ACCURACY ASSESSMENT OF LAND COVER MAPS DERIVED FROM MULTIPLE DATA SOURCES

Hilary Tribby, Research Assistant
Dr. Daniel R. Unger, Associate Professor
Dr. I-Kuai Hung, Assistant Professor
Dr. Hans Williams, Professor
Arthur Temple College of Forestry and Agriculture
Stephen F. Austin State University
Nacogdoches, Texas
unger@sfasu.edu

ABSTRACT

Maximum Likelihood (ML) and Artificial Neural Network (ANN) supervised classification methods were used to demarcate land cover types within IKONOS and Landsat ETM+ imagery. Three additional data sources were integrated into the classification process: Canopy Height Model (CHM), Digital Terrain Model (DTM) and Thermal data. Both the CHM and DTM were derived from multiple return small footprint LIDAR. Forty maps were created and assessed for overall map accuracy, user's accuracy, producer's accuracy, kappa statistic and Z statistic using classification schemes from U.S.G.S. 1976 levels 1 and 2 and T.G.I.C. 1999 levels 2 and 4. Results for overall accuracy of land cover maps derived from multiple sources ranged from 13.67 to 57.56 percent for U.S.G.S. level 2 and T.G.I.C. level 4 across ML and ANN classifications. Results for overall map accuracy ranged from 26.00 to 72.33 percent for U.S.G.S. level 1 and T.G.I.C. level 2 across ML and ANN classifications. Land cover maps, derived using ML classification methodology, were consistently more accurate than land cover maps derived using an ANN classification algorithm.

INTRODUCTION

Land cover maps derived from satellite imagery have been an important component of natural resource monitoring and management since the launch of the first Earth Resources Technology Satellite (later renamed Landsat) in 1972. Land cover classification enables researchers and managers to assess landscape components, track landscape changes over time and monitor wildlife habitat. Increasing the accuracy at which these classifications can be performed will provide decision-makers with more reliable data.

Numerous classification schemes have been developed for use with land cover and land use classifications. Early in satellite remote sensing Anderson and his colleagues (1976) developed and published a rubric by which to define dominant cover. Since Anderson et al. published this system for continental land cover classification; other regions have taken the principles set forth by the Anderson group and created more specific cover type classes. This is evident in the Interagency Land Use Land Cover Working Group (1999) publication where a classification scheme established particular classes for use in Texas at varying levels for the advent of high spatial resolution imagery.

Often a land cover map is produced solely from multispectral imagery, which is the traditional classification method. The neural network, an alternative to traditional classification methodology, easily incorporates multiple sources of data at varying spatial resolutions. Although important information can be extracted from a single satellite image, the inclusion of multiple sources and types of data may be best for analysis.

The purpose of this project was to evaluate ML and ANN classification methodologies and recommend which data sources should be integrated into image classifications in order to produce the most accurate land cover map. These results will aid decision-makers in determining what data should be included into a given land cover classification to provide the most accurate information possible.

BACKGROUND

Land cover classifications, typically derived from remotely sensed satellite imagery such as Landsat and IKONOS, are used to display the dominant regional cover type and provide foundational material for use in other analyses (Campbell, 2002). The accuracy of any derived land cover map is vital to managers since inaccurate data are harmful because decisions based on poor quality data are hazardous to the objectives of any mapping project (Congalton & Green, 1999).

LIDAR, an acronym for light detection and ranging, can be used to extract height and elevation data of the landscape's physical attributes. Forest characteristics such as canopy height, basal area and total above ground biomass were accurately predicted with large footprint LIDAR in the western United States (Means, Acker, Harding, Blair, Lefsky, Cohen & et al, 1999). In central Idaho a LIDAR derived canopy height model was used to model single and multi-story forest canopy with an overall accuracy of 97% and a kappa statistic of 0.89 (Zimble, Evans, Carlson, Parker, Grado & Gerard, 2003).

Since thermal infrared (TIR) data are captured at a different spatial resolution than the other Landsat ETM+ bands, researchers often discard the band entirely when performing an image classification (Linderman, Liu, Qi, An, Ouyang, Yang & et al., 2004; Lo & Choi, 2004). Since data contained in thermal bands are uncorrelated to the remaining Landsat ETM+ bands, thermal data should allow researchers the ability to more clearly delineate pixels that are not as easily classified due to confusion of the reflected electromagnetic energy. Price (1981) used the Landsat MSS thermal band to increase classification accuracy in the southeastern region of New York State but the TIR data were only useful in the separation of two classes: water and clouds. Price noted that a sensor problem and coarse spatial resolution may have accounted for the limited increase in image accuracy.

There are a number of schemes for class division that began with James Anderson and his need to create uniformity in a new unregulated science (Anderson, Hardy, Roach & Witmer, 1976). Although a standard scheme has been created, many scientists have ventured away from the rubric by creating particularized classes that are used for specific regions or instances. The late 1990's produced resurgence for the standardization of class categories in Texas. In 1999 the "Texas Land Classification System" was developed by the Texas Geographic Information Council's (TGIC) Interagency Land Use Land Cover Working Group (1999). TGIC's goal was to coordinate the development of land use land cover classification standards between federal, state and local entities. The group established three additional classification levels of vegetation that expanded the categories put forth by Anderson et al. in 1976.

Land cover classification methodologies, which have been well documented in the literature, typically follow either an unsupervised or supervised classification approach using a single source of digital imagery (Campbell, 2002; Jensen, 1996). In recent years, the artificial neural network of classification has increased in popularity since the method has the ability to produce better and faster results than classical statistical means as well as incorporate different sources of data (Atkinson & Tatnall, 1997; Gong, 1996). Even though many researchers have called the ANN the "black box" of remote sensing, it has a powerful process that can accomplish, with greater accuracy results, that which other classification methods cannot (Davalo & Naïm, 1991; Qui & Jensen, 2004).

METHODS

This project assessed the accuracy of land cover maps derived from Maximum Likelihood and Artificial Neural Network classification methodologies within the Forest Lake area of East Texas with multiple data sources and varying classification schemes (Figure 1). A Landsat scene of the study area, with an image acquisition date of July 21, 2000, was acquired from the Forest Resources Institute at Stephen F. Austin State University. An IKONOS image of the study area, with an image acquisition date of July 20, 2000, was purchased from Space Imaging, LLC. Once acquired each image was radiometrically corrected to acquire pure spectral signatures. Geometric validation of each image was verified using 25 randomly located points within the study area to visually ascertain individual pixel location agreement.

The data reduction technique of Principal Component Analysis (PCA) was applied to the Landsat ETM+ and IKONOS data. Duda and Canty (2002) found that reducing extraneous information contained within multispectral imagery using the PCA method produced more accurate results. In addition to creating PCA imagery of the study

Figure 1. Location of study site.

area, thermal data were extracted from the Landsat ETM+ data to derive a thermal spectral signature of the Forest Lake area. LIDAR data of the study area, acquired in three phases during June 2000, were used to create canopy height (CHM) and digital terrain models (DTM) of the Forest Lake region.

Forty land cover maps were generated in four phases; 5 maps per ML or ANN classification methodology within each U.S.G.S. levels 1 and 2 and T.G.I.C. levels 2 and 4 classification scheme (Figure 2). Data combinations used to generate each of the five maps were: PCA, PCA-CHM, PCA-DTM, PCA-Thermal and PCA-CHM-DTM-Thermal for all eight groups of five maps respectively.

T.G.I.C. level 4, which consists of 17 classes, includes tall grass, swamp, shrub wetland, short grass, river, mixed woodland, mixed shrub, mixed forest, marsh, lake, forest wetland, evergreen woodland, evergreen shrub, evergreen forest, deciduous woodland, deciduous shrub and deciduous forest. T.G.I.C. level 2, which consists of 7 classes, includes woody wetland, woodland, water, shrub land, natural herbaceous, forestland and emergent wetland.

U.S.G.S. level 2, which consists of 10 classes, includes wetland non-forest, wetland forest, shrub and brush, river, mixed rangeland, mixed forest, lake, herbaceous, evergreen forest and deciduous forest. U.S.G.S. level 1, which consists of 4 classes, consists of wetland, water, rangeland and forest.

Land cover map accuracy was performed by comparing 900 pixels in each image with corresponding high spatial resolution QuickBird imagery with an image acquisition date of November 24, 2002. Due to the historical nature of the study, and hence the inability to physically visit each random point in the field, high spatial resolution QuickBird imagery was the only imagery available to assess the accuracy of the derived land cover maps. The verification sites, which were located using stratified random sampling, were then validated against the remotely sensed images using a traditional error matrix (Jensen, 1996). Overall map accuracy, user's accuracy, producer's accuracy, kappa statistic and Z statistic were calculated for each error matrix.

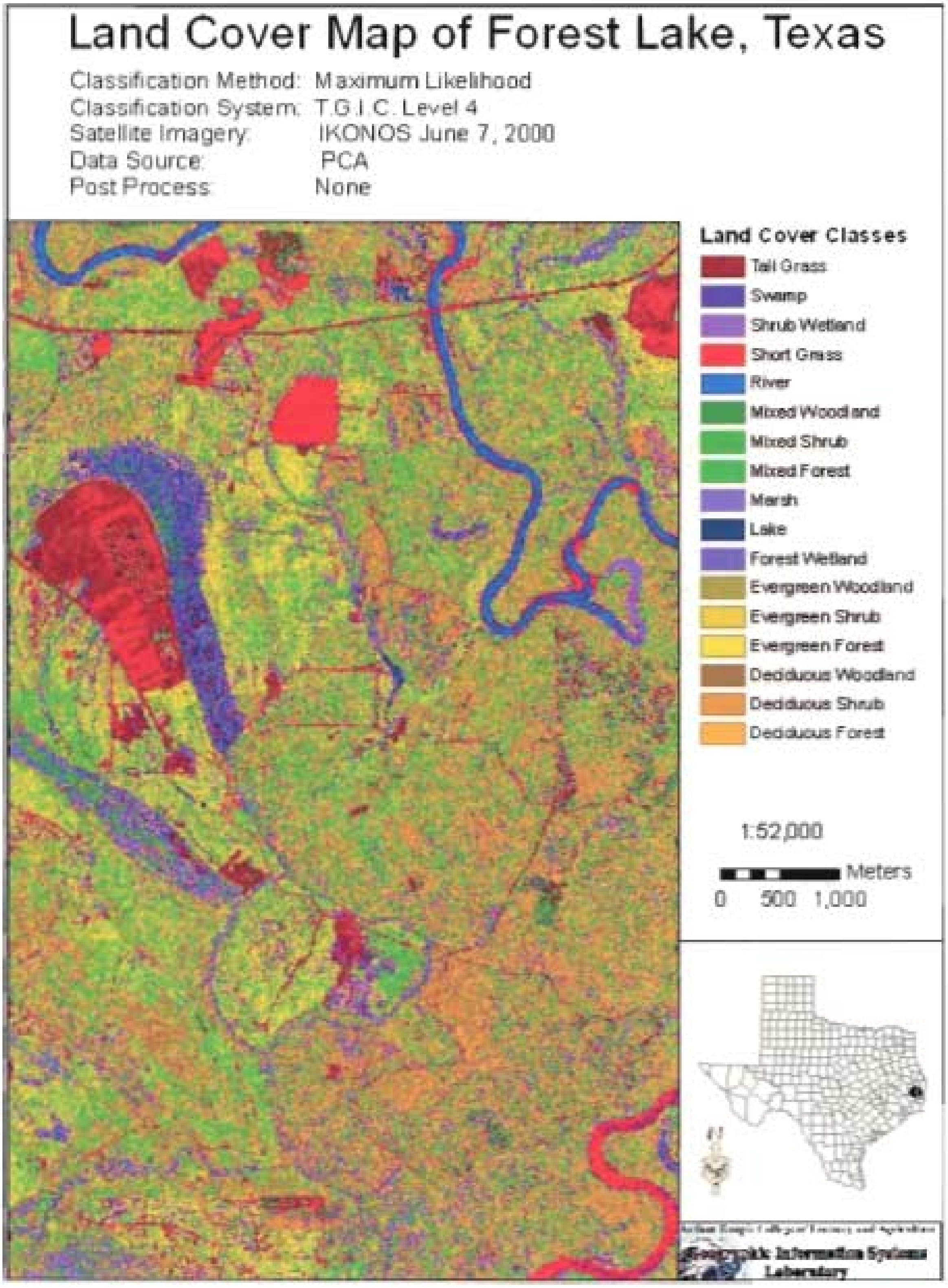


Figure 2. Land cover map example.

RESULTS

The land cover maps produced using IKONOS data with T.G.I.C. level 4 have extremely low overall map accuracy ranging from 13.67 to 36.89 percent (Figure 3). Land cover maps produced using Landsat ETM+ data at U.S.G.S. level 2 have higher levels of overall map accuracy, ranging from 30.33 to 57.56 percent, but are considerably lower than acceptable levels of overall map accuracy for any derived land cover map product (Figure 4).

After decreasing the T.G.I.C. classification scheme from level 4 to 2 for the IKONOS data, subsequently reducing the number of classes from seventeen to seven, overall map accuracy increased and ranged from 29.40 to 57.22 percent (Figure 5). This was an increase in overall map accuracy from T.G.I.C. level 4 but lower than acceptable levels of overall map accuracy. The PCA-CHM was the highest performing ML classification and some areas of confusion may have been decreased by the high resolution canopy height model. The ANN classification method did not produce such positive results. The most accurate ANN contained the PCA-CHM-DTM-Thermal data and had an overall map accuracy of 40.98 percent. The results, which were consistent between T.G.I.C. levels 4 and 2 using IKONOS data, indicate that the lowest performing maps were with the ANN classification.

After decreasing the U.S.G.S. classification scheme from level 2 to 1 for the Landsat ETM+ data, overall map accuracy increased and ranged from 26.00 to 72.33 percent (Figure 6). Although in one particular case, the map derived using PCA-Thermal data with the ANN algorithm actually decreased in overall map accuracy indicating that areas of confusion were not diminished with the broad definition of classes but may have increased. The highest performing ML land cover map at 72.33 percent overall map accuracy was created using all data combined (PCA-CHM-DTM-Thermal).

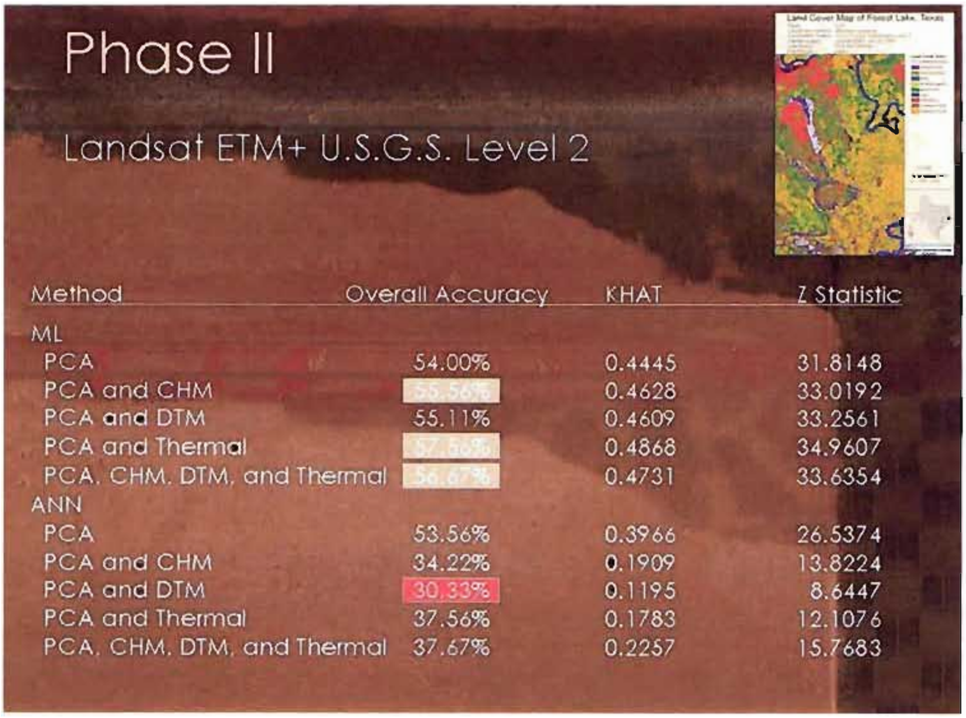
When comparing classification methodologies, the ML classifications consistently produced the most accurate land cover map compared to an ANN algorithm. The least accurate land cover map per set of five maps within each classification scheme was always an ANN produced land cover product. When comparing overall land cover map accuracy between image data sources, the midspatial resolution data of Landsat ETM+ consistently performed better than the high spatial resolution data of IKONOS.

Phase I

IKONOS T.G.I.C. Level 4

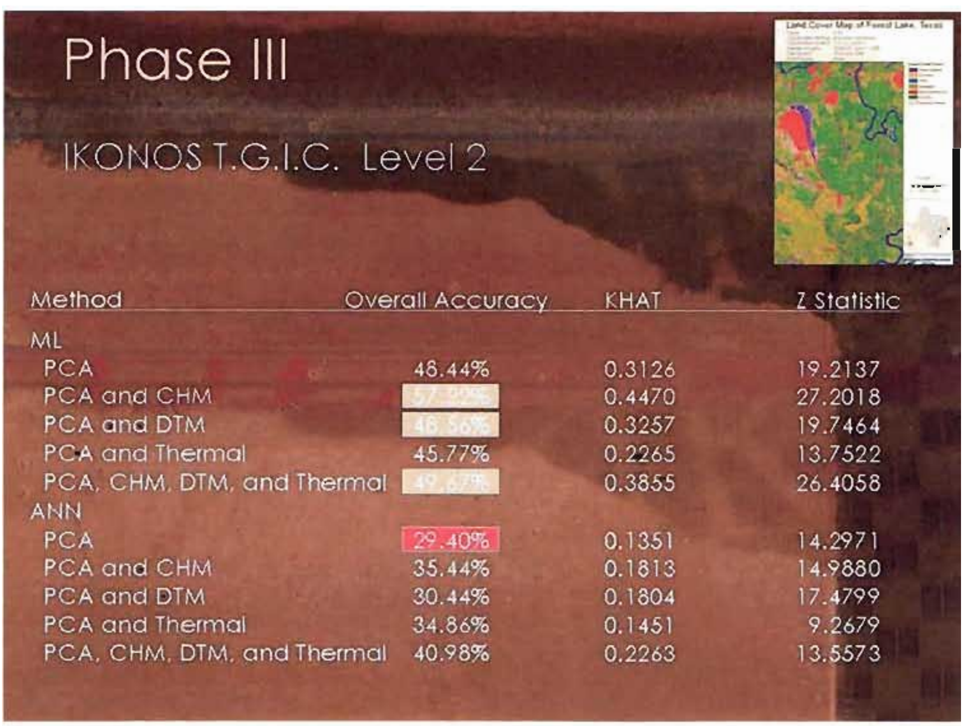
Method	Overall Accuracy	KHAT	Z Statistic
ML			
PCA	36.89%	0.3038	26.5027
PCA and CHM	20.89%	0.1503	17.2351
PCA and DTM	19.44%	0.1319	14.6103
PCA and Thermal	35.78%	0.2930	28.7369
PCA, CHM, DTM, and Thermal	22.67%	0.1503	15.1933
ANN			
PCA	23.67%	0.1891	24.6298
PCA and CHM	28.22%	0.1509	13.9583
PCA and DTM	17.89%	0.1192	15.2786
PCA and Thermal	34.00%	0.2670	26.0242
PCA, CHM, DTM, and Thermal	13.67%	0.0876	12.5116

Figure 3. IKONOS T.G.I.C. Level 4 results.



Method	Overall Accuracy	KHAT	Z Statistic
ML			
PCA	54.00%	0.4445	31.8148
PCA and CHM	55.56%	0.4628	33.0192
PCA and DTM	55.11%	0.4609	33.2561
PCA and Thermal	57.56%	0.4868	34.9607
PCA, CHM, DTM, and Thermal	56.67%	0.4731	33.6354
ANN			
PCA	53.56%	0.3966	26.5374
PCA and CHM	34.22%	0.1909	13.8224
PCA and DTM	30.33%	0.1195	8.6447
PCA and Thermal	37.56%	0.1783	12.1076
PCA, CHM, DTM, and Thermal	37.67%	0.2257	15.7683

Figure 4. Landsat ETM+ U.S.G.S. Level 2 results.



Method	Overall Accuracy	KHAT	Z Statistic
ML			
PCA	46.44%	0.3126	19.2137
PCA and CHM	57.22%	0.4470	27.2018
PCA and DTM	46.56%	0.3257	19.7464
PCA and Thermal	45.77%	0.2265	13.7522
PCA, CHM, DTM, and Thermal	49.67%	0.3855	26.4058
ANN			
PCA	29.40%	0.1351	14.2971
PCA and CHM	35.44%	0.1813	14.9880
PCA and DTM	30.44%	0.1804	17.4799
PCA and Thermal	34.86%	0.1451	9.2679
PCA, CHM, DTM, and Thermal	40.98%	0.2263	13.5573

Figure 5. IKONOS T.G.I.C. Level 2 results.

Phase IV

Landsat ETM+ U.S.G.S. Level 1

Method	Overall Accuracy	KHAT	Z Statistic
ML			
PCA	68.67%	0.4134	19.3566
PCA and CHM	70.44%	0.4443	20.4265
PCA and DTM	72.00%	0.4840	20.8969
PCA and Thermal	69.67%	0.4355	20.0346
PCA, CHM, DTM, and Thermal	72.33%	0.4864	20.6836
ANN			
PCA	64.89%	0.4328	20.1101
PCA and CHM	42.89%	0.1984	11.1062
PCA and DTM	39.56%	0.0813	4.3250
PCA and Thermal	26.00%	0.0828	6.3673
PCA, CHM, DTM, and Thermal	64.33%	0.3287	12.1545

Figure 6. Landsat ETM+ U.S.G.S. Level I results.

CONCLUSIONS

The results show that the addition of multiple sources of data did not habitually produce results that are better than traditional multispectral image classifications. The most noticeable result was the significant difference between the performances of both image classification methods. The Maximum Likelihood method, when comparing overall map accuracy, consistently produced a more accurate land cover product than an ANN algorithm. There were isolated instances where the ANN performed as well as the Maximum Likelihood method, but not regularly.

The classification scheme was another important factor when considering the results. The land cover maps produced at lower, or more general levels of classification, were more accurate than those maps produced at higher levels of analysis.

In general, maps created from Landsat ETM+ imagery within the Maximum Likelihood classification method using both U.S.G.S. level 1 and 2 were most accurate. Maps created from IKONOS imagery within the Artificial Neural Network classification method using both T.G.I.C. level 4 and 2 were least accurate, possibly due to increased pixel size, shadow confusion, and the normally distributed data that performs better within a Maximum Likelihood classification.

Future studies should look at vertical forest structure within the classification process since the vertical arrangement of vegetation could aid in the discrimination of different yet spectrally very similar land cover categories. Other classification methods should be used as well, i.e. fuzzy, decision-tree, and mixed methods. One might also look at the classification accuracies of data across software programs to ascertain possible differences in software algorithms.

REFERENCES

Anderson, J. R., Hardy, E. E., Roach, J. T. & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data (Geological survey professional paper 964). Washington, DC: U.S. Government Printing Office.

Atkinson, P. M. & Tatnall, A. R. L. (1997). Neural networks in remote sensing. *Int. J. Rem. Sens., 18,* 699-709.

Campbell, J. B. (2002). *Introduction to remote sensing* (3rd ed.). New York: Guilford Press.

Conglaton, R.G. & Green, K. (1999). *Assessing the accuracy of remotely sensed data: Principles and practices.* Boca Raton, FL: Lewis Pubs.

Davalo, E. & Naïm, P. (1991). *Neural networks* (A. Rawsthorne, trans.). London: Macmillan Press, Ltd.

Duda, T. & Canty, M. (2002). Unsupervised classification of satellite imagery: Choosing a good algorithm. *Int. J. Rem. Sens., 23,* 2193-2212.

Gong, P. (1996). Integrated analysis of spatial data from multiple sources: Using evidential reasoning and an artificial neural network for geological mapping. *Photo. Eng. & Rem. Sens., 62,* 513-523.

Texas Geographic Information Councils, Interagency LULC Working Group (1999). *Texas land classification system: Recommendations for new land use land cover datasets for Texas.* Available at: www.dir.state.tx.us/tgic/pubs/lulcreport. pdf Retrieved on March 17, 2005.

Jensen, J. R. (1996). *Introductory digital image processing: A remote sensing perspective* (2nd ed). Upper Saddle River, NJ: Prentice Hall.

Linderman, M., Liu, J., Qi, J., An, L., Ouyang, Z., Yang, J. & et al. (2004). Using artificial neural networks to map the spatial distribution of understory bamboo from remote sensing data. *Int. J. Rem. Sens., 25,* 1685-1700.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B. & et al. (1999). Use of large-footprint scanning airborne LIDAR to estimate forest stand characteristics in the Western Cascades of Oregon. *Rem. Sens. Env., 67,* 298-308.

Price, J. C. (1981). The contribution of thermal data in Landsat multispectral classification. *Photo. Eng. & Rem. Sens., 47,* 229-236.

Qui, F. & Jensen, J. R. (2004). Opening the black box of neural networks for remote sensing classification. *Int. J. Rem. Sens., 25,* 1749-1768.

Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C. & Gerard, P. D. (2003). Characterizing vertical forest structure using smallfootprint airborne LIDAR. *Rem. Sens. Env., 87,* 171-182.