Conterminous U.S. and Alaska Forest Type Mapping Using Forest Inventory and Analysis Data

B. Ruefenacht, 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.

Abstract

Classification-trees were used to model forest type groups and forest types for the conterminous United States and Alaska. The predictor data were a geospatial data set with a spatial resolution of 250 m developed by the U.S. Department of Agriculture Forest Service (USFS). The response data were plot data from the USFS Forest Inventory and Analysis program. Overall accuracies for the conterminous U.S. for the forest type group and forest type were 69 percent (Kappa = 0.66) and 50 percent (Kappa = 0.57), respectively. The overall accuracies for Alaska for the forest type group and forest type were 78 percent (Kappa = 0.69) and 67 percent (Kappa = 0.61), respectively. This is the first forest type map produced for the U.S. The forest type group map is an update of a previous forest type group map created by Zhu and Evans (1994).

B. Ruefenacht and M.V. Finco are with the USDA Forest Service, Remote Sensing Applications Center, 2222 West 2300 South, West Valley City, UT 84119 (bruefenacht@fs.fed.us).

M.D. Nelson is with the USDA Forest Service, St. Paul, MN 55108.

R. Czaplewski is with the Rocky Mountain Research Station, Fort Collins, CO 80526.

E.H. Helmer is with the International Institute of Tropical Forestry USDA Forest Service, Río Piedras, PR 00926.

J.A. Blackard is with the U.S. Army Corps of Engineers, Fort Worth District, Fort Worth, TX 76102, and formerly with the Rocky Mountain Research Station, Ogden, UT 84401.

G.R. Holden is with the Cibola National Forest, Albuquerque, NM 87113, and formerly with the USDA Forest Service, North Central Research Station, St. Paul, MN 55108.

A.J. Lister is with the Northeastern Research Station, Newtown Square, PA 19073.

D. Salajanu is with the Southern Research Station, Knoxville, TN 37919.

D. Weyermann is with the Pacific Northwest Research Station, Portland, OR 97205.

K. Winterberger is with the Pacific Northwest Research Station, Anchorage, AK 99503.

Introduction

The United States Department of Agriculture Forest Service (USFS) Forest Inventory and Analysis (FIA) program has been in continuous operation since 1930. The mission of FIA is to inventory the renewable forest and rangeland resources of the U.S. To inventory these resources, FIA has placed plots throughout the U.S. at an intensity of approximately one plot per 2,000 ha (6,000 acres) (Forest Inventory and Analysis, 2004). FIA uses an annual rotating panel system where between 10 to 20 percent of each state's FIA plots are sampled every year. From this plot data, the FIA program produces annual reports in the form of tabular data at the county and state level. This information is freely available to the public, but the original plot locations are not available due to provisions of the Food Security Act of 1985 (7 U.S.C. 2276).

FIA is legally required to provide summarized or analyzed data that are readily available and targeted at different audiences. One of the ways to accomplish this objective is to provide geospatial modeled products using FIA plot data and remote sensing imagery. Blackard *et al.* (2008) developed a forest/non-forest map and an above-ground live forest biomass map for the conterminous U.S., Alaska, and Puerto Rico derived from modeling FIA plot forest/non-forest and biomass variables as functions of 250 m resolution geo-spatial database. Observed biomass values from an independent test data set were favorably correlated to the predicted biomass values with correlation coefficients ranging between 0.40 to 0.78. Additionally, 21 States' modeled biomass estimates fell within 10 percent of the plot-based biomass estimates. Classification accuracies for the forest/non-forest product ranged from 80 to 98 percent. Thus, modeling FIA plot attributes as functions of remote sensing images and GIS data layers effectively scales plot-based forest attributes to national maps.

Blackard *et al.* (2008) used classification and regression-trees (CART) to model biomass and forest/non-forest. Using CART for land-cover classification is becoming popular (DeFries and Chan, 2000; DeFries *et al.*, 1998; Friedl and Brodley, 1997; Friedl *et al.*, 1999; Hansen and DeFries, 1996). CART procedures have several advantages over more traditional classification procedures, such as, supervised and

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 11, November 2008, pp. 1379–1388.

0099-1112/08/7411–1379/$3.00/0

unsupervised algorithms (Lillesand and Kiefer, 2000; Pal and Mather, 2003). Classification-trees are non-parametric, and as such, do not require assumptions about data distributions and can handle non-linear relationships between variables. They can also allow for missing data values, handle both numerical and categorical data, and incorporate many different data layers. The hierarchal structure of classification-trees makes interactions between data layers easier to interpret. Classification-trees are significantly less labor intensive than other classification techniques and can be used efficiently for large land-cover classifications (Friedl *et al.*, 1999; DeFries *et al.*, 1998). With quality training data, the accuracies of classification-trees are either similar to or better than supervised and unsupervised classification techniques (Lawrence and Wright, 2001; Friedl *et al.*, 1999; Friedl and Brodley, 1997; Hansen *et al.*, 1996).

Others have used remote sensing imagery and FIA plot data to create national mapping products. Zhu and Evans (1994) produced a forest type group map covering the entire United States and Puerto Rico. This forest type group map was produced using Advanced Very High Resolution Radiometer (AVHRR) imagery collected in 1992 and FIA plot data, which was the first attempt to create a forest type group map of the U.S. The procedure used to create the forest type group map involved several iterations of unsupervised classification algorithms, spectral signature evaluation, masking, and recoding.

Forest type group and forest type are two FIA plot variables. Eyer (1980) defined 145 forest types, which are aggregations or pure stands of forest trees. The FIA program uses a modified version of Eyer's (1980) forest type classification scheme. FIA combined some of the Eyer (1980) forest types and others were redefined for a total of 142 forest types (Table 1). Eyer (1980) grouped the forest types into 20 forest type groups, which classification scheme came from the USFS Renewable Resources Evaluation Group program. FIA uses a similar forest type group classification scheme. FIA defined eight new forest type groups for a

Table 1. List of the USDA Forest Service Forest Inventory and Analysis Forest Type Groups and Forest Types with Their Associated Codes

White/Red/Jack Pine Group	**100**
Jack Pine	101
Red Pine	102
Eastern White Pine	103
Eastern White Pine/Eastern Hemlock	104
Eastern Hemlock	105
Spruce/Fir Group	**120**
Balsam Fir	121
White Spruce	122
Red Spruce	123
Red Spruce/Balsam Fir	124
Black Spruce	125
Tamarack	126
Northern White-cedar	127
Longleaf/Slash Pine Group	**140**
Longleaf Pine	141
Slash Pine	142
Loblolly/Shortleaf Pine Group	**160**
Loblolly Pine	161
Shortleaf Pine	162
Virginia Pine	163
Sand Pine	164
Table-mountain Pine	165
Pond Pine	166
Pitch Pine	167
Spruce Pine	168
Pinyon/Juniper Group	**180**
Eastern Red Cedar	181
Rocky Mountain Juniper	182
Western Juniper	183
Juniper Woodland	184
Pinyon Juniper Woodland	185
Douglas-fir Group	**200**
Douglas-fir	201
Port Orford Cedar	202
Ponderosa Pine Group	**220**
Ponderosa Pine	221
Incense Cedar	222
Jeffrey Pine/Coulter Pine/Bigcone Douglas Fir	223
Sugar Pine	224
Western White Pine Group	**240**
Western White Pine	241
Fir/Spruce/Mountain Hemlock Group	**260**
White Fir	261
Red Fir	262
Noble Fir	263
Pacific Silver Fir	264
Engelmann Spruce	265
Engelmann Spruce/Subalpine Fir	266
Grand Fir	267
Subalpine Fir	268
Blue Spruce	269
Mountain Hemlock	270
Alaska-yellow-cedar	271
Lodgepole Pine Group	**280**
Lodgepole Pine	281
Hemlock/Sitka Spruce Group	**300**
Western Hemlock	301
Western Red Cedar	304
Sitka Spruce	305
Western Larch Group	**320**
Western Larch	321
Redwood Group	**340**
Redwood	341
Giant Sequoia	342
Other Western Softwoods Group	**360**
Knobcone Pine	361
Southwest White Pine	362
Bishop Pine	363
Monterey Pine	364
Foxtail Pine/Bristlecone Pine	365
Limber Pine	366
Whitebark Pine	367
Misc. Western Softwoods	368
California Mixed Conifer Group	**370**
California Mixed Conifer	371
Exotic Softwoods Group	**380**
Scotch Pine	381
Australian Pine	382
Other Exotic Softwoods	383
Norway Spruce	384
Introduced Larch	385
Oak/Pine Group	**400**
Eastern White Pine/Northern Red Oak/White Ash	401
Eastern Redcedar/Hardwood	402
Longleaf Pine/Oak	403
Shortleaf Pine/Oak	404
Virginia Pine/Southern Red Oak	405
Loblolly Pine/Hardwood	406
Slash Pine/Hardwood	407
Other Pine/Hardwood	409
Oak/Hickory Group	**500**
Post Oak/Blackjack Oak	501
Chestnut Oak	502
White Oak/Red Oak/Hickory	503
White Oak	504
Northern Red Oak	505
Yellow-poplar/White Oak/Northern Red Oak	506
Sassafras/Persimmon	507
Sweetgum/Yellow-poplar	508
Bur Oak	509
Scarlet Oak	510

Yellow-poplar	511	Paper Birch	902
Black Walnut	512	Gray Birch	903
Black Locust	513	Balsam Poplar	904
Southern Scrub Oak	514	**Alder/Maple Group**	**910**
Chestnut Oak/Black Oak/Scarlet Oak	515	Bigleaf Maple	912
Red Maple/Oak	519	**Western Oak Group**	**920**
Mixed Upland Hardwoods	520	Gray Pine	921
Oak/Gum/Cypress Group	**600**	California Black Oak	922
Swamp Chestnut Oak/Cherrybark Oak	601	Oregon White Oak	923
Sweetgum/Nuttall Oak/Willow Oak	602	Blue Oak	924
Overcup Oak/Water Hickory	605	Deciduous Oak Woodland	925
Atlantic White-cedar	606	Evergreen Oak	926
Baldcypress/Water Tupelo	607	Coast Live Oak	931
Sweetbay/Swamp Tupelo/Red Maple	608	Canyon Live Oak/Interior Live Oak	932
Elm/Ash/Cottonwood Group	**700**	**Tanoak/Laurel Group**	**940**
Black Ash/American Elm/Red Maple	701	Tanoak	941
River Birch/Sycamore	702	California Laurel	942
Cottonwood	703	Giant Chinkapin	943
Willow	704	**Other Western Hardwoods Group**	**950**
Sycamore/Pecan/American Elm	705	Pacific Madrone	951
Sugarberry/Hackberry/Elm/Green Ash	706	Mesquite Woodland	952
Silver Maple/American Elm	707	Cerocarpus Woodland	953
Red Maple/Lowland	708	Intermountain Maple Woodland	954
Cottonwood/Willow	709	Misc. Western Hardwood Woodlands	955
Oregon Ash	722	**Tropical Hardwoods Group**	**980**
Maple/Beech/Birch Group	**800**	Sable Palm	981
Sugar Maple/Beech/Yellow Birch	801	Mangrove	982
Black Cherry	802	Other Tropical	989
Cherry/Ash/Yellow-poplar	803	**Exotic Hardwoods Group**	**990**
Hard Maple/Basswood	805	Paulownia	991
Elm/Ash/Locust	807	Melaluca	992
Red Maple/Upland	809	Eucalyptus	993
Aspen/Birch Group	**900**	Other Exotic Hardwoods	995
Aspen	901		

total of 28 forest type groups (Table 1). The forest type group and forest type are determined for a plot in the following manner. Each tree on the plot is placed into an appropriate forest type group. The stocking values, which are individual trees' contributions to the total stocking of the stand, of the trees within the forest type groups are summed. Using a decision tree, the final forest type is assigned to each plot (Arner *et al.*, 2003). Since the forest type group and forest type classification schemes are hierarchical, the final forest group is determined by the forest type.

The main objective of this study was to examine the feasibility of using low resolution imagery, such as 250 m Terra MODIS imagery, and FIA plot data to produce national mapping products in a timely, efficient, and accurate manner. The FIA plot variables chosen for this project were forest type group and forest type. Even though the forest type group is determined by the forest type, and thus, the forest type group can be obtained by simply aggregating the forest types, the variables were modeled separately. Since the forest type group is a more general classification scheme than the forest type classification scheme, the forest type group final result was expected to have higher accuracy than the forest type final result. If only the forest types were modeled and the forest type groups were created by aggregating the forest types, the inaccuracies of the forest types will be compounded negatively affecting the forest type group accuracies. For instance, it is expected that the forest types in the oak/hickory group to have low accuracies due to the spectral similarities between the types in that group. If the oak/hickory forest type group was created by aggregating the types within that group, the oak/hickory forest type group would have low accuracy as well. However, the forest type group might have a high accuracy value if classified separately. Furthermore, the forest type group and forest type products might be used by different groups for different purposes. For these reasons, forest type groups and forest types were modeled separately creating two completely independent products. This means that a forest type pixel might not correspond to the forest type group assigned to the same pixel. If this is a concern or presents a problem to users, users can simply aggregate the forest types creating their own forest type group product.

Because CART does have the capability to handle large datasets and produce accuracies similar to other techniques, CART was chosen to model forest type groups and forest types for Alaska and the conterminous U.S. If the techniques developed for creating these spatial products are effective, perhaps additional FIA variables can be modeled and made available to the public.

Methods

The area of each FIA plot is categorized into a single condition or multi-conditions based on owner class and land class, which includes forest, non-forest, and water (Forest Inventory and Analysis, 2004). If an FIA plot is comprised of multi-conditions, the proportion of each condition occurring within the plot is calculated. For this study, all FIA plots with at least 50 percent of the plot area categorized into the land class forest condition were used for the modeling procedure. The land class forest condition is defined as greater than 0.4 ha (1 acre) in size, greater than 37 m (120 feet) in width, having or has been at least 10 percent stocked by trees of any size in the past or where stocking cannot be determined (e.g., western woodlands), having or has been at least five percent crown cover by trees of any size in the past, having an undisturbed understory, and not subjected to uses that prevent normal tree regeneration and succession (Forest Inventory and Analysis, 2004). According to this definition, transitional plots (i.e., plots temporarily cleared of trees) are categorized under the forest condition even though few or no trees currently exist on the plot. Transitional plots were used in the modeling procedure. It is unknown exactly how many transitional plots were actually used, but transitional plots probably comprised less than one percent of the total number of plots.

For the mapping of the forest type group and forest type, the FIA plot data were collected between 1978 and 2004. The majority of the plot data (55 percent) were collected between 2000 and 2004; 36 percent of the plot data were collected between 1990 and 1999, and nine percent were collected pre-1990.

A geospatial data set consisting of 269 remote sensing images and GIS layers with a spatial resolution of 250 m served as the predictor variables in the modeling of the forest type group and forest type for the conterminous U.S. All images and layers were projected to the Albers Conical Equal Area NAD27 projection. For Alaska, 19 geospatial remote sensing images and GIS layers were available, and most of the images and layers had a native spatial resolution of 250 m. Those with a native spatial resolution greater than 250 m underwent either nearest neighbor resampling if the data were categorical or bilinear interpolation resampling if the data were continuous.

The National Land Cover Database (NLCD), elevation, slope, and aspect data were at 30 m spatial resolution. NLCD was recoded into five classes: deciduous, developed, evergreen forest, mixed forest, shrubland, and woody wetland (Vogelmann *et al.*, 2001). To rescale the NLCD to 250 m, the percent of each of these NLCD classes occurring within a 250 m pixel was calculated.

The 30 m elevation data set was rescaled to 90 m. Mean elevation was calculated for a 3×3 window resulting in a 270 m mean elevation data set, which was resampled to 250 m using bilinear interpolation. Slope and aspect were derived from the 90 m elevation data set. The 90 m slope data set was resampled to 250 m using bilinear interpolation. The aspect data set was recoded into four categories: (a) 0 to 90 degrees, (b) 91 to 180 degrees, (c) 181 to 270 degrees, and (d) 271 to 360 degrees. The maximum aspect for a 3×3 window was calculated resulting in a 270 m dominant aspect product, which was resampled to 250 m using nearest neighbor. Using the 90 m recoded aspect data set, a focal variety aspect data set was produced by calculating the number of unique values within a 3×3 window, resulting in a 270 m product, which was resample to 250 m using nearest neighbor.

Soils (STATSGO), climate, and ecoregions variables were also included in the geo-spatial data set. The soils GIS layers were obtained from the National Resources Conservation Services (Miller and White, 1998). The climate data were obtained from DAYMET (Thornton *et al.*, 1997). The DAYMET variables used were annual and monthly average precipitation, monthly maximum and minimum temperature, and annual and monthly average temperature. The climate data resolution was 1 km. The climate data were rescaled to 250 m resolution using bilinear interpolation. For the conterminous U.S., Bailey's ecoregions (Bailey, 1989; Bailey and Hogg, 1986) were used and for Alaska, unified ecoregions of Alaska (Nowacki *et al.*, 2001) were used. The STATSGO, Bailey's ecoregions, and unified ecoregions of Alaska vector data were converted to raster data and scaled to 250 m.

The rest of the data layers consisted of Terra MODIS eight-day, Terra MODIS 32-day, and Terra MODIS-derived products such as enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), and Terra MODIS vegetation continuous fields from 2001 (Hansen *et al.*, 2003). To capture phenology changes, each of these MODIS products, except for the Terra MODIS 32-day imagery, came from three time periods: spring, summer, and fall. Because of the difficulty in finding cloud-free MODIS eight-day imagery, several 32-day MODIS images were used that covered spring, summer, and fall. Because of striping in band 5 in all the Terra MODIS eight-day and 32-day imagery, all Terra MODIS band 5 layers were excluded from analyses.

Many of these data layers are correlated. However, classification-trees can handle complex relationships between variables and can determine which data layers most accurately predict classes. Even though two data layers are correlated, there are differences between them. Classification-trees have the capability to use variations in data layers to develop models. The models that result are often complex. However, complex models are appropriate when the goal of a classification is accuracy rather than characterizing the relationships between the classes and the data layers. Consequently, correlated variables can be used.

All imagery, except for Alaska, was pre-stratified in order to prevent errors such as an eastern U.S. forest type occurring in the western U.S. The imagery was stratified using the USGS NLCD 65 mapping zones for the conterminous U.S. (Homer and Gallant, 2001). The reason why the NLCD mapping zones were chosen as opposed to Bailey's or Omernik ecoregions (Omernik, 1987) is because the NLCD mapping zones, which are loosely based on Omernik ecoregions, were developed specifically for remote sensing classification. The NLCD mapping zones takes into account the spectral variability occurring within each mapping zone. Thus, using NLCD mapping zones for pre-stratification seems more appropriate than other ecoregions. Alaska was modeled separately from the conterminous U.S. and was not further stratified. The reason why Alaska was not stratified is FIA data does not exist for large portions of Alaska. If Alaska was stratified, there would be many strata with few or no points.

Some evidence suggests that increasing sample sizes increases classification-tree accuracy (Pal and Mather, 2003). Seventeen NLCD mapping zones had fewer than 200 FIA forested plots. To increase the number of FIA plots for these NLCD mapping zones, these 17 zones were merged with adjacent NLCD mapping zones creating a total of 43 mapping zones (Figure 1). Instead of re-ordering the mapping zones from 1 to 43, the original NLCD mapping zones numbers were retained because the numbers are nominal data and frequent users of the NLCD mapping zones associate the numbers to specific geographic areas. When merging the 17 zones, the lowest NLCD mapping zone number was retained, while the other numbers were eliminated. Even though Figure 1 displays mapping zones 1 to 66, there are actually only 43 mapping zones. Each mapping zone and Alaska was modeled independently of each other.

For each mapping zone and Alaska, the FIA plot data were intersected with the geo-spatial data set creating a modeling data set. For the conterminous U.S., 83,519 FIA forested plots were used, and 5,392 FIA forested plots were used for Alaska. A random ten percent of the plots were withheld from model development and were used for accuracy assessment purposes.

Rulequest's See5® software package (**http://www.rulequest.com**), which is a commercial version of C4.5 (Quinlan, 1993), was used to develop the classification-trees for the forest type group and forest type classifications. Classification-trees recursively divide data into smaller groups on the basis of tests performed at the nodes in the trees. The tests used are learning algorithms developed within the pattern-recognition and machine-learning communities. At the ends of the trees, a value is assigned.

Boosting can significantly increase classification-trees accuracies (Friedl *et al.*, 1999; Pal and Mather, 2003). Boosting creates multiple iterations of classification-trees.

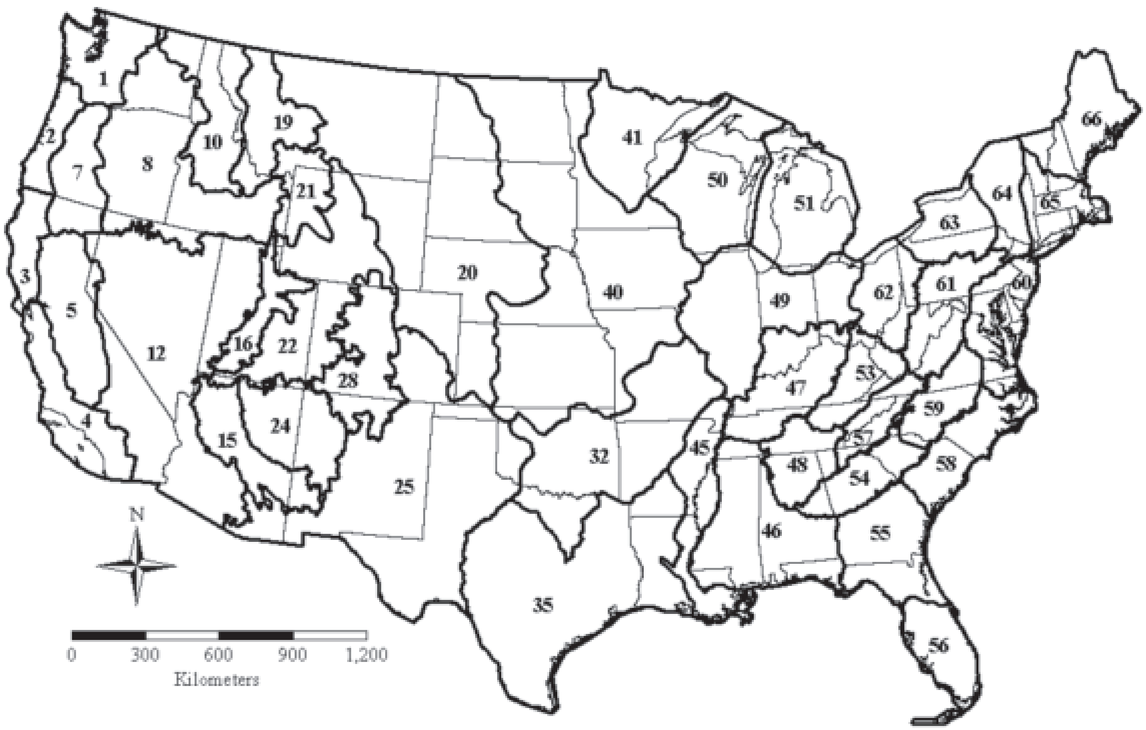

Figure 1. Forest Type Group and Forest Type Mapping Zones. The thicker black lines outline the 43 mapping zones, which were used to model forest type and forest type group for the conterminous United States. The mapping zone numbers are nominal, and thus, are not meant to be sequential.

For the first iteration, no weighting occurs. For all the other iterations, weights are assigned to each training observation. The weights assigned are based upon the misclassifications from the previous iteration. Each iteration tries to correct the errors from the previous iteration. Voting is used to generate the final classifier (DeFries and Chan, 2000; Freund and Schapire, 1996; Quinlan, 1996; Friedl *et al.*, 1999). Various studies have found 10 iterations to be the recommended number for both remote sensing studies and non-remote sensing studies (Friedl *et al.*, 1999; Freund and Schapire, 1997; Pal and Mather, 2003). For this study, the boosting option was set at 10 iterations.

Classification-trees can grow very large and complex, causing the classification-trees to overfit the training data. This can lead to poor accuracies when the classification-trees are applied if there were errors or noise in the training data. To alleviate this problem, classification-trees are pruned making them more general and flexible when classifying data not included in the training data set. The methodology See5® uses to prune is error-based pruning (Hall *et al.*, 2003; Mingers, 1989; Quinlan, 1987; Quinlan, 1993). To control the amount of pruning, the user can set the "pruning certainty factor." Changing this value affects the size and accuracy of the classification-tree. For this project the pruning certainty factor was set at 25, which is the recommended value determined by Quinlan (1993), and corroborated by Hall *et al.* (2003).

Forest type group and forest type classifications were created from the See5® classification-trees and the geospatial data set by integrating the See5® public domain code available from **http://www.rulequest.com** with ERDAS Imagine® software (version 8.6). The See5® public domain code has the ability to produce confidence values, which are expressions of the confidence of the classifications produced from the See5® models. Spatial confidence products for the forest type group and forest type were produced for the conterminous U.S. and Alaska. These spatial data products produced from See5®, covered the conterminous U.S. and Alaska regardless of the presence of forest. The non-forest areas were masked using the forest/non-forest mask produced by FIA scientists (Blackard *et al.*, 2008).

If a testing data set is specified, See5® creates standard error matrices (Congalton and Green, 1999). Using these matrices, overall accuracies and kappas were calculated for each zone and for Alaska. As an additional accuracy assessment, using FIA's mapmaker program (**http://ncrs2.fs.fed.us/4801/fiadb**) the current FIA state summaries of forest type group and forest type areas were compared to state summaries generated from the See5® modeled forest type group and forest type classifications.

Results and Discussion

The conterminous U.S. and Alaska forest type group classifications are shown in Plates 1 and 2. The conterminous U.S. and Alaska forest type classifications are not shown due to the amount of detail, but they along with the forest type group map, the confidence maps, associated metadata, and accuracy tables for each mapping zone are available at **http://fsgeodata.fs.fed.us**.

For the western U.S., the most abundant forest type group was pinyon/juniper group (22 million hectares) followed by Douglas-fir group (20 million hectares). The pinyon/juniper group occurred throughout the arid western U.S. and the

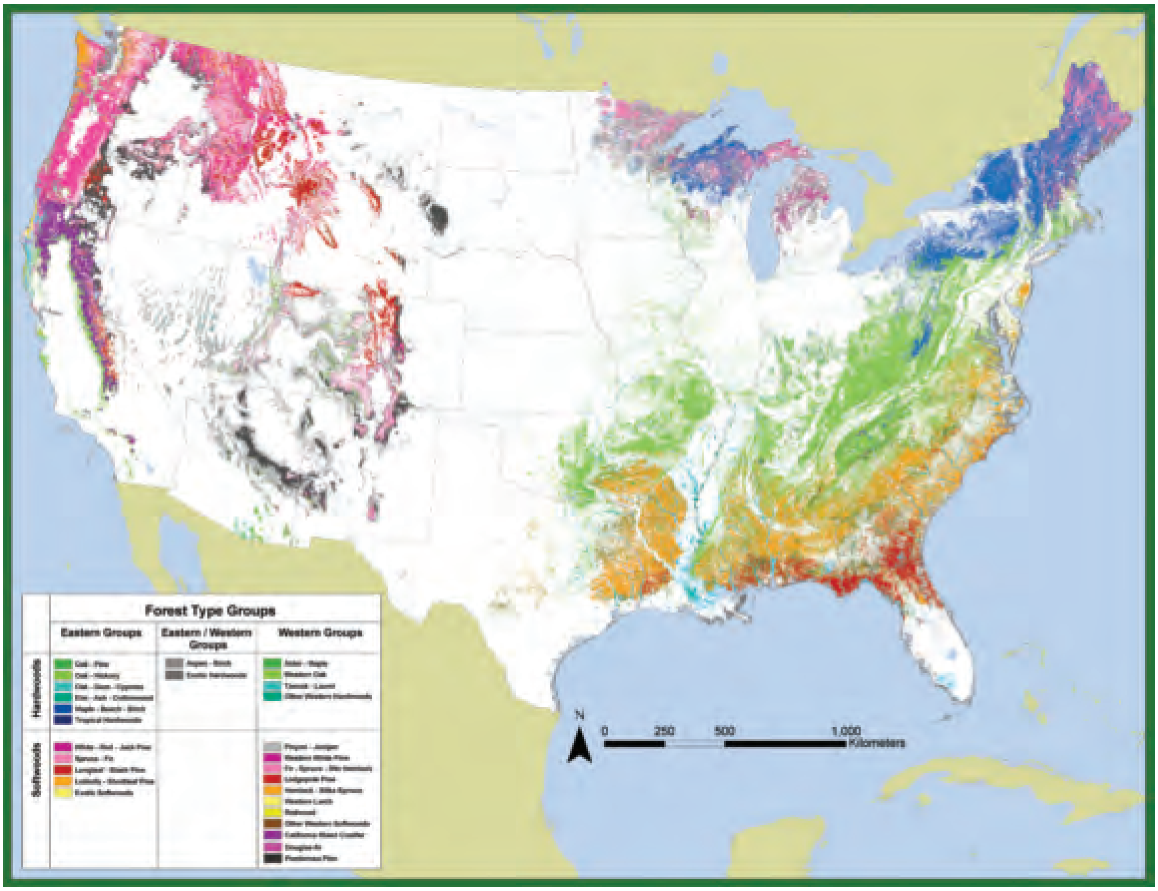

Plate 1. Forest type group map for the conterminous United States.

Douglas-fir group was prevalent in the Pacific Northwest and the northern Rocky Mountains. Other predominant western forest type groups included fir/spruce/mountain hemlock group (15 million hectares), ponderosa pine group (12 million hectares), and lodgepole pine group (7 million hectares).

The most abundant forest type in the western U.S. was Douglas-fir (22 million hectares) followed by pinyon/juniper woodland (18 million hectares). Other predominant western U.S. forest types were ponderosa pine (12 million hectares), lodgepole pine (7 million hectares), and California mixed conifer (5 million hectares). Note that the California mixed conifer forest type consists of a conglomerate of conifers including Douglas-fir (*Pseudotsuga menziesii*), ponderosa pine (*Pinus ponderosa*), sugar pine (*Pinus lambertiana*), incense cedar (*Libocedrus decurrens*), and white fir (*Abies concolor*). FIA only recognizes this type in California. Even though these tree species occur in neighboring states of California, the tree species are grouped into other types.

For the eastern U.S., the most abundant forest type group was oak/hickory group (67 million hectares) followed by loblolly/shortleaf pine group (31 million hectares), and maple/beech/birch group (24 million hectares). The oak/hickory group occurred throughout the eastern U.S., but primarily in the mid-eastern states. The loblolly/shortleaf pine group was predominant in the south and the maple/beech/birch group was predominant in the northeast. Other common eastern forest type groups included aspen/birch group (13 million hectares), oak/gum/cypress group (11 million hectares), and oak/pine group (8 million hectares). The most abundant forest type in the eastern U.S. was loblolly pine (34 million hectares) followed by white oak/red oak/hickory (29 million hectares).

For Alaska, the most abundant forest type group was spruce/fir group (49 million hectares) followed by aspen/birch group (8 million hectares) and hemlock/Sitka spruce group (4 million hectares). The spruce/fir group and aspen/birch group occurred throughout the interior of Alaska and hemlock/Sitka spruce group occurred mainly in southeastern Alaska. The most abundant forest type in Alaska was white spruce (28 million hectares) followed by black spruce (20 million hectares) and paper birch (7 million hectares). All of these forest types occurred primarily in the interior of Alaska.

For the conterminous U.S., most forest type groups and forest types had low confidence values. The forest type group with the highest confidence was the pinyon/juniper group, which had 80 percent of the pixels with greater than 70 percent confidence. The next highest was the oak/hickory group with 61 percent of the pixels with greater than 70 percent confidence. That was followed by the maple/beech/birch group with 54 percent of the pixels with greater than 70 percent confidence. The forest type with the highest confidence was pinyon juniper woodland with 76 percent of the pixels with greater than 70 percent confidence. The next highest forest type was mesquite with 58 percent of the pixels with greater than 70 percent confidence.

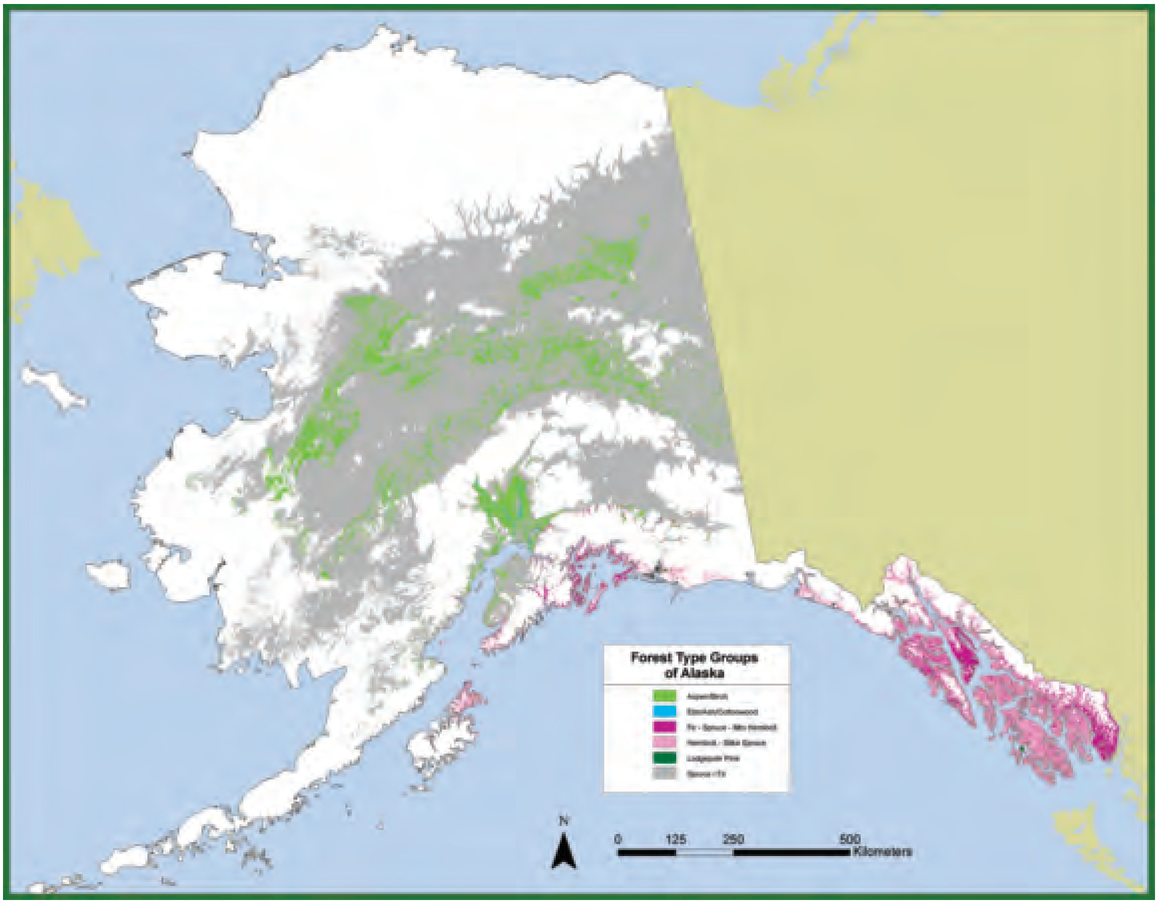

Plate 2. Forest type group map for Alaska.

For Alaska, the confidence values were higher. The forest type group with the highest confidence was the spruce/fir group, which had 85 percent of the pixels with greater than 70 percent confidence. The next highest was the hemlock/Sitka spruce group with 68 percent of the pixels greater than 70 percent confidence. That was followed by the fir/spruce/mountain hemlock group with 62 percent of the pixels greater than 70 percent confidence. The forest type with the highest confidence was Sitka spruce with 73 percent of the pixels with greater than 70 percent confidence. The next highest was mountain hemlock with 69 percent of the pixels with greater than 70 percent confidence. This was followed by white spruce with 64 percent of the pixels with greater than 70 percent confidence.

One of the advantages of using classification-trees is the ease of interpretation of the relationships between the independent variables and the dependent variable. For this project, the See5® output produced 88 classification-trees: a forest type group classification-tree and a forest type classification-tree for Alaska and for each of the 43 mapping zones. Each of these 88 classification-tree outputs consisted of ten classification-trees produced by the boosting option. Each individual classification tree consisted of more than 20 levels, and each of the 88 classification tree outputs are over 50 pages in length. Thus, the amount of data prohibits detailed examination of these classification-trees. For those interested in the models, they are available from the principal author. It is possible, however, to present general summaries of the mapping zones and Alaska classification-trees.

For the forest type groups occurring in the western U.S., topography variables were the most frequently used variables in the classification-trees followed by the spring and fall dates of the MODIS EVI imagery and percent tree cover from the MODIS vegetation continuous fields. It is well known that temperature and precipitation play vital roles in the distribution and growth of vegetation (Burns and Honkala, 1990a; Burns and Honkala, 1990b). Topography in the western U.S. has a strong influence on precipitation and temperature. Topography variables should be prevalent in the classification-trees for the western U.S. mapping zones.

Topographical relief in the eastern U.S. is relatively minor compared to the western U.S., and climate is less influenced by topography in the eastern U.S. Variables associated with plant growth such as climate and soils should be used more frequently in the classification-trees for the eastern U.S. than the topography variables. However, instead of the climate variables being the most frequently used, the spring, summer, and fall dates of the MODIS EVI imagery as well as the fall date of the MODIS NDVI image were the most frequently used variables in the forest type group classification-trees for the eastern U.S. Topography variables were the second most frequently used variables in the classification-trees. NDVI indicates density of plant growth and is frequently used in vegetation mapping. EVI is similar to NDVI, but EVI optimizes the vegetation signal improving sensitivity in high plant density regions. The differences in the EVI between the spring, summer, and fall could be surrogates for climatic data and this could explain the abundance of these variables in the forest type group classification-trees for the eastern U.S.

The forest type classification-trees for the western U.S. showed similar variable usage as the forest type group classification-trees for the western U.S. The topography variables were the most frequently used variables in the classification-trees followed by the spring, summer, and fall MODIS EVI imagery.

The topography variables were the most frequently used variables for the forest type classification-trees for the eastern U.S. This was unexpected because the topography variables were not expected to have a large influence in the eastern U.S. The second most frequently used variables for the forest type classification-trees for the eastern U.S. was the spring, summer, and fall MODIS EVI imagery, which were also significant variables for the forest type group classification-trees for the eastern U.S.

For Alaska, the most frequently used variables for both the forest type group classification-trees and the forest type classification-trees were the topography variables. Alaska does have high topographical relief. The elevation ranges from 6,200 m (20,000 feet) to 0 m. Topography is expected to have a large influence in the forest type groups and forest type. The second most frequently used variable in both classification-trees was the unified ecoregions.

Table 2 shows the forest type group and forest type accuracies along with the associated kappas for each mapping zone. The highest forest type accuracy was 81 percent and the lowest was 34 percent. All of the mapping zones, except for one, with forest type accuracy greater than or equal to 75 percent occurred in the sparsely forested arid western U.S. All of the mapping zones with forest type accuracy less than 40 percent occurred in the northeast U.S., an area high

Table 2. Forest Type Group and Forest Type Overall Accuracies and Kappas for the 43 Mapping Zones and Alaska. The Mapping Zones Refer to the Mapping Zones in Figure 1

Zone	Forest Type Group Overall Accuracy	Kappa	Forest Type Overall Accuracy	Kappa
Alaska	78%	0.69	67%	0.61
1	66%	0.50	56%	0.38
2	66%	0.24	66%	0.19
3	52%	0.36	55%	0.40
4	71%	0.42	41%	0.25
5	60%	0.46	51%	0.38
7	69%	0.61	64%	0.56
8	61%	0.49	57%	0.47
10	54%	0.33	53%	0.37
12	92%	0.63	81%	0.54
15	78%	0.59	73%	0.60
16	62%	0.55	49%	0.42
19	60%	0.44	63%	0.49
20	75%	0.57	75%	0.59
21	62%	0.50	57%	0.46
22	86%	0.67	79%	0.64
24	88%	0.34	81%	0.48
25	80%	0.67	73%	0.60
28	66%	0.60	55%	0.48
32	79%	0.52	56%	0.38
35	70%	0.56	63%	0.47
40	67%	0.47	47%	0.34
41	70%	0.58	62%	0.48
45	66%	0.45	47%	0.35
46	55%	0.40	51%	0.38
47	72%	0.34	60%	0.44
48	69%	0.52	53%	0.47
49	52%	0.18	42%	0.23
50	66%	0.55	56%	0.50
51	66%	0.46	54%	0.38
53	86%	0.10	51%	0.41
54	79%	0.67	77%	0.67
55	75%	0.67	66%	0.60
56	59%	0.42	45%	0.29
57	84%	0.52	42%	0.31
58	68%	0.54	62%	0.49
59	70%	0.54	62%	0.56
60	65%	0.46	50%	0.41
61	79%	0.45	34%	0.16
62	69%	0.17	37%	0.09
63	66%	0.27	35%	0.13
64	67%	0.32	43%	0.19
65	64%	0.46	37%	0.23
66	63%	0.42	39%	0.09

in tree species diversity. The eastern U.S. has a variety of pines and hardwoods, which are probably difficult to distinguish spectrally at a scale of 250 meters. The forest type group and forest type classification schemes were not designed for spectral analysis. There are forest types that are pure forest stands of a species, such as the longleaf pine forest type. These same species can also occur in mixed forests and be assigned a different forest type, such as the longleaf pine/oak forest type. These aforementioned forest types and others like them certainly reduced the accuracy. Forest type group and forest type classification schemes that considers spectral separability between classes would greatly improve the accuracy of these maps.

The highest forest type group mapping zone accuracy was 92 percent and the lowest was 52 percent. The forest type group mapping zone accuracies were always higher than the forest type mapping zone accuracies except for four mapping zones. The forest type groups are probably more easily distinguishable spectrally than the forest types. The forest type group accuracies showed no west to east trend of increasing accuracy.

The overall forest type group and forest type accuracies for the conterminous U.S. were 69 percent and 50 percent, respectively. For Alaska, the overall forest type group and forest type accuracies were 78 percent and 67 percent, respectively. To further verify the results, the total area for each forest type group and forest type was calculated for each state. These state area summaries were compared to FIA state summary tables. For the conterminous U.S., eight percent (40 out of 512) of the forest type groups and four percent (70 out of 1,777) of the forest types differed by more than ten percent in area between the FIA state summaries and the classification area estimates. For Alaska, three of the seven forest type groups and five of the 15 forest types differed by more than ten percent in area between the FIA state summaries and the classification area estimates. Thus, these forest type group and forest type mapping products compares favorably with other data sets at least at the state level.

The forest type group map produced by this project was compared to the forest type group map produced by Zhu and Evans (1994). Table 3 shows the percent agreement between the two forest type group maps. The oak/hickory forest type group had the high percent agreement at 74.38 percent and the western white pine forest type group had the lowest percent agreement at 0.01 percent. The western white pine forest type group is associated with forest types found within the Douglas-fir and the fir/spruce/mountain hemlock forest type groups (Forest Inventory and Analysis, 2004). The forest type group map produced by this project classified the western white pine forest type group into these other associated type groups. This occurred with all the other forest type groups with disagreements; the forest type groups were classified into other associated forest type groups. The Zhu and Evans (1994) forest type group map had a spatial resolution of 1 km. This new forest type group map with a spatial resolution of 250 m is able to distinguish finer differences in the forest type groups.

TABLE 3. PERCENT AGREEMENT BETWEEN THE ZHU AND EVANS (1994) FOREST TYPE GROUP MAP AND THE NEW FOREST TYPE GROUP MAP

Forest Type Group Name	Percent Agreement
Oak/Hickory	74.38%
Pinyon/Juniper	69.01%
Maple/Beech/Birch	64.16%
Fir/Spruce/Mountain Hemlock	58.45%
Douglas-fir	57.59%
Loblolly/Shortleaf Pine	56.87%
Aspen/Birch	46.90%
Longleaf/Slash Pine	45.30%
Spruce/Fir	39.31%
Oak/Gum/Cypress	34.89%
Ponderosa Pine	32.70%
Hemlock/Sitka Spruce	32.24%
Lodgepole Pine	31.70%
Redwood	27.51%
Elm/Ash/Cottonwood	16.16%
White/Red/Jack Pine	15.64%
Oak/Pine	8.53%
Western Larch	3.58%
Other Western Hardwoods	2.21%
Western White Pine	0.01%

Conclusions

The forest type map is the first national forest type map. This map along with the forest type group map can be used for many different applications such as assisting in pre-stratification for other vegetation mapping projects, habitat analyses, and assisting policy and decision makers. One application for the forest type group map involves the updating of the forest risk maps. The Zhu and Evans (1994) forest type group map was one of the key components for the development of the forest risk maps published in 2000 by the USFS Forest Health Monitoring (FHM) program (Lewis, 2002). The forest risk maps were non-site-specific and identified broad areas that had potential high risk of forest mortality or growth/volume loss from insects and diseases. The intent of these forest risk maps was to provide national scale information to policy makers to help determine national priorities. The forest type group map developed as part of this project will also aid in the development and improvement of forest risk maps. Because this new forest type group map is at a higher resolution (250 m) than the Zhu and Evans (1994) forest type group map, which had a resolution of 1 km, this new forest type group map will allow future forest risk maps to be used more specifically at larger scales and finer grain analyses.

This project effectively demonstrated the possibility of deriving national mapping products using FIA data and a geo-spatial database. Software programs were written to facilitate communication between ERDAS Imagine and the See5® software, which greatly enhanced the development of these national mapping products. See5® produced highly accurate forest type group maps. The forest type maps were less accurate because of some of the difficulty in separating out types which have similar spectral signatures. The forest type maps were still fairly accurate especially at the state level where the area of the forest types were similar to the area estimated by FIA.

Acknowledgments

We would like to thank Michael Hoppus, Rachel Riemann, Dennis Jacobs, and Gretchen Moisen for their intellectual contributions and support. We also are grateful for the contributions and support of the USDA Forest Service Forest Inventory and Analysis and the Forest Health Protection programs. We appreciate the insightful comments of three anonymous reviewers, which greatly improved the manuscript.

References

Arner, S.L., S. Woudenberg, S. Waters, J. Vissage, C. MacLean, M. Thompson, and M. Hansen, 2003. *National Algorithms for Determining Stocking Class, Stand Size Class, and Forest*

Type for Forest Inventory and Analysis Plots, U.S. Department of Agriculture Forest Inventory and Analysis, Washington, D.C., 65 p.

Bailey, R.G., 1989. Explanatory supplement to ecoregions map of the continents, *Environmental Conservation*, 16(4):307–309.

Bailey, R.G., and H.C. Hogg, 1986. A world ecoregions map for resource reporting, *Environmental Conservation*, 13(3):195–202.

Blackard, J., M. Finco, E. Helmer, G. Holden, M. Hoppus, D. Jacobs, A. Lister, G. Moisen, M. Nelson, R. Riemann, B. Ruefenacht, D. Salajanu, D. Weyermann, K. Winterberger, T. Brandeis, R. Czaplewski, R. McRoberts, P. Patterson, R. Tymcio, 2008. Mapping U.S. forest biomass using nationwide forest inventory data and Terra MODIS-based information, *Remote Sensing of Environment*, 112(4):1658–1671.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, 1984. *Classification and Regression Trees*, Wadsworth International Group, Belmont, California, 358 p.

Burns, R.M., and B.H. Honkala, 1990a. *Silvics of North America: 1. Conifers, Agriculture Handbooks 654*, U.S. Department of Agriculture, Forest Service, Washington, D.C., 675 p.

Burns, R.M., and B.H. Honkala, 1990b. *Silvics of North America: 2. Hardwoods, Agriculture Handbooks 654*, U.S. Department of Agriculture, Forest Service, Washington, D.C., 877 p.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, New York, 137 p.

DeFries, R.S., and J.C. Chan, 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data, *Remote Sensing of Environment*, 74(3):503–515.

DeFries, R., M. Hansen, J.R.G. Townshend, and R. Sohlberg, 1998. Global land cover classifications at 8 km spatial resolution: The use of training data derived from Landsat imagery in decision tree classifiers, *International Journal of Remote Sensing*, 19(16):3141–3168.

Eyre, F.H., 1980. *Forest Cover Types of the United States and Canada*, Society of American Foresters, Bethesda, Maryland, 148 p.

Forest Inventory and Analysis, 2004. *Forest Inventory and Analysis National Core Field Guide, Volume 1: Field Data Collection Procedures for Phase 2 Plots*, version 2.0, U.S. Department of Agriculture Forest Inventory and Analysis, Washington, D.C., 208 p.

Forest Health Protection, 2004. *Forest Insect and Disease Conditions in the United States 2003*, U.S. Department of Agriculture Forest Health, Washington, D.C., 156 p.

Freund, Y., and R.E. Schapire, 1996. Experiments with a new boosting algorithm, *Machine Learning Proceedings of the Thriteenth International Conference*, Morgan-Kaufman, San Francisco, California, pp. 148–156.

Freund, Y., and R.E. Schapire, 1997. A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, 55(1):119–139.

Friedl, M.A., and C.E. Brodley, 1997. Decision tree classification of land cover from remotely sensed data, *Remote Sensing of the Environment*, 61(3):399–409.

Friedl, M.A., C.E. Brodley, and A.H. Strahler, 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales, *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):969–977.

Hall, L.O., K.W. Bowyer, R.E. Banfield, S. Eschrich, and R. Collins, 2003. Is error-based pruning redeemable?, *International Journal on Artificial Intelligence Tools*, 12(3):249–264.

Hansen, M., R. DeFries, J.R. Townshend, M. Carroll, C. Dimiceli, and R. Sohlberg, 2003. *Vegetation Continuous Fields MOD44B, 2001 Percent Tree Cover, Collection 3*, University of Maryland, College Park, Maryland, 2001.

Hansen, M., R. Dubayah, and R. DeFries, 1996. Classification-trees: An alternative to traditional land cover classifiers, *International Journal of Remote Sensing*, 17(5):1075–1081.

Homer, C.G., and A. Gallant, 2001. *Partitioning the Conterminous United States into Mapping Zones for Landsat TM Land Cover Mapping*, U.S. Geological Survey, Sioux Falls, South Dakota, 7 p.

Lawrence, R., A. Bunn, S. Powell, and M. Zambon, 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis, *Remote Sensing of Environment*, 90(3):331–336.

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

Lewis, J.W., 2002. *Mapping Risk from Forest Insects and Diseases, FS-754*, U.S. Department of Agriculture Forest Service, Washington, D.C., 62 p.

Lillesand, T.M., and R.W. Kiefer, 2000. *Remote Sensing and Image Interpretation*, John Wiley and Sons, Inc., New York, 724 p.

Miller, D.A., and R.A. White, 1998. A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling, *Earth Interactions*, 2(2):1–26.

Mingers, J., 1989. An empirical comparison of pruning methods for decision tree induction, *Machine Learning*, 4(2):227–243.

Omernik, J.M., 1987. Ecoregions of the conterminous United States: Map (scale 1:7,500,000), *Annals of the Association of American Geographers* 77(1):118–125.

Pal, M., and P.M. Mather, 2003. An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment*, 86(4):554–565.

Quinlan, J.R., 1987. Simplifying decision trees, *International Journal of Man-Machine Studies*, 27(3):221–234.

Quinlan, J.R., 1993. *C4.5: Programs for Machine Learning*, Morgan-Kaufman, San Mateo, California, 302 p.

Quinlan, J.R., 1996. Bagging, boosting and C4.5, *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, American Association for Artificial Intelligence Press, Portland, Oregon, pp. 725–730.

Thornton, P.E., S.W. Running, and M.A. White, 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain, *Journal of Hydrology*, 190:214–251.

Vogelmann, J.E., S.M. Howard, L. Yang, C.R. Larson, B.K. Wylie, and J.N. Van Driel, 2001. Completion of the 1990's National Land Cover Data Set for the conterminous United States, *Photogrammetric Engineering & Remote Sensing*, 67(6):650–662.

Zhu, Z., and D.L. Evans, 1994. U.S. forest types and predicted percent forest cover from AVHRR data, *Photogrammetric Engineering & Remote Sensing*, 60(5):525–531.

(Received 06 June 2006; accepted 17 August 2006; revised 02 March 2007)