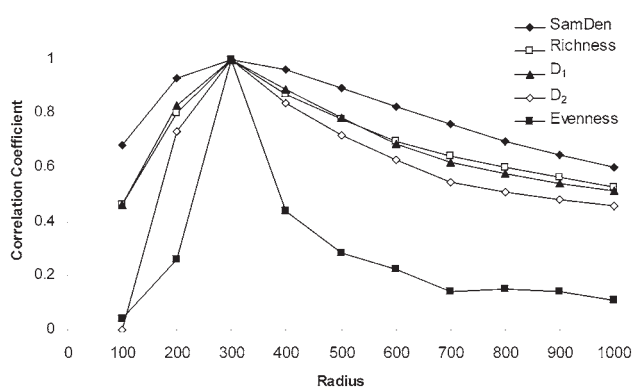

Figure 9. Correlation coefficient of SamDen, RICH, D_1, D_2, and EVEN between measures at a radius of 300 m and at other radii.

TABLE 4. THE PROPORTION OF UNRELIABLE SAMPLE OBJECTS

	Forest	Shrub	Herb	Others	All
Number of sample objects	2795	2263	1550	216	6824
% of unreliable sample objects	10.13%	14.85%	11.10%	15.74%	12.08%

4. The area of image objects is positively associated with the classification uncertainty. Image objects were generated from a region growing segmentation, which uses the local spatial variance of spectral value as the criterion to delineate objects. The object stops growing when the variance exceeds a predefined threshold. Therefore, the size of the objects reflects the local spatial variation of the image. Objects of small size always occur in spectrally heterogeneous areas, such as a highly mixed area, vegetation transitional area, or landscape boundary area. Objects of larger size indicate a spectrally homogeneous area with unitary spectral feature. It is reasonable that the per-case classification accuracy is high in spectrally homogeneous areas, i.e., objects of larger size.

Summary and Conclusions

The goal of this paper is to evaluate the impact of possible factors on the spatial variation of classification accuracy in an object-based vegetation mapping. Classification uncertainty provides a per-case basis classification accuracy to supplement the conventional general accuracy. Along the lines of previous work examining the impact of patch size, landscape characteristics, and terrain on accuracy in pixel-based classification (Steele *et al.*, 1998; Van der Wel *et al.*, 1998; Smith *et al.*, 2002; Smith *et al.*, 2003), we examined six categories of factors: object general membership, topography, sample object density, spatial composition, sample object reliability, and object features. Classification uncertainty derived from a bootstrap method is modeled as a function of these factors using a mixed linear model. The significant variables were identified by F-test ($p < 0.05$) and the parameter estimates of significant variables were resulted from REML fit. The effects of the explanatory variables are interpreted based on parameter estimates. We also investigated variation of spatial metrics with radius, and consequently compared their significance in the model.

The result supports the assumption that elevation, sample object size, sample object reliability, sample object density, and spatial composition all significantly influence classification uncertainty. Elevation, sample object density, evenness, richness, and area of image objects are positively related to the classification uncertainty. Diversity indices are negatively related to the classification uncertainty. The dummy variable, sample object reliability, indicates an unreliable sample would reduce the classification uncertainty by 7.1(%). The significance of those factors is consistent with that of comparable factors in the previous studies of pixel-based classification error. The results of this study helps in understanding the spatial allocation of classification error and suggests further improvement of classification through (a) using wet season images to classify herbs in low elevation areas, (b) adding samples in areas of very low sample object density, and (c) adjusting segmentation parameters to reduce the number of unreliable samples. Some research has reported that larger objects in the segmentation will achieve higher spectral separability (Wang *et al.*, 2004). However, we would not recommend increasing object size in the segmentation, although object size is positively related to classification uncertainty. Large objects are more likely to contain a spectral mixture from multiple landscape classes, and result in a larger number of unreliable sample objects.

Except for topography, all other five categories of effects are related to the object-based approach. With the increasing use of the object-based approach in analyzing high-resolution imagery, it is necessary to consider how to properly assess and interpret classification accuracy in an object-based classification. Conventional pixel-based classification uses the regular cell (pixel) as the processing unit, which is much smaller than the mapping or sampling unit. While in object-based classification, sample objects are not quite equivalent to the field sampling plots since one field plot always superimposes on a cluster of image objects. Therefore, the practice of accuracy assessment is directly or indirectly affected by sample object generation, sample object reliability, and object dimension. The impact of sample and sampling strategy on classification accuracy in object-based classification is much more complex than in pixel-based classification. The results of this study indicate there is a need to develop standardized procedures for measuring and comparing accuracies in object-based classification.

Acknowledgments

This research was supported by the National Park Service. We are grateful to Dave Schirokauer, Pam Van Der Leeden from NPS and Nick Clinton from U.C. Berkeley for providing us with help in field sample collection and valuable suggestions in the work. We also would like to thank three referees for their comments on the original manuscript.

References

Baatz, M., M. Heynen, P. Hofmann, I. Lingenfelder, M. Mimier, A. Schape, M. Weber, and G. Willhauck, 2004. *eCognition User Guide 4.0: Object Oriented Image Analysis*, Definiens Imaging GmbH, Munich, Germany, 485 p.

Bailey, T.C., and A.C. Gatrell, 1995. *Interactive Spatial Data Analysis*, Prentice Hall, Harlow, Essex, England, 413 p.

Burnham, K.P., and D.R. Anderson, 2002. *Model Selection and Inference: A Practical Information-Theoretic Approach*, Second edition, Springer, New York, 488 p.

The California Natural Diversity Database, 2003. The Vegetation Classification and Mapping Program, Department of Fish and Game-Wildlife and Habitat Data Analysis Branch, September 2003 edition, http://www.dfg.ca.gov/whdab/pdfs/natcomlist.pdf (last date accessed: 01 April 2008).

Campbell, J.B., 1981. Spatial correlation-effects upon accuracy of supervised classification of land-cover, *Photogrammetric Engineering & Remote Sensing*, 47(3):355–363.

Canters, F., 1997. Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification, *Photogrammetric Engineering & Remote Sensing*, 63(4):403–414.

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

Congalton, R.G., 1988b. Using spatial auto-correlation analysis to explore the errors in maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 54(5):587–592.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data*, CRC Press, 137 p.

Dicks, S.E., and T.H.C. Lo, 1990. Evaluation of thematic map accuracy in a land-use and land-cover mapping program, *Photogrammetric Engineering & Remote Sensing*, 56(9):1247–1252.

Efron, B., and R. Tibshirani, 1993. *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Efron, B., and R. Tibshirani, 1997. Improvements on cross-validation: The.632+ bootstrap method, *Journal of American Statistical Association*, 92(438):548–560.

Fisher, P.F., 1994. Visualization of the reliability in classified remotely-sensed images, *Photogrammetric Engineering & Remote Sensing*, 60(7):905–910.

Foody, G., 2002. Status of land-cover classification accuracy assessment, *Remote Sensing of Environment*, 80(1):185–201.

Foody, G., N. Campbell, N. Trodd, and T. Wood, 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification, *Photogrammetric Engineering & Remote Sensing*, 58(9):1335–1341.

Friedl, M.A., C. Woodcock, S. Gopal, D. Muchoney, A.H. Strahler, and C. Barker-Schaaf, 2000. A note on procedures used for accuracy assessment in land-cover maps derived from AVHRR data, *International Journal of Remote Sensing*, 21(5):1073–1077.

Gong, P., and P.J. Howarth, 1990. An assessment of some factors influencing multispectral land-cover classification, *Photogrammetric Engineering & Remote Sensing*, 56(5):597–603.

Gong, P., R. Pu, and J. Chen, 1996. Mapping ecological land systems and classification uncertainties from digital elevation and forest-cover data using neural networks, *Photogrammetric Engineering & Remote Sensing*, 62(11):1249–1260.

Hastie, T., R. Tibshirani, and J. Friedman, 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer-Verlag, New York, 533 p.

Hilborn, R., and M. Mangel, 1997. *The Ecological Detective: Confronting Models with Data*, Princeton University Press, Princeton, New Jersey.

Hill, M., 1973. Diversity and evenness – Unifying notation and its consequences, *Ecology*, 54(2):427–432.

Hoeting, J.A., R.A. Davis, A.A. Merton, and S.E. Thompson, 2006. Model selection for geostatistical models, *Ecological Applications*, 16(1):87–98.

Keeler-Wolf, T., 1999. *Field and Photo-interpretation Key to the Vegetation Alliances and Defined Associations from the Point Reyes National Seashore, Golden Gate National Recreation Area, San Francisco Municipal Water District Lands, and Mt. Tamalpais, Tomales Bay, and Samuel P. Taylor State Parks*, unpublished vegetation key.

Lewis, A., and M.F. Hutchinson, 2000. From data accuracy to data quality: using spatial statistics to predict the implications of spatial error in point data, *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing* (H.T. Mowrer and R.G. Congalton, editors), Ann Arbor Press, Chelsea, Michigan, pp. 17–35.

Littell, R.C., W.W. Stroup, and R.J. Freund, 2002. *SAS for Linear Models*, Fourth edition, SAS Publishing, Cary, North Carolina, SAS Institute, Inc., 466 p.

Lunetta, R.S., R.G. Congalton, L.K. Fenstermaker, J.R. Jensen, K.C. Mcgwire, and L.R. Tinney, 1991. Remote-sensing and geographic information-system data integration – Error sources and research issues, *Photogrammetric Engineering & Remote Sensing*, 57(6):677–687.

Moisen, G.G., D.R. Cutler, and T.C. Edwards, 2000. Generalized linear mixed models for analyzing error in a satellite-based vegetation map of Utah, *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing* (H.T. Mowrer and R.G. Congalton, editors), Ann Arbor Press, Chelsea, Michigan, pp. 37–43.

Pu, R.L., and P. Gong, 2004. Determination of burnt scars using logistic regression and neural network techniques from a single post-fire Landsat 7 ETM+ image, *Photogrammetric Engineering & Remote Sensing*, 70(7):841–850.

SAS Institute Inc., 2002. *SAS OnlineDoc® 9*, Cary, North Carolina, SAS Institute, Inc.

Silverman, B.W., 1998. *Density Estimation for Statistics and Data Analysis*, Second editon, Chapman & Hall/CRC Press, Boca Raton, Florida, 175 p.

Smith, J.H., S.V. Stehman, J.D. Wickham, and L.M. Yang, 2003. Effects of landscape characteristics on land-cover class accuracy, *Remote Sensing of Environment*, 84(3):342–349.

Smith, J.H., J.D. Wickham, S.V. Stehman, and L.M. Yang, 2002. Impacts of patch size and land-cover heterogeneity on thematic image classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 68(1):65–70.

Steele, B.M., J.C. Winne, and R.L. Redmond, 1998. Estimation and mapping of misclassification probabilities for thematic land-cover maps, *Remote Sensing of Environment*, 66(2):192–202.

Stehman, S.V., 1992. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 58(9):1343–1350.

Turner, M.G., R.H. Gardner, and R.V. O'Neill, 2001. *Landscape Ecology in Theory and Practice: Pattern and Process*, Springer, New York, 401 p.

Unwin, D., 1995. Geographical information systems and the problem of "error and uncertainty", *Progress in Human Geography*, 19(4):549–558.

Van der Wel, F.J.M., L.C. Van der Gaag, and B.G.H. Gorte, 1998. Visual exploration of uncertainty in remote-sensing classification, *Computers & Geosciences*, 24(4):335–343.

Van Oort, P.A.J., A.K. Bregt, S. De Bruin, and A.J.W. De Wit, 2004. Spatial variability in classification accuracy of agricultural crops in the Dutch national land-cover database, *International Journal of Geographical Information Science*, 18(6):611–626.

Wang, L., W.P. Sousa, and P. Gong, 2004. Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery, *International Journal of Remote Sensing*, 25(24):5655–5668.

Yu, Q., P. Gong, N. Clinton, G. Biging, M. Kelly, and D. Schirokauer, 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery, *Photogrammetric Engineering & Remote Sensing*, 72(7):799–811.

(Received 09 February 2007; accepted 13 August 2007; revised 30 October 2007)