CONTRIBUTION OF CLIMATE, SOIL, AND MODIS PREDICTORS WHEN MODELING FOREST INVASIVE SPECIES DISTRIBUTION USING FOREST INVENTORY 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

Forest inventory and analysis data are used to monitor the presence and extent of certain non-native invasive species. Effective control of its spread requires quality spatial distribution information. There is no clear consensus why some ecosystems are more favorable to non-native species. The objective of this study is to evaluate the relative contribution of geo-spatial predictor variables, individually and groups, to the overall classification accuracy of the model. The three selected major groups of geo-spatial data are MODIS satellite imagery, soil properties, and climate information. We combined predictor variables with forest inventory information, to model/classify the spatial distribution of privet invasive species. Models were separately developed for each group (MODIS group, soil group, climate group), and all data (269 layers) together. Forest inventory plot information and predictor variable information from each group were used to model spatial distribution of forested areas exhibiting the potential to contain privet. Classification results are based on a ten percent set aside dataset. Overall classification accuracy showed that the all data together model performs better than the stand-alone group predictive model. The climate predictive model performs better in identifying forest with privet among the stand alone (MODIS and soil) predictive models. There is no significant difference between the overall classification accuracy obtained from soil predictive model and the climate predictive model. However, there is a significant difference on overall classification accuracy at 90% and 95% significance level between the MODIS predictive model and both the soil and climate predictive models.

INTRODUCTION

Invasive species are both beneficial and detrimental to the forest ecosystem, whether in the forest or beyond the forest edge. Invasive species are an important source of food and cover for wildlife (migratory birds and big animals) in a variety of habitats, but also detrimental to the forest regeneration, species composition, soil processes, and is a fuel source for wild fire. A study of *Ligustrum lucidum* invasion on Argentina's secondary forest patches (Aragon and Groom 2003), shows that this invasive species is a prolific fruit producer capable of germinating and surviving in a very large range of forest environments. The finding shows a higher survival and growth rate in comparison to the native species that affects species composition.

Humans, birds and wildlife, catastrophic events, climate, and soil factors have an important contribution in the spread of invasive species. Martin (1999) suggested that competition for light negatively affects diversity of the understory in the eastern states, while Busch and Smith (1995) reported that a combination of light and water competition helps salt cedar invasion on many riparian sites in the southwest (Levine et al., 2003). Invasive species can affect ecosystem processes by altering soil nitrogen cycling through differences in litter quantity and quality as well as changes in soil organic matter (Evans et al., 2001). The study found that *Bromus tectorum* invasive species reduces the rate of nitrogen mineralization by having higher carbon nitrogen and lignin nitrogen ratios in litter compared to native species. The change in litter quality caused by *Bromus* was considered the responsible mechanism for the change in the rate of nitrogen transformations, by altering soil organic matter.

Effective control of invasive species spread requires reliable information concerning the rate of spread, dispersal capability, soil properties, as well as good spatial distribution knowledge. Satellite remote sensing is an important tool that offers forest managers reliable information related to spatial distribution of different invasive species throughout the ecological landscape. Results from previous studies have shown that overall and individual classification accuracy as well as discrimination/separability between categories/classes improves when ancillary data are used in the classification process as new layers associated with the satellite data. In our study, we are trying to incorporate ancillary data, such as soil properties, climate and relief information with the multitemporal satellite images in order to improve the separability and accuracy with which invasive species present in the forest understory can be predicted, modeled and mapped. The dataset for this study contains a stack of 269 geo-spatial data layers. These 269 layers can be divided into several individual major groups. The three largest groups (MODIS satellite imagery, soil properties, and climate data), are partitioned separately for use in this study.

The main objective of this study is to evaluate the relative contribution of predictor variables of the major groups (MODIS group, soil properties, and climate information group) to the overall classification accuracy, when each major group in combination with forest inventory plot information is separately used to model/classify privet (*Ligustrum* spp.) invasive species spatial distribution.

STUDY AREA

South Carolina was the test area for this study. It includes small portions of the USGS mapping zones 54, 57, 58, and 59. The study area consists of diverse landforms, land cover, and land use types such as forests, agricultural lands, open spaces, urban areas, low populated rural areas, bodies of water, and wetlands. Southern conifer forests are the dominant forest type followed by the mixed conifer-hardwood and hardwoods. Hardwood forests consist of mixed broadleaf species throughout the area. The study area is approximately 57-60% forest and includes 3434 Forest Inventory and Analysis (FIA) plot locations.

FOREST INVENTORY PLOTS

Southern FIA uses a network of inventory plots to collect information on the state of the forests. A field plot design consists of four subplots approximately 1/24 acre in size, and are used to collect data on trees with a diameter at breast height of 5 inches or greater. Each subplot contains a microplot of approximately 1/300 acre in size. Microplots are used to collect information data on seedlings and saplings.

In this study, 3434 forest inventory plots (forested, non-forested, water, and forested containing privet) were filtered from a complete five panel dataset. Each of the 3434 plots consists of one condition (100 percent forested, 100 percent non-forested), or may have a mixture of two or more conditions (forest, non-forest, water, forest with privet) within the same plot.

DATA BASE DESCRIPTION

The database consists of multi-temporal MODIS satellite data, ancillary data (climate data, topographic data, land cover, etc) and FIA plot data. Acquisition dates for the MODIS multi-temporal data were 2001 (spring, and summer), 2002 (spring, summer, and fall), and 2003 (summer); including other reflectance information from 2002 (spring, summer and fall). MODIS derivates such as NDVI, EVI, and percent of tree cover were also included in the study. Most of the satellite images used had a 250-meter spatial resolution. Those with a different spatial resolution have been resampled to a 250-meters spatial resolution using nearest neighbor resampling method for categorical/discrete data, and linear interpolation resampling method for continuous data. Climate data include temperature measures, average monthly, and annual precipitations derived from 4-km spatial resolution data. Soil data from STATSGO were converted to raster data and resampled to a 250-meters spatial resolution. 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 included as a separate group of layers in the database. National Land Cover Dataset (NLCD92, Vogelmann et al., 2001) provided info-data on percentage of forest (conifer, deciduous, mixed), shrub land, and woody wetland. The individual layers from MODIS and ancillary data were stacked into a single geo-spatial dataset that contains 269 layers of data (stack image file). The 269 layer stack was subdivided into several stand alone major groups of predictive variables (MODIS group, soil group, climate group, etc.). Stand alone groups and all data groups' predictor variables were linked with forest inventory plots to produce several individual datasets. Three stand alone major groups that were used in this study are MODIS group, soil group, and climate group. MODIS group consists of the following layers: MODIS seven band composites from spring, summer, and fall, NDVI, and EVI, reflectance layers from spring, summer, and fall, NLCD layers of conifer, broadleaf, shrubs, and woody wetland, MODIS percent tree cover, and MODIS fire layers. Soil group consists of soil properties layers such as dominant fragmentation, dominant soil texture, permeability, soil pH, plasticity, porosity, rock depth, rock volume, etc. Climate group contains temperature and precipitation variables such as monthly and annual precipitation, monthly average and annual temperatures, and monthly minimum and maximum temperatures. These datasets were linked to FIA plots to produce nonparametric decision tree classification models. Table 1 shows some of the 269 data layers used in this study. The database contains continuous and categorical variables.

Table 1. List of satellite and ancillary data layers used in the models

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 MODIS32-2002225 - 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 Indices Layers
Conus EVI- 2002097 image
Conus EVI- 2002321 image
Conus NDVI- 2002097 image
Conus NDVI- 2002225 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
Terrain information; Conus dominant aspect, Conus mean elevation, stream density
Conus MODIS fire points from 2001 and 2002
Soil data layers; available water capacity, permeability, soil bulk density, soil ph, soil plasticity, soil porosity, rock volume and soil texture.
USGS mapping zone images
Precipitation – annual and for each month
Temperature layers – averages, minimum and maximum temperatures.
. . . . . .
. . . . . .

SEE5 -REGRESSION TREE

See5 is a regression-tree software program, used to create nonparametric decision tree models. Two files are essential for running See5, and several other optional functions. The first essential file is the names file that lists the names 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_tavg301_albers.img: continuous.
usgs_mapping_zones.img: 0, 54, 58, 59.
ustmax01_albers.img: continuous.
ustmax02_albers.img: continuous.

The attribute that contains the target values (forest, non-forest, or forest with privet) to be classified and modeled, based on the values of the other predictors is the first row in the names file (e.g. FNF). Predictors contained in the name file are either continuous or defined by numeric values (categorical data). The final entry in the name file specifies whether a predictor is included or excluded from the classifier/model. 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 is used to evaluate the performance of the classifier/model. Data used to classify/model spatial distribution of privet invasive species was randomly split into 90% for data training and 10% for test files.

FOREST AND NON-FOREST INVASIVE SPECIES CLASSIFICATION

A set of 3434 forest inventory plots from a complete 5-year cycle of FIA data can be found in the geographic study area. Attributes selected for these plots include forest and non-forest conditions and forest conditions containing privet. Forest inventory plot data were linked with predictor variables from each layer of the three major groups (MODIS, soil properties, climate), and all data together group (269 layers) used in the study through the "Prepare FIA data for See5" program. The "Prepare FIA Data for Cubist/See5" tool extracts geospatial data information using FIA plot locations. The outputs of this program are three different files (test, data, names), used as input to the See5 program to create nonparametric decision tree classification models. These three files were produced for each of the three stand-alone groups of dataset, and for all data together (269 layers): thus producing twelve files for input to See5. The test file is the result of a randomly selection subset of ten percent of the plots in the dataset. The test file is used to evaluate the accuracy of the model, and it shows the classifying person how well the model can discriminate between the categories that need to be classified. In this case, the test file shows how good the models developed from predictive values of each stand-alone group are in recognizing and classifying privet invasive species category. Data file from each group was used, separately, in the See5 program to build a nonparametric decision tree classification model. The boosting option, set to ten trials, was the only See5 option used for this study. Each decision tree tries to correct the prediction error from the previous decision tree. This process continues for the pre-determined number of trials (ten trials for this study).

A sample of the output file from the See5 software program (Table 2) reports classification errors based on a confusion matrix produced for both training and test datasets. The decision tree obtained from the boosting option for each trial and dataset was used to classify and model spatial patterns of privet, the non-native invasive species, in association with forest cover. The classification model also includes forest, non-forest, and water. The final product is a single layer image map (predicted output image) showing spatial distribution of classes/categories that were modeled (Figure 1) and a confidence image that shows spatial distribution of the correct and misclassified areas. Confidence values ranged from zero to one.

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

```
Options:
      10 boosting trials
Class specified by attribute `fnf'
Trial 9:  Decision tree:
SubTree [S1]
conus_tm6_albers.img_band6 <= 1: 4 (9.1)
conus_tm6_albers.img_band6 > 1:
:...conus_tm3_albers.img_band3 <= 14:
SubTree [S2]
conus_reflectance_2002321.img_band4 > 507:
conus_reflectance_2002321.img_band4 <= 507: 1 (254.9/22.2)
:...conus_modis_percent_tree_cover.img > 80.56: 2 (13.6/5.8)
   conus_modis_percent_tree_cover.img <= 80.56:
      :...us_ppt04_apr.img <=10467: 2  (36.6/11.6)
          us_ppt04_apr.img > 10467:  1  (14.5 )
  conus_tm3_albers.img_band3 > 14:
    bdgrid_30m.img_band10 > 2.64:  1  (42.3/14)
    :...conus_modis_percent_tree_cover.img > 67.852: 1 (18.4/4.3)
       conus_modis_percent_tree_cover.img <= 67.852:
Evaluation on training data (464 cases):        Evaluation on test data (308 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(27.5%)
(a)  (b)     (c)     (d)     (e)  classified as            (a)   (b)    (c)   (d)    (e)  classified as
    1659                          (b): class 1                   156    15     4      1    (b): class 1
       1    926                   (c): class 2                    26    75     1           (c): class 2
       5            329           (d): class 3                    28    11     6           (d): class 3
       2      1             168   (e): class 4                     5     4            11   (e): class 4
```

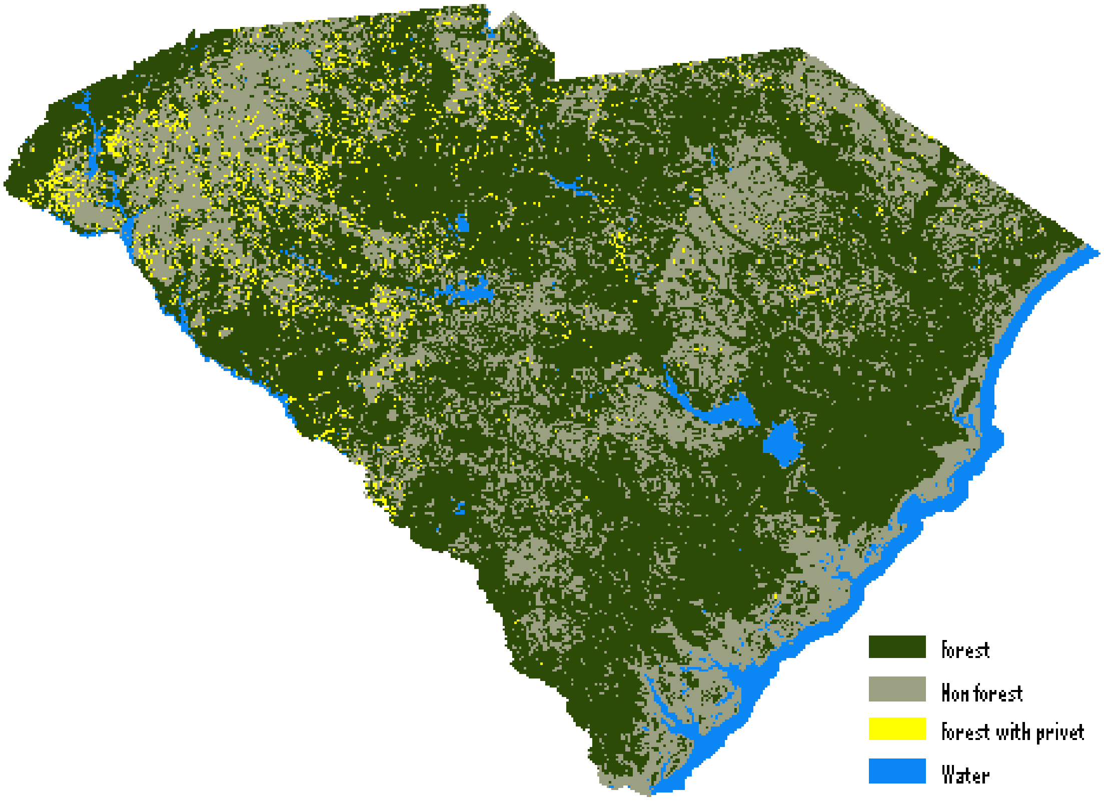

Figure 1. Forest, non-forest and forest with privet.

RESULTS AND DISCUSSIONS

Privet invasive species spatial distribution models have been developed for South Carolina, using forest inventory plot information and predictor variables information from each of the major groups (MODIS, soil, and climate) present in the stack image file. A total of 3434 forest inventory plots have been used separately with the predictor variable from each of the three major groups (MODIS, climate, soil) in See5. The See5 boosting option of 10 trials was used to classify and model the land cover into forest, non-forest, water, and forest with privet invasive species. Classification results from See5 for each trial are presented in Table 3. Overall classification accuracy (%PCC) increased from 60.06%, when only climate group (temperature and precipitation) information was used in the model, to 72.30% when all 269 stack image layers were used in the model. Results of the classification show that the overall accuracy and the producer and user accuracy of each classified category were affected by the predictive variables of the major groups used in the model. The individual use of just MODIS, soil properties, and climate groups alone resulted in smaller overall classification accuracy than when all groups were used in the model. Accuracy assessments performed were based on analysis of a contingency table (produced by the See5 algorithm) for the 10-percent set-aside dataset. This test is used to evaluate the predictive ability (validation) of the selected model, and not for the overall classification accuracy of the final output. Overall classification accuracy shows that the all data group prediction model performed slightly better than the individual group's alone predictive models (Tables 3). MODIS group has the highest overall classification accuracy (67.93%) among the individual group's alone predictive models. Producer and user accuracy at the category level is different from one group alone model to another. Predictive models from each group perform relatively poor in detecting the forestland cover containing privet. Among the three major groups the climate group alone model has the highest classification accuracy of forest containing privet (8.89% producer, and 40.00% user accuracy).

The Z test proposed by Cohen (1960) was used to test for significant differences between the classifications obtained from two different models. The results show that there is no significant difference between the classification accuracy obtained from the all groups model (269 layers), and the MODIS alone group model at 90% and 95% confidence level (Z=1.33 <1.64 and Z=1.33 < 1.96). There is no significant difference between the overall classification accuracy obtained from soil properties alone group model and climate alone group model (Z=0.189 <1.64). There is a significant difference between the MODIS alone group model classification and soil properties and climate alone groups models at both 90% and 95% confidence level (Z=2.59 >1.64, and Z=2.59 >1.96 for soil group, respectively Z=2.50 >1.64 and Z=2.50 >1.96 for climate group model). A significant difference in classification accuracy is also found between the all data group model (269 layers), and soil alone and climate alone models (Z=3.88 >1.96 and Z=3.84 >1.96).

Table 3. Classification accuracy of forest/non-forest and privet from test data by group variables

Type of Data Groups	Overall %PCC	Producer accuracy %				User accuracy %			
		Forest	Nonforest	Water	Privet	Forest	Nonforest	Water	Privet
All 269-layers	72.30	88.64	73.53	55.00	13.33	72.57	71.43	91.67	54.54
MODIS	67.93	84.66	69.61	50.00	6.67	69.30	66.98	100.0	25.00
Soil	60.35	89.20	36.28	55.00	2.22	62.30	51.35	78.57	33.33
Climate T&P	60.06	84.09	48.04	25.00	8.89	64.63	51.04	62.50	40.00

The small number of plots containing forest with privet, in the test file, makes the test unreliable for evaluating the predictive ability of the selected model. According to user, producer, and overall accuracy (Table 3) the all data group model performed better than the stand alone (MODIS, soil, and climate) models in modeling and classifying of each land cover classes. Both producer and user accuracies show a slightly higher classification accuracy of the forest class than for the non-forest. Both models have a high percent of misclassification among the forest with privet classes, even though the all data group model performed better than the stand-alone models.

CONCLUSIONS

Based on the work described, the all data group model (269 layers) performs better than the individual stand alone (MODIS, soil, and climate) models, in modeling spatial distribution of forest with privet. Privet classifications provide information on its spatial distribution throughout the forested area.

The climate group predictive model performed better than the MODIS and soil group models in classifying forest with privet. More research is needed to determine whether the climate group model performs better than the soil or the MODIS model when other invasive species plot data are used as input for the classification models (honeysuckle, roses, tree-of-heaven, etc.).

The spatial distribution pattern provides a visual assessment for the occurrence of high and low percent cover for these invasive species in forested stands.

FIA plot information ties See5 models to actual FIA plot measurements on the ground.

Low number of plots in the test file for forest with privet makes the test less reliable in evaluating how effective the models are in modeling invasive species. Thus, there is a need to compare the final classification output from each model with the field in order to validate the model.

Results suggest that FIA plot information can be used to yield good results in mapping spatial distribution of invasive species, showing potential areas with high and low densities of these species.

A recommendation is to revisit this, or a similar invasive species study.

REFERENCES

Aragon, R., and M. Groom, 2003. Invasion by Ligustrum lucidum in NW Argentina: early stage characteristics in different habitat types, *Biologia Tropical*, 51(1):59-38.

Blackard, J.A., M.V. Finco, E.H. Helmer, 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.C. 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, *Remote Sensing of Environment*, 112(2008):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.

Busch, D.E. and S.D. Smith, 1995. Mechanisms associated with decline of woody species in riparian ecosystem of the southwestern U.S., *Ecol. Monogr*., 65:347-370.

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*.

Cohen, Jacob, 1960. A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20(l):37-40.

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, 67 p.

Cooke, W.H. III, 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.

Evans, R.D., R. Rimer, L. Sperry, and J. Belnap, 2001. Exotic plant invasion alters nitrogen dynamics in an arid grassland, *Ecological Applications*, 11(5):1301-1310.

Levine, J.M., M. Vila, C.M. D’Antonio, J.S. Dukes, K. Grigulis, and S. Lavorel, 2003. Mechanisms underlying the impacts of exotic plant invasions, *Proceedings of the Royal Society*, 270:775-781.

Martin, P.H., 1999. Norway maple invasion of a natural forest stand: understory consequences and regeneration pattern, *Biol. Invasions*, 1:215-222.

Ruefenacht, B., M.V. Finco, M.D. Nelson, R. Czaplewski, E.H. Helmer, J.A. Blackard, G.R. Holden, A.J. Lister, D. Salajanu, D. Weyermann, and K. Winterberger, 2008. Conterminous U.S. and Alaska forest types mapping using forest inventory and analysis data, *Photogrammetric Engineering and Remote Sensing*, 74(11):1379–1388.

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.

Salajanu, D., and D.M. Jacobs, 2006. The effect of using complete and partial forested FIA plot data on biomass and forested area classifications from MODIS satellite data, American Society of Photogrammetry and Remote Sensing 2006 Annual conference, Reno, Nevada, May 1-5, 2006.