PREDICTING SPATIAL DISTRIBUTION OF PRIVET (*LIGUSTRUM SPP.*) IN SOUTH CAROLINA FROM MODIS AND FOREST INVENTORY PLOT DATA

Dumitru Salajanu, Forester/Remote Sensing
Dennis M. Jacobs, Research Forester
USDA Forest Service
Forest Inventory and Analysis
4700 Old Kingston Pike
Knoxville, TN 37919
dsalajanu@fs.fed.us
djacobs@fs.fed.us

ABSTRACT

Privet's aggressive competitive behavior causes environmental harm to the ecosystem by degrading species diversity and wildlife habitat. Effective control of its spread requires high-quality spatial distribution information. Our objective in this study was to evaluate the capability and reliability of using satellite imagery, ancillary, and forest inventory data to predict and map the spatial distribution of privet (*Ligustrum* spp); an invasive species. Two types of trials (one for forested area and the other for the entire state's area), and two datasets for each trial were used to model privet spatial distribution within forest. The first dataset in each trial has all forest plots with privet coded as present or not present, while in the second, each forest plot with privet has an assignment to one of four classes, based on the proportion of privet in the plot. Separately, each dataset was used in a decision tree classification process applied to MODIS satellite data (250-meter resolution) and ancillary data to classify and model privet spatial distribution. Classification results from the first trial (forested area) show a decrease in overall classification accuracy from 84.73 percent, when plots are coded for presence or non-presence of privet, to 78.64 percent, when each privet plot uses categorical codes based on field calls for average privet proportion present on the plot. The decrease in accuracy is the result of a combination of assigning privet classes and an insufficient number of plots in classes containing greater than 10 percent privet. Overall classification accuracy of 72.30 percent in the second trial decreased to 69.39 percent when the privet-presence class was expanded by using the field-call categories. The greatest increase in accuracy is exhibited when comparing classifications of the entire state area (69.39 percent accuracy) with the state's forested area (78.64 percent accuracy). In both trials, plots with privet were divided into four classes, based on the proportion of privet. The final output provides information on spatial distribution of privet in forest areas and at the state level. To improve model performance, there needs to be an increase in the number of forest plots containing greater than 10 percent privet cover.

INTRODUCTION

Introduced into the United States around 1800 as an ornamental shrub, privet became one of the most widespread invasive species in the Southern States. Due to aggressive competitive behavior, many land-cover types (forestland, grassland, pasture, water, etc.) are affected by invasive species; from the lower plain to the high mountains. In addition, invasive species cause economic and environmental harm to the ecosystem by degrading species diversity, displacing native vegetation, and producing changes in ground flora and wildlife habitat. Effective control of its spread requires good spatial distribution information**.**

Satellite remote sensing is an important tool for surveying vast areas of forestland and has been used in addressing the urban-rural interface, e.g., Dwyer and others (2001). Recent improvements in classification algorithms allow ancillary data, e.g., soils, average maximum and minimum temperatures, to be incorporated and used successfully to depict forest and biomass from widely spaced ground observations (Blackard et al 2008). Over the years, forest type classifications have been derived from an assortment of satellite data sensors with a variety of spatial and spectral resolutions. Results varied according to the classification algorithm, site location, number of classes used in each classification, etc.

Improvements in technology and classification algorithms allow ancillary data (slope, soil type, vegetation indices, merged information from sensors with different resolutions, etc.) to be incorporated into the original satellite data as new channels. Classification accuracy improved when the original spectral channels were combined with ancillary data as additional channels in the classification process (Ricchetti, 2000; Chavez, 1986; Borry et al, 1990; Pellemans et al, 1993; Vogelmann et al 1998; Salajanu and Olson, 2001). In the last several years, classification and regression tree analysis have been implemented in several software programs and were used in many remote sensing applications (Huang and Jensen, 1997; Lawrence and Wright, 2001; Cooke and Jacobs, 2005).

The inventory design of the Forest Inventory and Analysis National Program of the United States Department of Agriculture Forest Service (FIA) requires annual measurements on a portion of all land in order to form rotating panels. The Southern States strive for a 20% sample each year as part of a 5-year cycle. The panels of forest inventory plot data are used as ancillary information when classifying forest types and forest area, as well as modeling invasive species.

The main objective of this study is to evaluate the capability and reliability of combining Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery, ancillary data and forest inventory data, to predict and map the spatial distribution of privet (*Ligustrum* spp.); an invasive species primarily used in landscaping.

STUDY AREA

The test area for this study includes different portions of the United States Geological Survey (USGS) mapping zones 54, 55, 57, 58, and 59 covering the 46 counties of South Carolina. The test site consists of a large diversity of landforms (sandy beaches to coastal plains and hills to low mountains), soil types and land cover/use types. Hardwood forests are mixed throughout the State and are the dominant forest type followed by southern yellow pines (Smith et. al. 2004).

FOREST INVENTORY PLOTS

The forest inventory field plot design consists of four subplots approximately 1/24 acre in size, and delineate an area for collecting ecological forest vegetation data, primarily including trees with a diameter at breast height of 5 inches or greater (Figure 1). Each subplot contains a microplot of approximately 1/300 acre in size. Microplots are used to collect information data on seedlings and saplings.

For this study a subset of 3434 plots (2214 forested, 1029 non-forested, and 191 water) was filtered from a complete 5-panel statewide forest inventory for South Carolina. This subset includes all of the 100% forested and 100% non-forested plots, as well as mixed plots containing two or all three conditions in the same plot (forest and non-forest, forest and water, non-forest and water, and forest, non-forest and water), and it was used to produce a forest non-forest mask (map). Two datasets were created from this study subset. The first dataset consisted of forest inventory plots coded as forest, non-forest, water, and forested plots with privet. The second dataset consisted of forested plots and forested plots with privet. These two datasets were then used to generate two more datasets, for a total of four datasets. The 'forested plots with privet' category for the first two subsets was further subdivided into four classes, and coded based on the proportion of privet presence in the plot (trace< 1%, 1–10%, 11–50%, and 51–90%). These datasets were then used to model privet spatial distribution within forest.

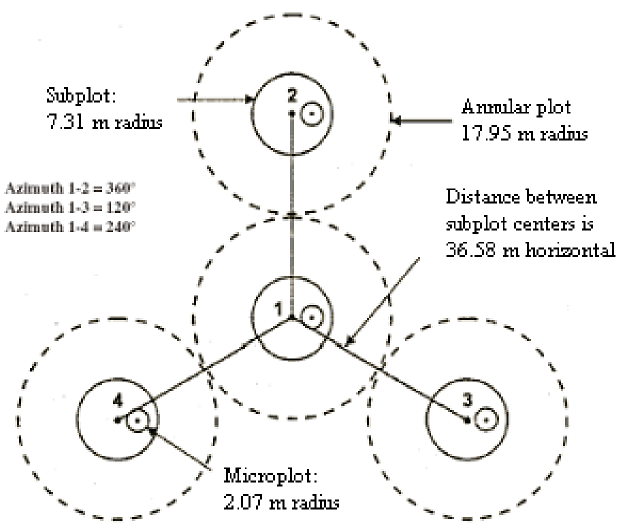

Figure 1. FIA field plot design.

DATA BASE DESCRIPTION

The database consists of multi-temporal MODIS satellite data, ancillary data (climate data, topographic data, land cover, etc), and FIA data. Climate data includes temperature measures, average monthly and annual precipitations derived from 4-km spatial resolution data. Soil information (such as porosity, permeability, ph, plasticity, rock volume, etc.) and topographic derivates from digital elevation models (DEM) such as elevation, slope, and dominant aspect were used as separate layers in the database. The National Land Cover Dataset (NLCD92, Vogelmann et al., 2001) provided info-data on the percentage of forest (conifer, deciduous, woody wetland, mixed), and shrub land. Layers from MODIS and ancillary data have been stacked to form a single image that contains 269 layers of data. This database was used for modeling the forest/non-forest and privet classifications. Table 1 shows some of the 269 layers that make up the database image. The layers in the database either existed as 250-meter resolution data or were re-sampled to 250-meters and projected to Albers Equal Area projection by personnel at the USFS Remote Sensing Applications Center in Salt Lake City (RSAC). The database contains continuous and categorical variables.

Table 1. List of Type of Layers Used to Map Forest, Non-forest and Privet.

Database Layers Description
MODIS 32 day composite imagery between 2001 and 2003
Conus MODIS32-2001097 - Bands 1 to 7
Conus MODIS32-2001193 - Bands 1 to 7
Conus MODIS32-2002129 - Bands 1 to 7
Conus MODIS32-2002257 - Bands 1 to 7
Conus MODIS32-2002321 - Bands 1 to 7
Conus MODIS32-2003161 - Bands 1 to 7
Conus Bailey’s Ecoregions image layer
MODIS Vegetation Indicies Layers
Conus EVI- 2002097 image
Conus EVI- 2002321 image
Conus NDVI- 2002097 image
Conus NDVI- 2002321 image
MODIS Vegetation Layer: MODIS –percent tree cover image
Reflectance layers from spring, summer and fall of 2002
Conus Reflectance – 2002097 – Bands 1 to 7
Conus Reflectance – 2002225 – Bands 1 to 7
Conus Reflectance – 2002321 – Bands 1 to 7
NLCD layers;
Conus NLCD – Percent conifer forest image
Conus NLCD – Percent deciduous forest image
Conus NLCD – Percent mixed forest image
Conus NLCD – Percent shrub land image
Conus NLCD – Percent woody wet land image
Terrain information; Conus dominant aspect, Conus mean elevation, stream density
Soil data layers; available water capacity, permeability, soil bulk density, soil ph, soil plasticity, soil porosity, rock volume and soil texture.
Precipitation – annual and for each month
Temperature layers – averages, minimum and maximum temperatures.

DECISION TREE – CUBIST AND SEE5

Cubist and See5 are regression tree software used to create decision tree classifications (forest/non-forest map) and models for modeling spatial distribution of privet invasive species. Two files are essential for running Cubist or See5, and several others are optional. The first essential file is the "names" file that lists the attribute names of each layer and describes the classes and attributes/predictors, as shown below.

FNF
FNF: 1, 2, 3, 4.
awc-250m.img-band1: continuous.
bdgrid-250m.img-band1: continuous.
conus-dvi-2002225.img: continuous.
conus-evi-2002097.img: continuous.
conus-modis32-2001097-albers.img-band1: continuous.
us_ppt01_jan.img: continuous.
us_ppt02_feb.img: continuous.
us_tavg301_albers.img: continuous.
us_tavg302_albers.img: continuous.
usgs_mapping_zones.img: 0, 54, 58, 59.
ustmax01_albers.img: continuous.
ustmax02_albers.img: continuous.
:
attributes excluded:
conus_modis32_2001097_albers.img_band5.
conus_modis32_2001193_albers.img_band5.

The first row/entry in the names file is the attribute (forest/non-forest, forest proportion) containing the target value to be classified/modeled based on values of the other predictors. Predictors contained in the names file are either continuous or defined by numeric categorical values. The final entry in the names file specifies whether a predictor is included or excluded from the classifier/modeler. The second essential file is the "data" file that provides information on the training data used to construct the decision tree model. A comma separates the values and the entry terminates with a period (http://www.rulequest.com). The "test" file is one of the optional files, and it is used to evaluate the performance of the classifier/modeler. In this study, data used to classify/model privet invasive species have been split randomly 90% and 10% into data training and test files.

FOREST NON-FOREST CLASSIFICATION

A decision tree algorithm in See5 was used to classify/model the land cover of South Carolina into forest, non-forest, water, and forest containing privet-invasive species, as well as four privet classes based on percent average of privet in the plot. A subset of forested, non-forested, and water plots was selected from a complete 5-year cycle of FIA data. The subset data was further used to produce two data sets that later were used in two different trials. The first trial used only forested area of South Carolina, while the second trial used the entire state's area to classify/model spatial distribution of privet invasive species within forest. Forest proportion in each of the 3434 plots was used as a predictor to create a forest mask in Cubist and to mask out non-forested area and keep the forested area. The forested area was used later (in See5) with two different datasets to classify/model privet invasive species presence, and its spatial distribution within the forested area. The following two datasets were used in the first trial: 1) MODIS satellite data, ancillary data, forested plots, and forested plots with privet; and 2) MODIS satellite data, ancillary data, forested plots, and forested plots with privet divided into 4 classes (trace< 1%, 1-10%, 11-50%, 51-90%; each code based on the field-call proportion of privet in each plot).

Two datasets were used for the second trial. The first: 1) MODIS satellite data, ancillary data, and all inventory plots (forested, non-forested, water, and forested plots with privet). The second: 2) MODIS satellite data, ancillary data, and all inventory plots (forested, non-forested, water, and forested plots with privet; but plots with privet were coded into the four classes as in the second dataset in the first trial, based on the average proportion of privet in the plot). Separately, each data set was used in See5 to create a decision tree model and classify the land cover into forest, non-forest, water, and forest with privet invasive species. A software tool was developed (for ERDAS Imagine) to convert remote sensing and GIS layers to See5 and Cubist data file formats by the Remote Sensing Applications Center (RSAC) in Salt Lake City.

Ten percent of the data was randomly set-aside (in each data set) for accuracy assessment. The boosting option in See5, set to ten trials was the only one used in the study. Each decision tree trial, produced by boosting, tries to correct the prediction error from the previous decision tree. The final output file from the See5 software program reports classification errors based on a confusion matrix produced for both training and test datasets (Table 2). Shown below is a sample of the See5 output.

Table 2: A sample of the See5 output showing the misclassifications.

```
Options:
    10 boosting trials
Class specified by attribute `fnf'
Trial 9:  Decision tree:
conus_modis32_2001097_albers.img_band7 > 954: 3 (4.8/0.8)
conus_modis32_2001097_albers.img_band7 <= 954:
:...conus_modis32_2003161_albers.img_band1 <= 220: 3 (2.1)
    conus_modis32_2003161_albers.img_band1 > 220: 1 (25.9/2.9)
  SubTree [S36]
conus_nlcd_percent_decidforest.img > 23: 3 (4.5/0.6)
conus_nlcd_percent_decidforest.img <= 23:
:...bdgrid_250m.img_band8 > 1.45: 1 (26.7/2.7)
    bdgrid_250m.img_band8 <= 1.45:
    :...l48_trasp.img <= 36: 1 (4)
  SubTree [S38]
conus_modis32_2001193_albers.img_band2 <= 4618: 2 (18)
conus_modis32_2001193_albers.img_band2 > 4618: 3 (3/1)
      l48_trasp.img > 36: 3 (7.9/1.6)
Evaluation on training data (428 cases):     Evaluation on test data (285 cases):
Trial      Decision Tree                      Trial      Decision Tree
        Size      Errors                              Size     Errors
  0      389     197 ( 6.4%)                  0       389     132(38.5%)
  1      279     401(13.0%)                   1       279     127(37.0%)
  2      312     393(12.7%)                   2       312     152(44.3%)
  3      304     388(12.6%)                   3       304     141(41.1%)
  4      306     406(13.1%)                   4       306     131(38.2%)
  5      304     371(12.0%)                   5       304     142(41.4%)
  6      310     384(12.4%)                   6       310     151(44.0%)
  7      283     445(14.4%)                   7       283     142(41.4%)
  8      328     371(12.0%)                   8       328     139(40.5%)
  9      300     403(13.0%)                   9       300     150(43.7%)
  boost           9( 0.3%)   <<              boost            95(27.7%)
(a)  (b)  (c)  (d)  (e)  classified as         (a)  (b)  (c)  (d)  (e)   classified as
                         (a): class 0                                    (a): class 0
     1659                (b): Forest                156   15    4    1   (b): Forest
        1  926           (c): Non-forest             26   75    1        (c): Non-forest
        5       329      (d): Privet                 28   11    6        (d): Privet
        2    1       168 (e): Water                   5    4        11   (e): Water
```

The decision tree obtained from boosting for each trial and dataset was used in the RSAC "Apply See5" module in ERDAS Imagine to model/classify the land cover into forest, non-forest, water and forest with privet as a function of the modeling dataset. The final product is a single layer image map (predicted output image) with values representing the variables (forest/non-forest, and privet) that were modeled (Figures 2 and 3) and a confidence image that shows spatial distribution of the correct and misclassified areas. Confidence values range from zero to one.

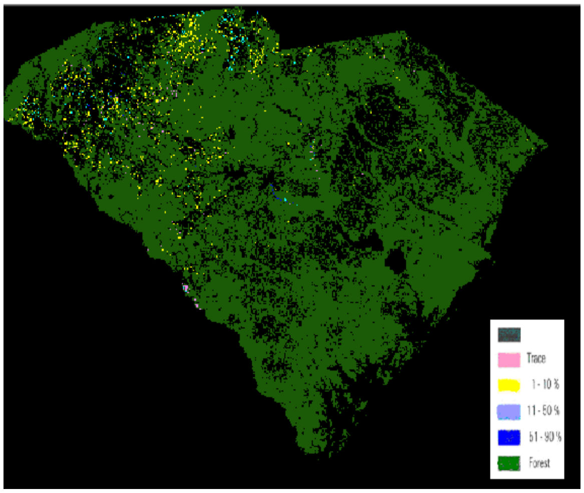

Figure 2. Privet spatial distribution from MODIS imagery overlaying green forest mask.

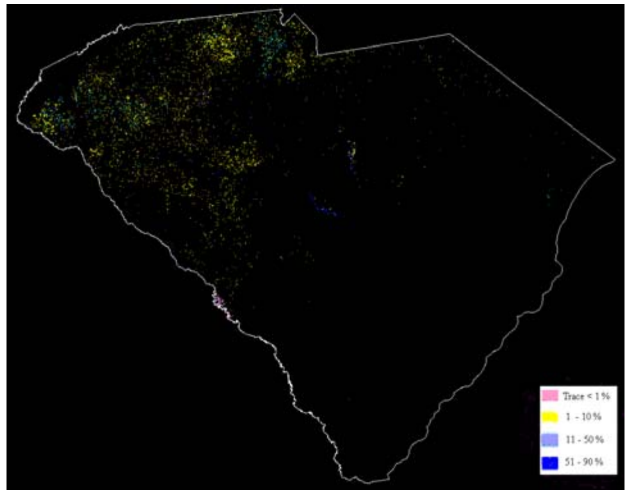

Figure 3. Privet spatial distribution from MODIS imagery – South Carolina.

RESULTS AND DISCUSSIONS

Privet invasive species spatial distribution was modeled for South Carolina from MODIS satellite data and forest inventory plot data. Two types of trials (one for forested area and the other for the entire state area) and two datasets of information were used in each trial to model privet invasive species spatial distribution. The two datasets used for forested only area consists of: 1) MODIS satellite data, ancillary, forested plots and forested plots with privet, and 2) MODIS satellite data, ancillary data, forested plots and forested plots with privet coded differently based on the proportion of privet in each plot. Accuracy assessment is based on analysis of a contingency table produced by the See5 program for the 10-percent set-aside dataset. The overall classification accuracy (%PCC) increases from 72.30 percent, when the entire state area is modeled, to 84.73 percent when only forested area of the state is modeled and all the privet plots used in the classification are grouped in a single class/category. Classification results of the two trials (forest only, and entire state's area) are presented in Table 3 and summarize the producer, user, and overall accuracy (%PCC) in South Carolina. Results of the first trial (forested area) show a decrease in overall classification accuracy from 84.73 percent, when plots with privet have the same code, to 78.64 percent when each privet plot is coded differently based on average privet proportion present in the plot. The decrease in accuracy is the result of a combination of assigning privet classes and an insufficient number of plots in classes containing greater than 10 percent privet. Results of the second trial (state level) also show a decrease in overall classification accuracy (%PCC), from 72.30 percent to 69.5 percent, as the FIA plots with privet used in the model/classification are coded based on the proportion of privet in the plot.

Table 3. Classification accuracy of forest/non-forest and privet.

Trial Types	Overall %PCC	Producer accuracy %				User accuracy %			
		Forest	Privet invasive species	Non-forest	Water	Forest	Privet Invasive species	Non-forest	Water
Entire South Carolina	72.30	88.64	13.33	73.53	55.00	72.56	54.55	71.43	91.67
Forested Area	84.73	96.57	10.71	*	*	87.11	33.33	*	*

When forest inventory plots with privet used in each trial have been grouped in four classes based on the proportion of privet in the plot (Table 4), overall classification accuracy increased from 69.5 percent (second trial-state) to 78.64 percent when only forested area was modeled. In both trials, results from the contingency table (test data) show poor performance of the models for privet classes that show only trace of privet in the plot.

Table 4. Classification accuracy of forest/non-forest and privet classes.

Trial	PCC %	Producer accuracy %							User accuracy %						
		Privet %				Forest	Non forest	H_2O	Privet %				Forest	Non forest	H_2O
		Trace	1-10	11-50	51-90				Trace	1-10	11-50	51-90			
1	78.64	*	4.55	*	*	98.77	*	*	*	33.3	*	*	79.70	*	*
2	69.39	*	4.17	*	*	87.50	74.23	57.89	*	20.0	*	*	71.96	63.7	100

A very low number of forested plots with privet in class four (51–90% privet) in both the model and test data is the primary cause of poor classification accuracy for this category. From a total of 13 plots in class four (51–90%), only one plot became part of the test set for trial 1 and six plots fell in the test set for trial 2 (to test the performance of each model). The second test set for each trial is not representative of the privet class/category four.

It is important to note that classification accuracy does not indicate anything about the purity of the classified pixels (250 by 250 m) that correspond to FIA plots.

CONCLUSIONS

This dataset is part of a product that was developed with the intention of using a full five-year cycle of FIA data. Privet at the state level provides information on how forest with privet is spatially distributed throughout the entire statewide landscape. The spatial pattern allows a visual assessment of privet distribution below the county level to help show potential areas containing high and low incidence of privet invasive species.

A larger number of plots with privet is needed in class four (13 privet plots) and class three in order to evaluate the performance of the model. The small class 4 can be combined with class 3 and a third trial performed on each of these two new datasets.

FIA plot information ties See5 and Cubist models to actual FIA phase 2 plot measurements on the ground.

Even though the increases in accuracy were not particularly large, they are meaningful when used in different forest applications.

REFERENCES

Blackard, J.A, M.V. Finko,E.H. Hemler, G.R. Holden, M.L Hoppus,D.M. Jacobs,A.J. Lister, G.G.Moisen, M.D Nelson, R.Riemann, B.Ruefenacht, D. Salajanu, D.L. Weyermann, K.J. Winterberger, T.J.Brandeis, R.L.Czaplewski, R.E. McRoberts, P.L.Patterson, and R.P. Tymcio, 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information (MODIS Data), *Remote Sensing of Environment*, Vol. 112/4 pp. 1658 – 1677.

Borry, F. C., B. P DeRoover, R. R. DeWulf, and R. E. Goossens, 1990. Assessing the value of monotemporal SPOT-1 imagery for forestry applications under Flemish conditions, *Photogrammetric Engineering and Remote Sensing*, 56(8):114–1153.

Chavez, P .S., Jr., 1986. Digital merging of Landsat TM and digitized NHAP data for 1:24,000 scale image Mapping, *Photogrammetric Engineering and Remote Sensing*, 52(10):1637-1646.

Chavez, P. S Jr., S. C. Sides, and J. A. Anderson, 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic, *Photogrammetric Engineering and Remote Sensing*, 57(3):295–303.

Conner, R. C., T. Adams, B. J. Butler, W. A. Bechtold, T. G. Johnson, S. N. Oswalt, G. Smith, S. W. Wolf, and C.W. Woodall, 2004. *The State of South Carolina's Forests, 2001*, pp. 67.

Cooke, W. H., D. M. Jacobs, 2005. Rapid classification of Landsat TM imagery for phase 1 stratification using the automated NDVI threshold supervised classification (ANTSC) methodology, In: *Proceedings of the Fourth Annual Forest Inventory and Analysis Symposium*, 2002 November 19–21 New Orleans, LA.

Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station, pp. 81-86.

Guo, X., K. P. Price, and J. Stiles, 2003. Grasslands discriminant analysis using Landsat TM single and multitemporal data, *Photogrammetric Engineering and Remote Sensing*, 69(11):1255–1262.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan, 2004. Development of a 2001 national land-cover database for the United States, *Photogrammetric Engineering and Remote Sensing*, 70(7):829–840.

Lawrence, R. L., and A. Wright, 2001. Rule-based classification system using classification and regression tree (CART) analysis, *Photogrammetric Engineering and Remote Sensing*, 67(10):1137–1142.

Lunetta, R.S., J. G. Lyon, B. Guindon, and C. D. Elvidge, 1998. North American landscape characterization dataset development and data fusion issues, *Photogrammetric Engineering and Remote Sensing,* 64(8):821-829.

Ma, Z., and C. E. Olson Jr., 1989. A measurement of spectral overlap among cover types, *Photogrammetric Engineering and Remote Sensing*, 55(10):1441-1444.

Pellemans, A. H. J. M., R. W. L. Jordans, and R. Allewijn, 1993. Merging multispectral and panchromatic SPOT images with respect to the radiometric properties of the sensor, *Photogrammetric Engineering and Remote Sensing*, 59(1):8 –87.

Ricchetti, E., 2000. Multispectral satellite image and ancillary data integration for geological classification, *Photogrammetric Engineering and Remote Sensing*, 66(4):429–435.

Salajanu, D., 1992. Utilization of SPOT panchromatic data for forest cover/land use classification. A Master's Thesis in School of Natural Resources, The University of Michigan, Ann Arbor, MI., pp. 88.

Salajanu, D., and C. E. Olson Jr., 2001. The significance of spatial resolution: Identifying forest cover from satellite data, *Journal of Forestry*, 99(6):32-38.

Skidmore, A. K., B. J. Turner, and E. Knowles, 1997. Performance of a neural network: Mapping forests using GIS and remotely sensed data, *Photogrammetric Engineering and Remote Sensing*, 63(5):501-514.

Smith, W. B., P. D. Miles, J. S. Vissage, and S. A. Pugh, 2004. Forest resources of the United States, 2002, Gen. Tech. Rep. NS-241, St. Paul MN: U.S. Department of Agriculture, Forest Service, North Central Research Station, pp. 137.

Vogelmann, J. E., T. Sohl, and S. M. Howard, 1998. Regional characterization of land cover using multiple sources of data, *Photogrammetric Engineering and Remote Sensing,* 64(1):45-57.

Vogelmann, J. E., S. Howard, L. Yang, C. Larson, B. Wylie, and N. Van Driel, 2001. Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources, *Photogrammetric Engineering and Remote Sensing*, 67(4):650-661.

Wayman, J. P., R. H. Wynne, J. A. Scrivani, and G. A. Reams, 2001. Landsat TM-based forest area estimation using iterative guided spectral class rejection, *Photogrammetric Engineering and Remote Sensing*, 67(10):1155–1166.