A COMPARISON OF INTER-ANALYST DIFFERENCES IN THE CLASSIFICATION OF A LANDSAT ETM+ SCENE IN SOUTH-CENTRAL VIRGINIA

John Iiames, Research Biologist
US Environmental Protection Agency
Office of Research and Development
Research Triangle Park, NC 27612
iiames.john@epa.gov

Dr. Russell G. Congalton, Professor
University of New Hampshire
Dept. of Natural Resources
Durham, NH 03824
russ.congalton@unh.edu

ABSTRACT

This study examined inter-analyst classification variability based on training site signature selection only for six classifications from a 10 km^2 Landsat ETM+ image centered over a highly heterogeneous area in south-central Virginia. Six analysts classified the image at the 30 m ETM+ resolution varying only location and number of training sites. These classifications were then degraded to coarser resolutions, assigning the dominant land cover to the new cell resolution. Analyst-to-analyst differences were noted at the varying scales as well as overall accuracy assessment results compared to a land cover map digitized from an August 3, 2002 Ikonos panchromatic image. Results indicated that highest accuracies for all six analysts occurred at the 450 m scale resolution (i.e. 20.25 ha), corresponding to a 364 m^2 (13.25 ha) average patch size for all classes. Spectral separability for training site data was analyzed for each of the six classifications. These tests included a Euclidean Distance, Transformed Divergence, and Jeffries-Matusita Distance evaluation. All spectral separability tests pointed to areas of class confusion within each interpretation, but prediction of an analyst accuracy ranking based on separability amongst all six interpretations was not achieved. This study was initiated to examine land cover variability between analysts as it applies to the process of creating leaf area index (LAI) surface maps used in the validation of medium resolution LAI products.

Note: The views expressed in these Proceedings are those of the individual authors and do not necessarily reflect the views and policies of the United States Environmental Protection Agency (EPA). Scientists in EPA have prepared the EPA sections, and those sections have been reviewed in accordance with EPA's peer and administrative review policies and approved for presentation and publication.

INTRODUCTION

A number of image processing considerations are required for the extraction of meaningful information from remotely sensed data (Jensen 1996). These processes include image acquisition, exploration (spectral pattern and histogram analysis), preprocessing (radiometric and geometric correction), classification (categorical definitions and classification approach), and accuracy assessment (sampling design). The quality of a classified image is determined by the cumulative error associated with these intermediate steps. Errors incurred along any point of this process may have an additive, multiplicative, or negating effect on the end classification. One important component associated with the supervised classification approach is the selection of training sites for extraction of spectral signatures emanated from specific cover types. High quality (pure) training site selection is vital in the classification process to train the classification algorithm for land cover mapping extended over the entire image. The highest likelihood of pixel membership to a particular cover class is determined by the analysis of the multivariate statistical parameters derived from the representative training sites. By holding all other areas of the classification process constant, this research was designed to test the effect of varying training site selections (location and number) from six analysts performing a supervised classification on a Landsat ETM+ image.

Interest in classification differences of Landsat ETM+ imagery exists in the creation of underlying leaf area index (LAI) surface maps used in the validation of the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI product (MOD15A2). The MODIS LAI product is a 1 km global data product composited over an 8-day period and is derived from a three-dimensional radiative transfer model driven by a atmosphere corrected surface reflectance product (MOD09), a land cover product (MOD12) and ancillary information on surface characteristics. The United States Environmental Protection Agency has initiated validation research in the deciduous evergreen needle leaf biome classification (MOD09), in a regional study located in the southeastern United States investigating MODIS LAI inputs into atmospheric deposition and biogenic modeling processes. The MODIS LAI validation process involves: (1) the collection of ground-based LAI measurements via direct or indirect means, (2) the correlation of these LAI estimates with spectral values retrieved from high resolution imagery (20 m - 30 m), and (3) the aggregation of these high-resolution LAI maps to 1 km spatial resolution, thus matching the resolution of the MODIS product and thereby enabling a comparison between the two LAI values (1 km ETM^+ LAI v. 1 km MODIS LAI) (Morisette et al. 2006). Preliminary findings of analyst variability in the classification process with respect to training site differences and scaling these land cover values to the coarser cell resolutions is presented in this paper.

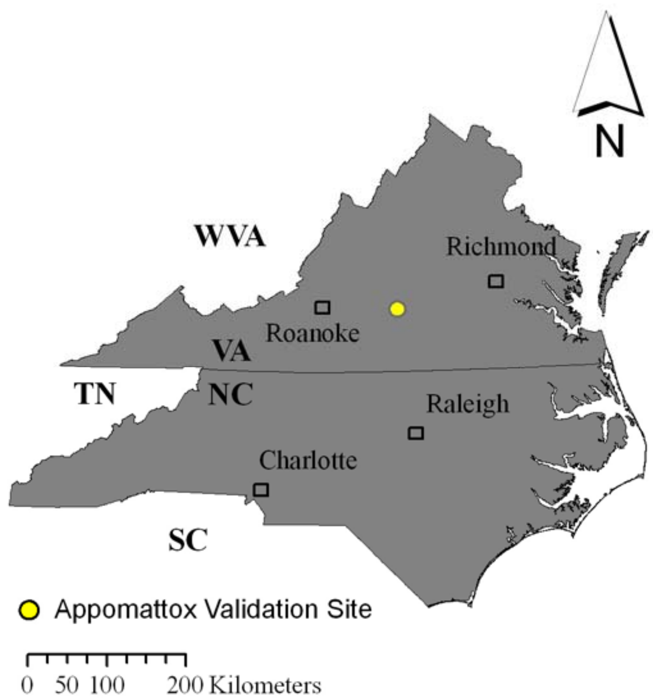

Figure 1. Site location map for LAI validation site in Virginia.

METHODS

Inter-analyst classification variability based on training signature selection was assessed from six analyst-derived classifications generated from a 4-band, 10 km^2 Landsat ETM+ image centered on a site located in south-central Virginia (37.2185^o, 78.8778^0) (Figure 1). Six graduate-level remote sensing students from the Department of Natural Resources within the University of New Hampshire classified this image using Leica Erdas Imagine 8.7 imaging software. In this study, analysts only controlled the training site selection for their classifications. The six 30 m cell-based classifications were scaled-up to varying cell resolutions (150 m – 1200 m) and then were compared to one another in order to assess the level of agreement. An Ikonos reference data set, developed from heads-up digitization of land cover within the area of interest, was likewise scaled to the same range of cell resolutions (150 m-1200 m), then compared to the six analyst classifications to determine overall accuracy values at the particular cell resolutions. The methods and results from the classification and training site assessment are presented in this paper.

Classification

In order to test inter-analyst differences, all classifications differed only with respect to training site location and the number of iterations. Analysts were provided a 4-band (30 m) image to classify using a supervised classification approach implementing the parametric maximum likelihood algorithm. The 4-band image was comprised of original and derived bands from the August 12, 2002 ETM+ image. The bands chosen for analysis included ETM+ bands 5 and 7, SQRT(4/3), and tNDVI.

Training and data was provided to each analyst. Each received the image to be classified, reference images (Ikonos 1.0 m panchromatic and 1.0 m color infrared (CIR) digital orthophoto quarter-quadrangles (DOQQs)), screen captured images in TIFF format depicting specific land cover types, and directions. The land cover classes and definitions are as follows:

- Deciduous Forest – at least 25% of the area is covered by trees and more than 75% of the crown class is deciduous
- Coniferous Forest – at least 25% of the area is covered by trees and more than 75% of the crown class is coniferous
- Mixed Forest - at least 25% of the area is covered by trees and neither conifers or deciduous trees have more than 75% of the crown class
- Water – lakes, ponds, or reservoirs
- Other Vegetation – areas having less than 25% tree coverage and more than 10% other vegetation such as agricultural lands, urban grasses, and transitional vegetation (i.e., clearcut harvests).

Training site signatures were collected for all classes, making certain to attain class invertability. The parametric maximum likelihood algorithm was used to assign class values to each pixel.

Classification Assessment

The six classifications were compared to one another and to reference land cover data sets derived from Ikonos imagery.

Classification-to-Classification Comparison. All classifications were evaluated at eight different resolutions (150 m, 300 m, 450 m, 600 m, 750 m, 900 m, 1050 m, and 1200 m). At each resolution across all six classifications, the percent total agreement, percent agreement between forest and non-forest classes, and the percent agreement of deciduous and coniferous within the forest classes were determined.

Classification-to-Reference Comparison. Overall accuracy was determined for all six classifications for each of the eight resolutions comparing the classified land cover values with the Ikonos reference reference values. The reference data set was a 19.4 km^2 land cover map digitized from an August 3, 2002 Ikonos image. Proportions of the five land cover classes (deciduous, coniferous, mix, water, and other vegetation) were tabulated for all eight resolutions within this reference area. Pixel homogeneity was determined across all resolutions. Error matrices were then constructed only for the 150 m resolution for each of the six classifications comparing the classified land cover values with the Ikonos reference values

Training Site Assessment

Spectral separability for training site data was analyzed for each of the six classifications. These tests included a Euclidean Distance, Transformed Divergence, and Jeffries-Matusita Distance evaluation. The Euclidean Distance evaluation is a statistical measure of the spectral distance between the mean vectors of a signature pair. The Euclidean Distance measurement between signatures did not incorporate measures of variation about the mean, but did give a rough estimate of spectral separation. The Transformed Divergence analysis exhibited computationally simpler empirical measurements than the Jeffries-Matusita analysis. The Jeffries-Matusita analysis displays a more theoretical soundness than the Transformed Divergence in that it directly related to the upper bounds of classification errors (Richards, 1986). Signatures were defined as totally separable if the calculated divergence was equal to the appropriate upper bound.

Along with the above mentioned divergence measures of separability, bi-spectral box plots were constructed for each operator classification to investigate overlap percentages between classes. For all pair-wise band combinations which included: TM 5-TM7; TM 5-SQRT(4/3); TM 5-tNDVI; TM 7-SQRT(4/3); TM 7-tNDVI; SQRT(4/3)-tNDVI, bi-spectral box plots were created from variability measured within each training signature averaged overall

all training sites for a particular land cover class. This variability was added to the variability of the training site means across all training sites for a particular land cover class (Figure 2)

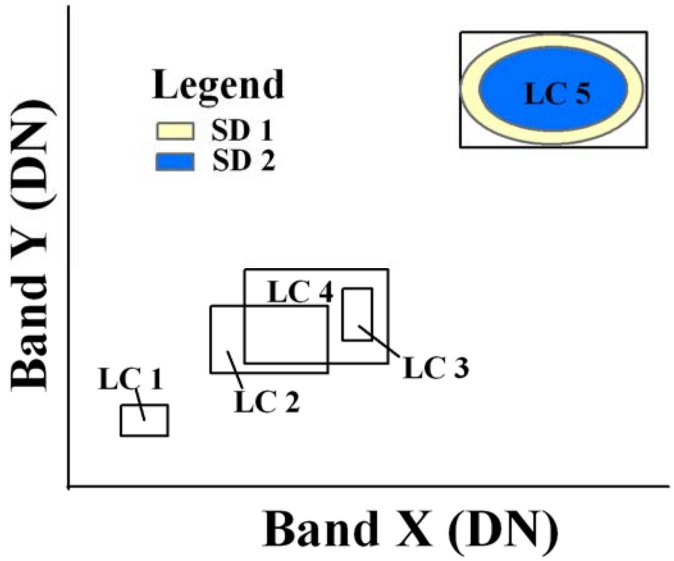

Figure 2. Calculation of the percentage of training site overlap. Each of the five land cover (LC) box plots are created from variance associated within each independent training site (SD 1) and variance measured across all training sites (SD 2). To arrive at a value SD 1, standard deviations were acquired for each independent training site within a given land cover class. These standard deviations were then averaged to give an overall mean value, resulting in the added variability visualized by the outer yellow ring in land cover 5. The majority of variability is found in SD 2, which was a measurement of the standard deviation of the means of each training site contributing to the land cover 5 classification.

PRELIMINARY RESULTS

Classification-To-Reference Assessment

Reference Data Set Description. The 19.4 km^2 digitized Ikonos imagery provided reference land cover data sets at the eight resolutions for approximately 20% of the 100 km^2 area covered by the Landsat ETM+ subset scene (Table 2). Degrading the cell resolution over the 150 m to 1200 m range, the percent land cover changed from a fairly even distribution between the deciduous (33.7%), coniferous (35.3%), and other vegetation (21.7%) at the 150 m resolution to only occurrences of deciduous (50%) and coniferous (50%) land cover classes at the coarsest resolution (1200 m) (Figure 3). Overall and class percent homogeneity were recorded for each cell resolution (90 m cell resolution included in this analysis). Cells of 100% one land cover class decreased from 54.7% at the finer resolution (150 m) cell size to 0% at the 600 m cell resolution (Figure 4). From 90 m to 600 m, the percent of cells with one class type between 75 and 99 percent of the total averaged 28.4%, after which dropped to near-zero percent at the coarser cell resolutions (> 750 m). There was a general increase of cells having one class type at least 51%, but less than 75% over decreasing cell resolution up to 750 m, then a decrease at the coarsest resolution (1050 m).

Table 2. Land cover reference data sets by cell resolution

	Cell Resolution (m)							
	150	300	450	600	750	900	1050	1200
Cell Area (Ha)	2.25	9	20.3	36	56.3	81	110.3	144.0
# Cells Analyzed	836	180	84	32	21	16	15	8
Reference Area (Km^2)	18.8	16.2	17	11.5	11.8	12.2	16.5	11.5
% LC 1 (Deciduous)	34.8	35.2	34.2	36.6	35.5	34.4	35.0	36.6
% LC 2 (Coniferous)	35.5	35.7	36.6	39.7	39.6	42.3	35.2	39.6
% LC 3 (Mix)	8.6	8.5	8.7	9.5	9.5	8.5	8.5	9.5
% LC 4 (Water)	< 1.0	1.3	1.3	1.7	1.8	1.7	1.3	1.8
% LC 5 (Other Vegetation)	21.2	19.3	19.2	12.5	13.5	13.1	20.0	12.5

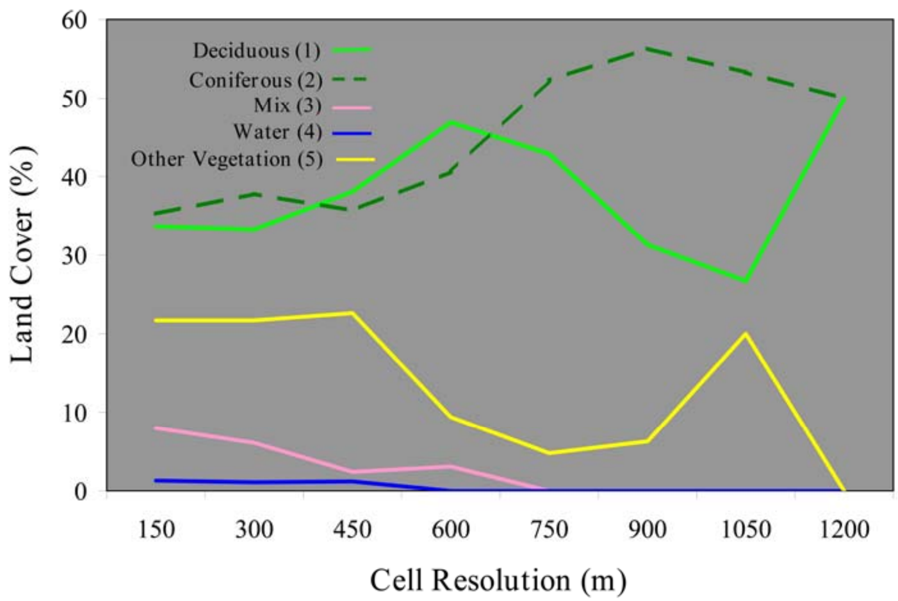

Figure 3. Land cover variability across eight resolutions.

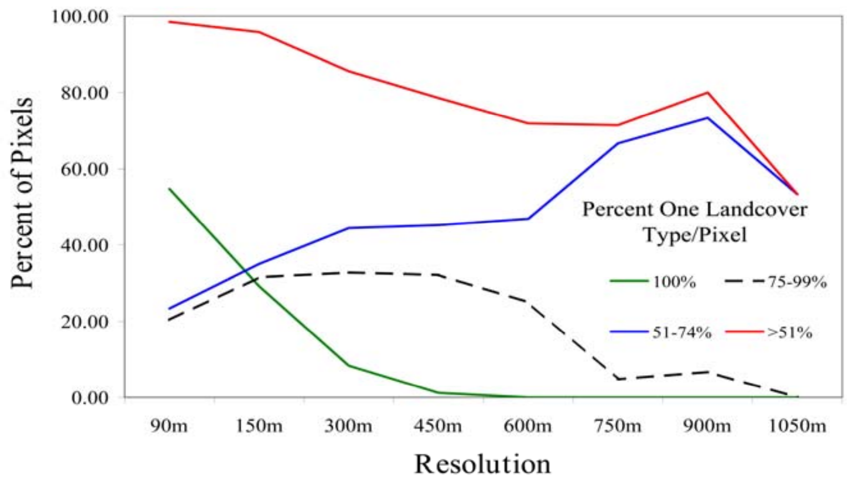

Figure 4. Land cover heterogeneity of multi-resolution analysis.

Assessment. Pixel-to-pixel comparisons were completed at cell resolutions greater than 150 m. Error matrices were developed for each analyst for the 150 m cell resolution not for cell resolutions greater than 150 m due to the low reference sample number per class. A summary of overall accuracies across all cell resolutions is found in Table 4. Highest accuracies from four out of the six analysts were achieved at the 450 m cell resolution. Also, cell by cell agreement between analysts increased with increased resolution (i.e. smaller cell size) (Figure 4).

Table 4. Overall classification accuracy results for six analysts across 8 cell resolutions (150 m – 1200 m) as compared to 19.4 km^2 reference classification.

Cell Size (m)	Analyst 1	Analyst 2	Analyst 3	Analyst 4	Analyst 5	Analyst 6
1200	37.5	37.5	75.0	37.5	50.0	75.0
1050	40.0	53.3	60.0	40.0	60.0	86.7
900	20.0	66.7	46.7	33.3	53.3	66.7
750	42.9	42.9	57.1	33.3	61.9	66.7
600	46.9	65.6	68.8	43.8	59.4	75.0
450	57.1	69.0	76.2	54.8	75.0	79.8
300	58.3	65.0	69.4	50.6	64.4	73.3
150	57.5	64.8	67.7	53.9	63.6	70.7

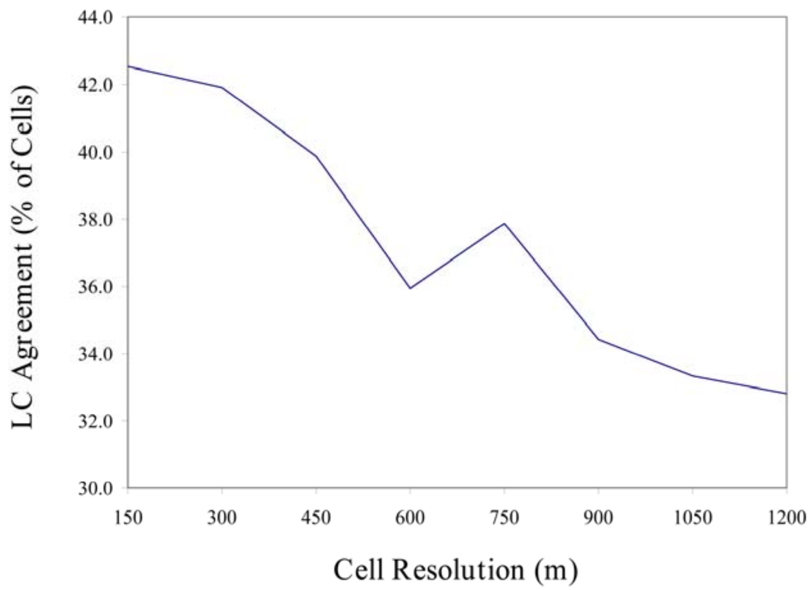

Figure 4. Percent overall agreement between six classifications across varying scales.

Error matrices from the 150 m assessment depicted a range of accuracies from 53.9% to 70.7% (Table 4). Across classes 1 and 2, omission errors exceeded commission errors, yet the opposite occurred with both error types in classes 4 and 5 (Figure 5). For example, class 4 (water) had an omission error of 27% with no commission error for all analysts, as would be expected. As expected, the majority of the omission errors for all analysts occurred within the forested classes (deciduous, coniferous, or mix). The kappa analysis revealed that five out of the six analyst derived classifications displayed moderate agreement between the reference data and the classification (Landis and Koch 1977) (Table 5).

Table 5. Kappa analysis of all six analyst-derived classifications

Analyst	# of 150 m cells	# of cells analyzed	Correct	Accuracy	Kappa	Agreement
1	836	836	481	57.5	0.46	Moderate
2	836	836	542	64.8	0.54	Moderate
3	836	836	566	67.7	0.57	Moderate
4	836	836	451	53.9	0.39	Poor
5	836	836	532	63.6	0.51	Moderate
6	836	836	591	70.7	0.59	Moderate

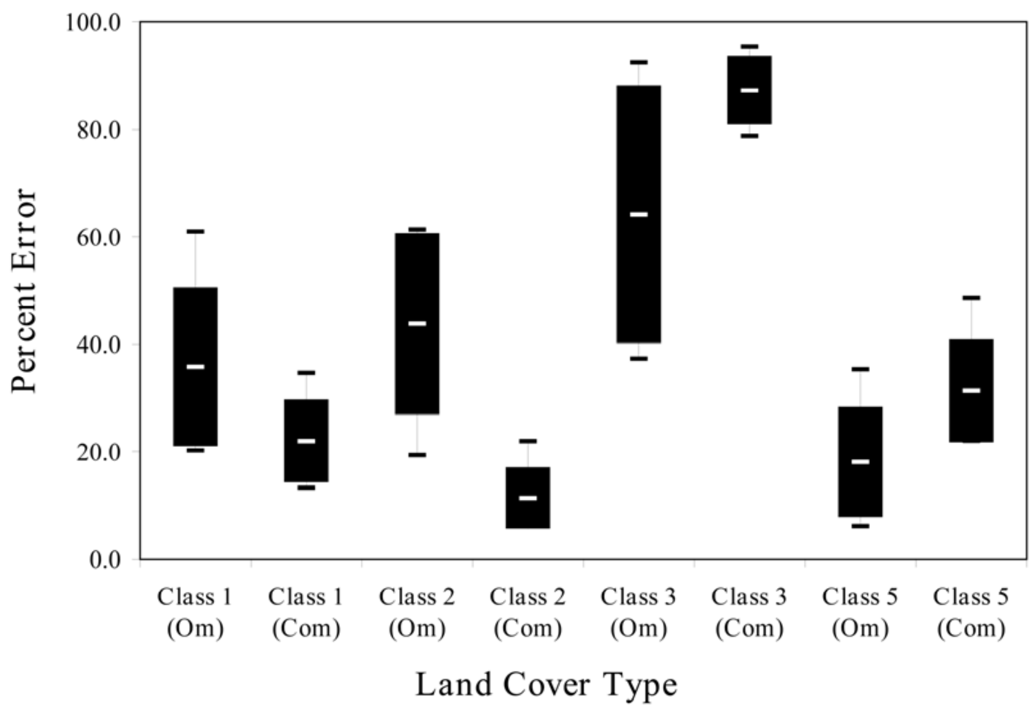

Figure 5. Omission and Commission error for all analyst derived classifications across all land cover classes at the 150 m resolution.

Training Site Assessment.

Classification differences between analysts can be attributed to training site selection for all five classes. The six analysts chose between 27 and 60 training sites with a minimum of 5 training sites per class. Iterations varied between 3 and 5 for all analysts. Across all analysts, spectral pattern analysis revealed that the water and other vegetation classes were widely separated from the each other and the forested classes across all 4 bands. However, the forested classes (deciduous, coniferous, mix) displayed overlap and very little separation across all 4 bands. Separability tests were inconclusive in the prediction of the best classification. However, the lowest accuracy classification was identified with all thee tests (Euclidean, Transformed Divergence, and Jeffries-Matusita. The percent overlap analysis did detect the top four classifications with a minimum overlap ranging between 4.4% and 20.4%.

DISCUSSION

The range of classification differences in this study was 16.8% at the 150 m cell resolution. The lowest overall accuracy was 53.9% (analyst 4) compared to the highest accuracy of 70.7% (analyst 6). The combination of analyst limited experience and not having other post-processing options available to them resulted in low overall accuracies. An example of one area of large confusion in the training site selection was found with the analysis from analyst 4. A significant number of pixels (35.2%) were committed by analyst 4 to the “other vegetation” (LC 5) category. However, the reference data set showed 21.2% attributed to this LC 5 category. The impact of this error could be significant in use of this data for estimating forest biometric parameters, for example. One parameter, leaf area index (LAI) ranges between 1.5 and 3.0 for the deciduous and coniferous land cover type, whereas the “other vegetation” land cover type typically averages below an LAI value of 1.0. Committing forested pixels to other vegetation pixels would certainly negatively skew the resulting LAI surface map.

Examination of the training site information for all six analysts revealed that class confusion could be predicted using all the separability measures, however gradations of accuracy could not be determined. Also, training site confusion between classes could not be further segmented into potential omission and commission errors. It was beneficial to utilize all separibilty measures for determination of the best and worst classification.

ACKNOWLEDGEMENTS

The authors would like to thank the analysts from UNH who provided their time and effort to produce these classifications: Jennifer Bourgeault, Jesse Bishop, Chris Czarnecki, Shawn Herrick, and Tina Cormier.

REFERENCES

Jensen, J.R. (1996). Introductory digital image processing: a remote sensing perspective, Prentice-Hall, Inc., Upper Saddle River, NJ

Landis, J., and G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics. Vol 33, pp. 159-174.

Morisette, J.T., Baret, F., Privette, J.L., Myneni, R.B., Nickeson, J., Garrigues, S., Shabanov, N., Weiss, M., Fernandes, R., Leblanc, S., Kalacska, M., Sánchez-Azofeifa, G., Chubey, Rivard, M.B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A., Lewis, T., Iiames, J., Colombo, R., Meroni, M., Busetto, L., Cohen, W., Turner, D., Warner, E., Petersen, G.W., Seufert, G., and R. Cook (2006). Validation of global moderate-resolution LAI Products: a framework proposed within the CEOS Land Product Validation subgroup. *IEEE Transactions of Geoscience and Remote Sensing*, (In Press)

Richards, J. A., Remote Sensing Digital Image Analysis: An Introduction, Springer-Verlag, Berlin, 1986.