Improving Classification Accuracy Assessments with Statistical Bootstrap Resampling Techniques

Keith T. Weber, GIS Director, Idaho State University GIS Center, 921 S. 8th Ave. Stop 8104, Pocatello, Idaho 83209-8104, webekeit@isu.edu

Jackie Langille, Central Washington University, 2203 Brooksfield Ellensburg, WA 98926, langille@geology.cwu.edu

ABSTRACT

The use of remotely sensed imagery to generate land cover models is common today. Validation of these models typically involves the use of an independent set of ground-truth data which are used to calculate an error matrix resulting in estimates of omission, commission, and overall error. However, each estimate of error contains a degree of uncertainty itself due to 1) conceptual bias, 2) location/registration and co-registration errors, and 3) variability in the sample sites used to produce and validate the model. In this study, focus was not placed upon describing land cover mapping techniques, but rather the application of bootstrap resampling to improve the characterization of classification error, demonstrate a method to determine uncertainty from sample site variability, and calculate confidence limits using statistical bootstrap resampling of 500 sample sites acquired within a single Landsat 5 TM image. The sample sites represented one of five land cover categories (water, roads, lava, irrigated agriculture, and rangelands) with each category containing 100 samples. The sample set was then iteratively resampled (n=200) and 65 sites were randomly selected (without replacement) for use as classification training sites while the balance (n=35) were used for validation. Imagery was subsequently classified using a maximum likelihood technique and the model validated using a standard error matrix. This classification-validation process was repeated 200 times. Confidence intervals were then calculated using the resulting omission and commission errors. Results from this experiment indicate that bootstrap resampling is an effective method to characterize classification uncertainty and determine the effect of sample bias.

KEYWORDS: Remote sensing, accuracy assessment, uncertainty, error, confidence intervals

INTRODUCTION

Classification of remotely sensed imagery provides a means to model the earth's surface (Lillesand and Kiefer 2004). Studies implementing this technology have ranged from mapping weeds (Glenn et al., 2005; Lass et al., 2005; Gokhale and Weber, 2006) to modeling wildfire susceptibility (Ercanoglu et al., 2006), forecasting urban growth (McMahan et al., 2002), and a multitude of other applications. In theory, and based upon the results of other scientists (Boardman, 1993; Ray and Murray, 1998; Roberts et al., 1998) classification accuracy is a function of a pixel's spectral similarity to the training sites (pixels with a high proportion of target material). As a result, pixels having high proportions of the target (more spectrally pure) will generally classify correctly more often than pixels containing a lower proportion of target material (spectrally mixed), assuming uniform reflectance characteristics of all objects within each pixel (Mundt et al., 2006).

Regardless of the scale, scope, or resolution of image classification, error assessment is a critical step in the analysis and interpretation process (Stehman and Czaplewski, 1998;, Janssen and van der Wel, 1994). The development of current error assessment techniques has occurred in three stages (Congalton and Green, 1999): 1) no constraint, 2) qualitative general assessments (non-site specific), and 3) quantitative site specific assessments. Longley et al. (2005) similarly describe uncertainty and its assessment as beginning with one's conception of the real-world as this affects how one perceives, observes, defines, and categorizes the world. In turn, one's paradigm influences, and perhaps even drives, measurements and representations which consequently affect nominal classification accuracy assessments. We fully realize this inherent error exists and is an important consideration when one attempts to model natural systems using crisp logic (Sauder et al., 2003). Thus, the quantification of classification error contains both true classification error (the classification algorithm made a mistake) and conception error (how one observes the natural world during the acquisition of training sites) which are inextricably linked together.

Today, classification accuracy is frequently determined using an error matrix of predicted (classified) versus known (reference) occurrences of a target (Congalton, 1991). These tabulations produce estimates of omission, commission, and overall error, and may also be used to calculate statistical measures of accuracy such as Kappa, Tau, and Z-statistics (Titus et al., 1984; Ma and Redmond, 1995; Congalton and Green, 1999; Foody 2002, 2004). While an error matrix is commonly used to tabulate classification accuracy, the number of samples necessary to validate the model is debated, because sample size relates directly to power of analysis (Cohen, 1992, Taylor and Gerrodette, 1993), which is a function of sample site variability, and total project cost. Congalton (1991) stated that "sufficient samples must be acquired to be able to adequately represent the confusion" (i.e., the categories and types of errors committed in the model). One option to accommodate high variability and to assess its effect on classification results is the use of bootstrap resampling. Applying bootstrap resampling does not imply that classifications should be made with an insufficient number of samples. Rather, bootstrap resampling presents a methodology to make better use of the samples available and more easily determine the adequacy of the existing sample set. The authors note that this paper does not attempt to determine minimum sample size as other resources are currently available to aid in this decision such as the recent work published by Van Niel et al. (2005) and Foody et al (2006) and Foody and Mathur (2006).

To improve the reliability of error calculations and better characterize uncertainty, statisticians have applied various techniques including bootstrap resampling (Good, 2005; Efron, 1979). This technique has also been applied with spatial data (Mahalanobis, 1946; Hall, 2003) and remote sensing studies (Okeke and Karniele, 2006; Verbyla and Boles 2000; McMahan and Weber 2003). The general idea behind bootstrap resampling is that with a sufficiently large set of samples one can treat the sample set as representative of a population. By dividing the sample set into classification and validation sub-sets one can perform a classification and then validate the resulting classification to arrive at an estimate of classification accuracy using only a single sample set. To reduce the effects of bias (i.e., the difference in the type of samples chosen for training versus validation), the sample set can be iteratively resampled to draw new classification and validation sub-sets. Using this process a more robust estimation of uncertainty can be achieved as the user retains the use of all samples and consequently, statistical power, while producing a better characterization of classification error where confidence intervals are calculated directly (cf. Snedecor and Cochran, 1967). This paper explores a parallel approach to classification error assessment and applies iterative bootstrap resampling to improve the characterization of classification error and better represent bias within the sample set (Jones, 1956; Efron, 1979; Good, 2005). An interesting alternative to this and other methods based upon the standard error matrix is the fuzzy error matrix described by Binaghi et al. (2005) and the calculation of misclassification probability as described by Steele et al (1998). The latter applies Kriging and a lattice of training locations to construct an accuracy model similar to a contour map. This tool, however, has not been as broadly applied or accepted as the error matrix.

METHODS

To explore the use of iterative bootstrap resampling to improve error assessment of remote sensing classifications, we designed an experiment using 500 sample points collected from one of five distinct land cover categories (water, roads, lava, irrigated agriculture, and rangelands, respectively) within a study area in southeastern Idaho (Figure 1). One-hundred points (pixels) were acquired within each category using standard on-screen digitizing and one Landsat 5 TM scene (path 39, row 30) acquired on 7-July-2003. One meter digital orthophotos were used to ensure each sample point was clearly located within the correct land cover type. In addition, sample points were corroborated using field observations made during the summer of 2003 (n=253). For each class, sixty-five sample points were randomly selected without replacement for use as training sites while the remainder (n=35) were used for validation. These points were then rasterized using the same spatial parameters as the satellite imagery used for classification (28.5 x 28.5 m).

Landsat imagery was prepared for classification by generating a normalized difference vegetation index (NDVI) and principal component analysis (PCA) images (Eastman, 2003) from raw radiance values (brightness values or DN). The first two PCA bands (PCA-C1 and PCA-C2) contained >95% of the unique spectral variance, therefore the NDVI and two PCA raster images were used for classification. This approach is justified because previous studies have found these data reliable (Rouse et al., 1973; Kriegler et al., 1969) and have applied these data for land cover characterization with positive results (Ingebritsen and Lyon 1985, Singh 1989, Collins and Woodcock 1996, Savage 2005). Two-hundred classification-validation iterations were completed using maximum-likelihood classification in Idrisi Kilimanjaro software (Eastman, 2003). For each iteration, accuracy was assessed using standard error matrices. Each classification-validation iterations included 65 training samples and 35 validation samples, respectively. The goal of this study was not to produce perfect classification results, but rather to explore the use of bootstrap resampling to better assess classification error and bias within the sample set while applying well documented classification techniques. We assumed that sample sites exhibited some degree of autocorrelation (Tobler 1970, Miller 2004) but due to the size of the study area (approximately 35,144 km^2) and distribution of sample sites, spatial autocorrelation was not considered a confounding factor. Geo-registration and co-registration of training and validation sample sites were also considered and explored as a source of error. To explore potential error we used on-screen techniques where we zoomed into each site (approximate scale = 1:1,000) and using both Landsat imagery and NAIP orthoimagery (true color 1-m spatial resolution) were able to verify the category assigned to each site. It should be noted that the categories used in this study were general (water, roads, lava, irrigated agriculture, and rangelands) and thereby, easily verified and differentiated from one another. Thus, the above errors were considered minimal for two reasons; 1) sites were acquired using careful on-screen digitizing and verification and 2) the size of each pixel (28.5 x 28.5 m) made it relatively easy to place a sample site point within it, thereby minimizing co-registration error (Weber, 2006).

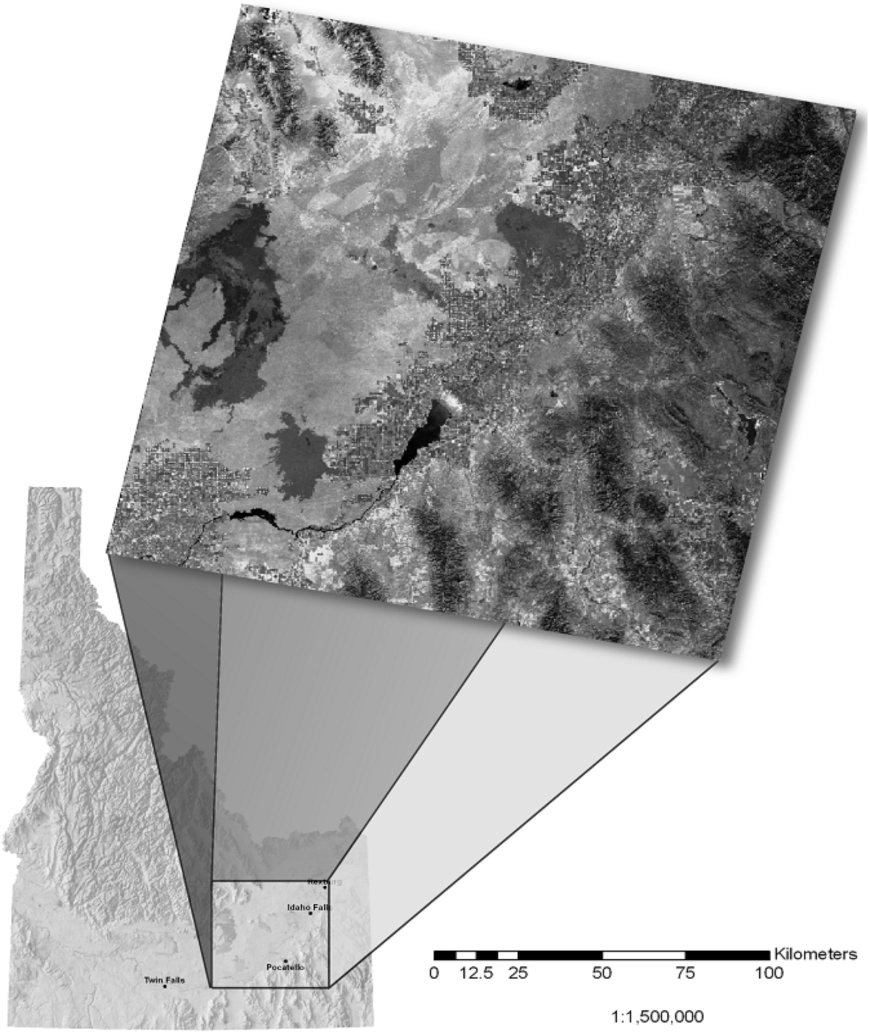

Figure 1. The study area location and Landsat 5 TM imagery used in this research (path 39, row 30, acquired 7-July-2003).

The resulting omission and commission errors from each classification error assessment were then entered into a Microsoft Excel spreadsheet. To determine the minimum number of iterations required for classification error characterization, various statistical techniques have been used such as the cumulative variance approach described by Mundt et al. (2005). We applied an approach similar to Mundt et al. (2005) where cumulative standard deviation (standard error) was graphed for omission and commission errors. The resulting graphs were then investigated to determine the first point at which a “practical sill” has been achieved. The practical sill was defined as the point where the curve developed by the standard error plot has dropped to within 5% of the asymptote, or conversely, has dropped 95% of the difference observed between maximum and minimum standard error (Isaaks and Srivastava 1990) (Equation 1). The number of classification-validation iterations required to achieve the “practical sill” can then be determined by calculating a line of best fit using either inverse or exponential functional forms. This was accomplished in Microsoft Excel by fitting a trend line to the curve and using the equation of the line and its R^2 value. If the fit is sufficient (we suggest a threshold R^2 of 0.70) the minimum number of iterations is then calculated by finding X when Y (i.e., S_p) is known. Alternatively, the number of iterations can be read from the figure itself. Ultimately, all methods are in some ways subjective in that the user must decide the threshold or point at which inflection occurs or where the variance curve “flattens”. Using a

Microsoft Excel spreadsheet containing all classification/validation iterations (n=200) also enabled us to directly calculate the 95% confidence interval describing classification error (Good, 2005, NBII, 2006).

$$R = SE_{max} - SE_{min}$$
$$S_p = SE_{max} - 0.95R \qquad \text{Eq. 1}$$

Where R is the range and S_p is the practical sill.

RESULTS AND DISCUSSION

The resulting error from two of the 200 classification iterations were randomly selected and given in Tables 1 and 2. These classification results illustrate the variability in accuracies where a randomly selected subset (n=65) of training sites were used for classification. The variability between the two example classification results illustrates a scenario where one or more of the target features exhibit overlapping spectral signatures. To visualize the characteristics of these spectral signatures (n=65), we used spectral comparison charts where each cover class' minimum, maximum, and mean value was displayed for each band (e.g., PCA-C1, PCA-C2, and NDVI). Referring to the signature comparison charts (Figure 2) one can see the water class overlaps with or encompasses several other classes. We speculate this is because some of the water training sites may have been located in shallow water where vegetation, rocks, or substrate affected the spectra. However this did not adversely affect the goal of this study as the authors sought to understand the effect of training site selection and resampling selections on classification accuracy. Indeed this variability and difficulty aided the author's in their goal. Further, some of the training sites may have been in areas where algae or turbidity affected the spectra. In cases like this, where seperability (Richards, 2005) is problematic, one can apply statistical purification (McKay and Campbell, 1982; Eastman, 2003) to the training sites and/or use canonical correlation analysis (Eastman, 2003) of the imagery instead of the principal components analysis used in this study. Purifying spectral signatures effectively makes the signatures more homogenous or uniform by removing training sites that fall outside the typicality threshold. This may not be a viable option, however, if reducing the sample size (i.e., training sites) results in loss of statistical power (Cohen 1992, Taylor and Gerodette, 1993) as it would have been in this case.

Table 1. Sample of percent classification error where 65 points were randomly selected for classification and 35 points were used for validation from a total sample set of 100 points per class.

Class	Water	Roads	Lava	Irrigated Agriculture	Rangelands	Sum	Commission Error (%)
Water	2	0	0	0	0	2	0
Roads	4	25	1	0	2	32	22
Lava	7	0	33	0	0	40	18
Irrigated Agriculture	3	0	0	31	0	34	9
Rangelands	7	3	0	0	29	39	26
Unclassified	12	7	1	4	4	28	
Sum	35	35	35	35	35	175	
Omission Error (%)	94	29	6	11	17		31

Table 2. Sample of percent classification error where a second set of 65 randomly selected points were used for classification and 35 points were used for validation from the same sample set of 100 points per class.

Class	Water	Roads	Lava	Irrigated Agriculture	Rangelands	Sum	Commission Error (%)
Water	35	0	0	0	0	35	0
Roads	0	32	1	0	0	33	3
Lava	0	0	34	0	0	34	0
Irrigated Agriculture	0	0	0	34	0	34	0
Rangelands	0	1	0	0	35	36	2
Unclassified	0	2	0	1	0	3	
Sum	35	35	35	35	35	175	
Omission Error (%)	0	8	2	2	0		2

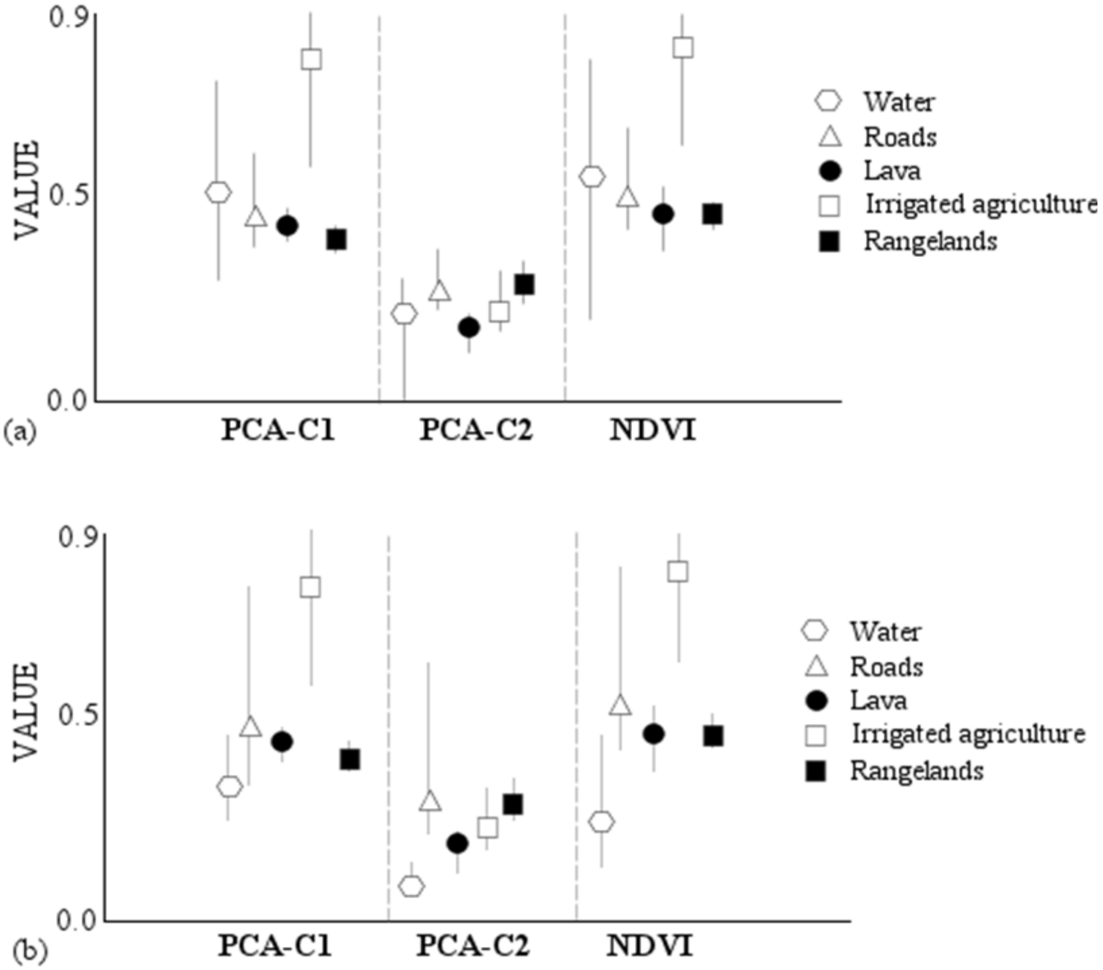

Figure 2. (a,b) Minimum, maximum, and mean spectral signature values for the training sites (n=65) used in this study. Fig 2a represents signatures used to produce classification results reported in table one. Fig. 2b represents signatures used to produce classification results reported in table two.

Based upon the results of bootstrap resampling, a 95% confidence interval was calculated using all classification-validation iterations (n=200) (Table 3). Not surprisingly, the water class had very broad commission and omission errors. Other classes performed better resulting in a tighter range of commission and omission errors.

Table 3. Calculated error (at 95% confidence) for each class based upon 200 classification-validation iterations.

Class	Commission error	Omission error
Water	0.00 - 0.50	0.01 - 0.97
Roads	0.00 - 0.26	0.08 - 0.31
Lava	0.00 - 0.27	0.00 - 0.14
Irrigated agriculture	0.00 - 0.23	0.02 - 0.18
Rangelands	0.00 - 0.26	0.00 - 0.20

The use of additional sample points (pixels) or the use of sample polygons instead of sample points may have improved classification results and reduced the number of iterations required to achieve the “practical sill”. However, the methods used to explore and understand training site variability (bootstrap resampling) and determine the practical sill would remain unchanged.

By inspecting graphs of cumulative standard error (SE) for omission error (producer’s accuracy), the minimum number of classification-validation iterations required to capture classification variability was determined as the point where SE met the practical sill (Figure 3). The least number of iterations required to achieve the “practical sill” was 82 iterations for the lava land cover class. Table 4 gives calculated error at 95% confidence where only the first 82 classification-validation iterations were used for all cover classes. Notice that the other cover classes responded differently and required from 91 to 656 classification-validation iterations to achieve their “practical sill” (Table 5). While the authors only performed 200 classification-validation iterations, the value of 656 was derived synthetically using the regression line was fitted to SE curve, and solving for the number of iterations required to achieve the practical sill. To minimize repetition in the rest of this paper, only results from the lava cover class will be discussed. Further, as this technique was designed to improve the characterization of classification error and determine the effect of sampling bias due to training site heterogeneity, we will focus upon omission error (producer’s accuracy).

Table 4. Calculated error (at 95% confidence) for the each land cover class based upon 82 classification-validation iterations[1].

Class	Commission error (%)	Omission error (%)
Water	0.0 – 75.0	4.0 – 97.0
Roads	2.0 – 23.0	11.0 – 27.0
Lava	0.0 – 25.0	0.0 – 15.0
Irrigated agriculture	0.0 – 21.0	3.0 – 18.0
Rangelands	2.0 – 26.0	3.0 – 23.0

1- Eighty-two classification-validation iterations were used as the minimum number of iterations necessary to achieve the “practical sill” derived from the standard error curve for the lava cover class (figure 3).

Table 5. The minimum number of classification-validation iterations required to achieve the practical sill[1]

Class	Iterations required	R^2 for line of best fit
Water	210[2]	0.894
Roads	102	0.380
Lava	82	0.989
Irrigated agriculture	656[2]	0.470
Rangelands	91	0.923

1. Based upon the cumulative standard error calculated from omission error.
2. These values were derived synthetically using the regression equation to estimate the number of iterations required to achieve the practical sill.

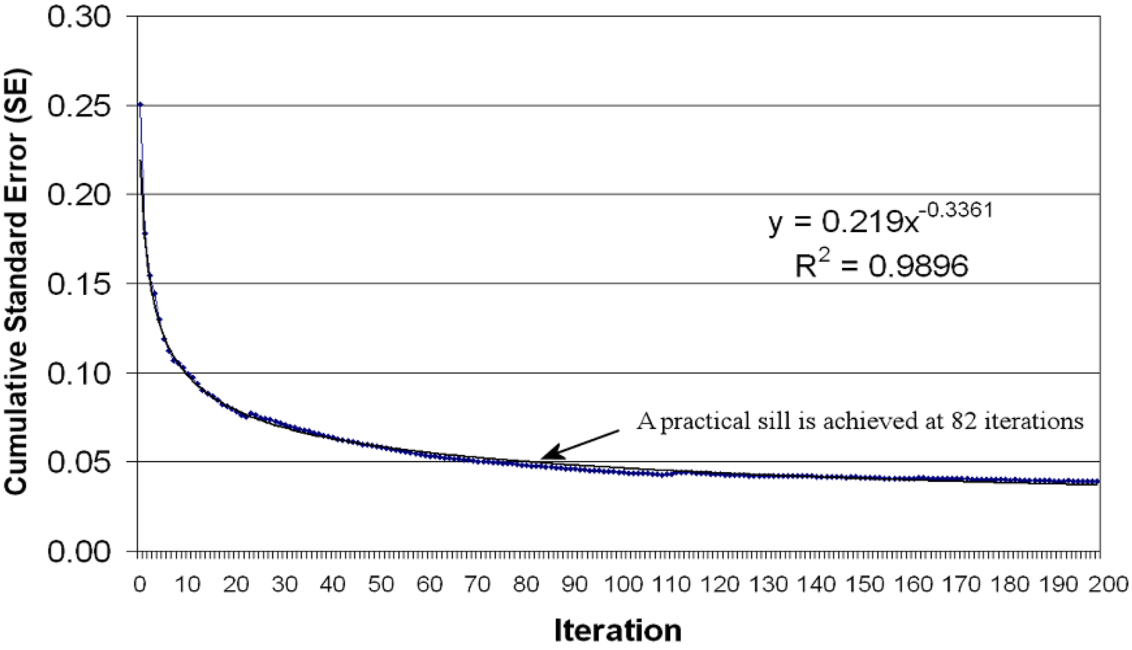

Figure 3. Example of how the minimum number of iterations required to capture the variability in classification error was determined (note: this graph represents cumulative standard error of omission error for the lava cover class). The equation describing the exponential line of best fit is given along with the R-squared value.

The difference between calculated error using 200 versus 82 classification-validation iterations is shown in Figure 4 for the lava land cover class. Based upon these results we concluded that 82 classification-validation iterations were an adequate number to characterize classification error within this cover class. It should be pointed out that –like the other land cover classes used in this paper-- other data sets may require more or perhaps fewer classification-validation iterations. Following the procedures described in this paper, one can easily determine a “practical sill” from SE curves and hence, the minimum number of iterations required for classification error characterization. From these data, one can infer a great deal about sample set bias as the number of iterations required is a function of training site variability (high variability training sites will require additional iterations) along with imagery characteristics such as radiometric, spatial, and spectral resolution, as well as scale of analysis.

CONCLUSIONS

The study was performed to better understand the effect of training site selection and subset resampling on classification accuracy. As a result of this focus, the authors were not concerned about classification accuracy or the development of a reliable land cover mapping technique. Rather, this study focused upon the application of bootstrap resampling to improve the characterization of classification error and hence, uncertainty is a time-consuming process. However, based upon the results of this study, bootstrap resampling greatly improved our understanding of classification uncertainty because variability of classification accuracy was determined, which led to a better understanding of the influence of training site heterogeneity (sample bias). This study indicates that after calculating standard error for any number of classification-validation iterations, if one notices a cumulative standard error that approximates an exponential curve it should be clear that the heterogeneity within the training sites is large and may have significant effects upon classification results. Further, if the asymptote or practical sill is not reached with the number of iterations performed (note the associated low R^2 value (Table 5) for the line fitted to the cumulative SE curve), this indicates the training sites are extremely heterogeneous, requiring the application of either 1) spectral signature purification, 2) spectral mixture analysis, 3) some form of data reduction technique, or 4) additional sample points. We note that this process, like all other statistically-based techniques, requires a sufficient number of observations per class/category to allow subsets to be created that retain statistical power of analysis. Bearing this in mind, bootstrap resampling can be applied to improve the characterization of classification error regardless of the type of imagery, satellite sensor, or classification methodology used.

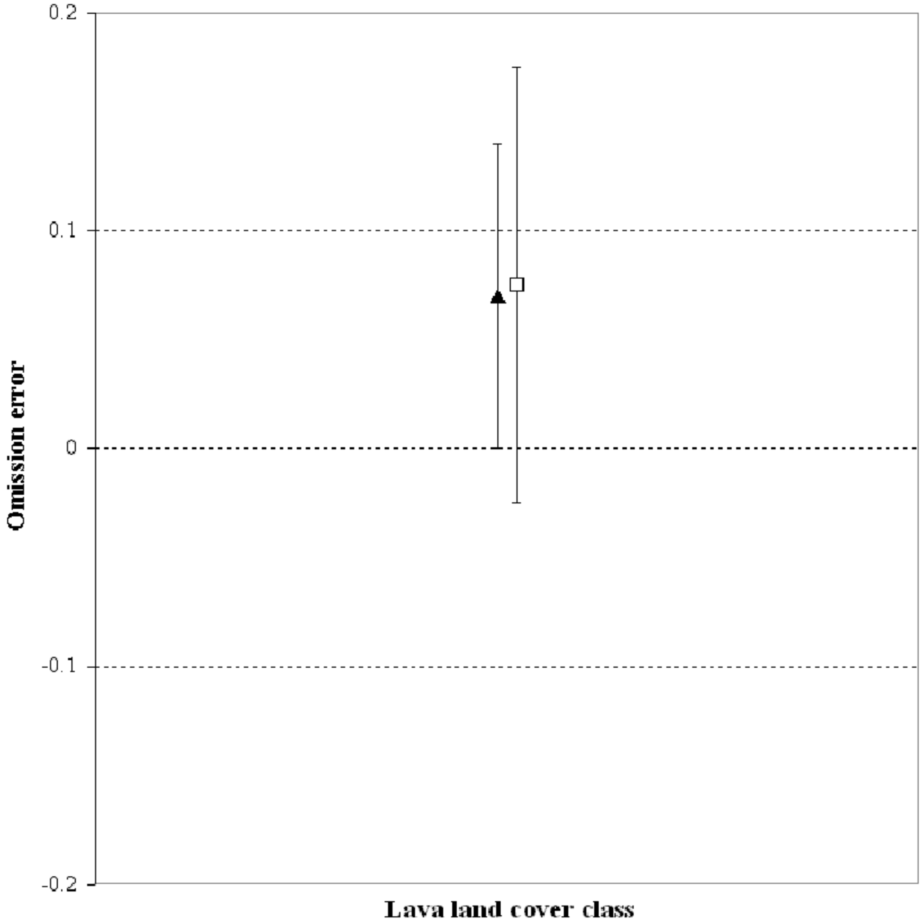

Figure 4. Differences between calculated omission error based upon 200 classification-validation iterations (triangles) and 82 classification-validation iterations (squares). Note: the marker symbol is used to indicate the mean while the whiskers represent a 95% confidence interval of the estimated distribution of errors.

ACKNOWLEDGEMENTS

This study was made possible by a grant from the National Aeronautics and Space Administration Goddard Space Flight Center. ISU would also like to acknowledge the Idaho Delegation for their assistance in obtaining this grant and the assistance of Nancy Glenn, Jacob Mundt, John Welhan and Teri Peterson for their assistance and advice.

LITERATURE CITED

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

Boardman J.W. 1993. Automating spectral unmixing of AVIRIS data using convex geometry concepts. In: Summaries of the Fourth Annual JPL Airborne Geoscience Workshop. pp. 11-14. JPL Publication 93-26, Arlington, VA.

Cohen, J. 1992. Statistical Power Analysis for the Behavioral Sciences. LEA. 567 pp.

Collins, J.B. and C.E. Woodcock. 1996. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. Remote Sensing of Environment, 56: 66-77.

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

Congalton R.G. and K. Green. 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Boca Raton: Lewis Publishers. 160pp.

Eastman R. J. 2003. Idrisi Kilimanjaro Guide to GIS and Image Processing. Clark University Laboratory.

Efron B. 1979. Bootstrap methods: Another Look at the Jackknife. Annals of Statistics. 7(1):1-26

Ercanoglu, M., K. T. Weber, J. M. Langille, and R. Neves. 2006. Modeling Wildland Fire Susceptibility Using Fuzzy Systems. GIScience and Remote Sensing 43(3):268-282.

Foody G. M. 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment, 80, 185-201.

Foody G. M. 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering and Remote Sensing, 70, 627-633.

Foody, G. M., and A. Mathur. 2006 The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. Remote Sensing of Environment. 103(2006) 179-189.

Foody, G. M., A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd. 2006. Training set size requirements for the classification of a specific class. Remote Sensing of Environment, 104 (2006) 1-14.

Glenn N.F., J. T. Mundt, K. T. Weber, T. S. Prather, L. W. Lass, and J. Pettingill. 2005. Hyperspectral data processing for repeat detection of small infestations of leafy spurge. Remote Sensing of Environment, 95, 399-412.

Gokhale, B. and K. T. Weber. 2006. Modeling Cheatgrass using Quickbird Imagery. Pages 77-84 in K. T. Weber (ed.), Final Report: Detection, Prediction, Impact, and Management of Invasive Plants using GIS. 196 pp.

Good, P. I. 2005. Resampling Methods: A Practical Guide to Data Analysis. 3rd Edition. Birkhauser. 218 pp.

Hall, P. 2003. A short prehistory of the boostrap. Statistical Science. 18(2):158-167.

Ingebritsen, S.E. and R. J. P. Lyon. 1985. Principal components analysis of multitemporal image pairs. International Journal of Remote Sensing, 6(5): 687-696.

Isaaks, E. H. and Srivastava, R. M. 1990. Introduction to Applied Geostatistics. Oxford University Press, NY. 592 pp.

Janssen, L. L. F. and F. J. M. van der Wel. 1994. Accuracy Assessment of Satellite Derived Land-Cover Data: A Review. Photogrammetric Engineering and Remote Sensing, 60(4): 419-426.

Jones, H. L. 1956. Investigating the properties of a sample mean by employing random subsample means. J. Amer. Statist. Assoc. 51:54--83.

Kriegler, F.J, W. A.Malila, R. F. Nalepka, and W. Richardson. 1969. Preprocessing transformations and their effects on multi-spectral recognition. In: Proceedings of the Sixth International Symposium on Remote Sensing of Environment. University of Michigan, Ann Arbor, MI., USA, 97-131.

Lass, L. W., T. S. Prather, N. F. Glenn, K. T. Weber, J. T. Mundt, J. Pettingill. 2005. A review of remote sensing of invasive weeds and example of the early detection of spotted knapweed (Centaurea maculosa) and babysbreath (Gypsophila paniculata) with a hyperspectral sensor. Weed Science, 53:242-251

Lillesand, T. M. and R. W. Kiefer. 2000. Remote Sensing and Image Interpretation. 4th Ed. John Wiley and Sons, New York, NY 724 pp.

Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 2005. Geographic Information Systems and Science. 2nd Ed. Wiley 515 pp.

Ma, Z. and R. L. Redmond 1995. Tau Coefficient for Accuracy Assessment of Classification of Remote Sensing Data. Photogrammetric Engineering and Remote Sensing, 61(4):435-439.

Mahalanobis, P. C. 1946. Sample Surveys of Crop Yields in India. Sankhya. 7:269-280.

McKay, R.J. and N. A. Campbell. 1982, Variable selection techniques in discriminant analysis II: Allocation, British Journal of Mathematical and Statistical Psychology, 35, 30-41.

McMahan, J. B., K. T. Weber, and J. D. Sauder. 2002. Using Remote Sensing Data in Urban Sprawl and Green Space Analyses. Intermountain Journal of Sciences 8(1):30-37.

McMahan, J. B. and K. T. Weber. 2003. Validation Alternatives for Classified Imagery. Pages 70-73 in K. T. Weber (Ed), Final Report: Wildfire Effects on Rangeland Ecosystems and Livestock Grazing in Idaho. 209pp.

Miller, H. J. 2004. Tobler’s First Law and Spatial Analysis. Annals of the Association of American Geographers. 94(2):284-289.

Mundt, J.T., N. F. Glenn, K. T. Weber, T. S. Prather, L. W. Lass, J. A. Pettingill. 2005. Discrimination of hoary cress and determination of its detection limits via hyperspectral image processing and accuracy assessment techniques. Remote Sensing of Environment 96(1), 509-517.

Mundt, J. T., N. F. Glenn, K. T. Weber, and J. A. Pettingill. 2006. Assessing Detection Limits and Confidence from Classification Accuracy. Remote Sensing of the Environment. 105 (2006) 34-40.

NBII. 2006. Accuracy Assessment Procedures: Computational Issues. URL: http://biology.usgs.gov/npsveg/aa/sect5.html visited 9-Feb-2006.

Okeke, F. and A. Karnieli. 2006. Methods for fuzzy classification and accuracy assessment of historical aerial photographs for vegetation change analyses. Part I: Algorithm development. International Journal of Remote Sensing. 24(1-2): 153-176.

Ray T.W. and B. C. Murray B.C. 1998. Non-linear spectral mixing in desert vegetation. Remote Sensing of Environment, 55, 59-64.

Richards, J.A. 2005. Remote Sensing Digital Image Analysis. Springer-Verlag, New York. 360pp.

Roberts D.A., M. Gardner, R. Church, S. L. Ustin, G. J. Scheer, and R. O. Green. 1998. Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models. Remote Sensing of Environment, 65, 267-279.

Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351-1, 309-317.

Sauder, J., J. B. McMahan, and K. T. Weber. 2003. Fuzzy Classification of Heterogeneous Vegetation in a Complex Arid Ecosystem. African Journal of Range and Forage Science, 20(2):126.

Savage, S. L. 2005. Vegetation Dynamics in Yellowstone’s Northern Range: 1985-1999. MS Thesis, Montana State University. 139pp.

Singh, A. 1989. Review article: digital change detection techniques using remotely sensed data. International Journal of Remote Sensing, 10(6): 989-1003.

Snedecor, G. W., and W. G. Cochran. 1967. Statistical Methods. The Iowa State University Press, Ames, Iowa. 524pp.

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. 1998. Selecting and Interpreting Measures of Thematic Classification Accuracy. Remote Sensing of Environment. 62:77-89.

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

Taylor, B. L. and T. Gerodette. 1993. The Uses of Statistical Power in Conservation Biology: the Vaquita and Northern Spotted Owl. Conservation Biology 7(3):489-500.

Titus, K., J. A. Mosher, and B. K. Williams. 1984. Chance-corrected Classification for use in Discriminant analysis: ecological applications. Am. Midl. Nat. 111:1-7.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46:234–40.

Van Niel, T. G., T. R. McVicar, and B. Datt. 2005. On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification. Remote Sensing of Environment, 98(4) 468-480.

Verbyla, D. L. and S. H. Boles. 2000. Bias in Land Cover Change Estimates Due to Misregistration. International Journal of Remote Sensing. 21(18):3553-3560.

Weber, K. T. 2006. Challenges of Integrating Geospatial Technologies into Rangeland Research and Management. Rangeland Ecology and Management 59:38-43.

[THIS PAGE LEFT BLANK INTENTIONALLY]