FOREST TYPE MAPPING OF THE INTERIOR WEST

Bonnie Ruefenacht, Remote Sensing Analyst
Remote Sensing Applications Center
USDA Forest Service
2222 West 2300 South
West Valley City, Utah 84119
bruefenacht@fs.fed.us

Gretchen G. Moisen, Research Forester
Jock A. Blackard, Forester
Forest Inventory and Analysis
Rocky Mountain Research Station
USDA Forest Service
507 25th Street
Ogden, UT 84401
gmoisen@fs.fed.us
jablackard@fs.fed.us

ABSTRACT

This paper develops techniques for the mapping of forest types in Arizona, New Mexico, and Wyoming. The methods involve regression-tree modeling using a variety of remote sensing and GIS layers along with Forest Inventory Analysis (FIA) point data. Regression-tree modeling is a fast and efficient technique of estimating variables for large data sets with high accuracy levels. If the methods developed in this paper are successful, they will be applied to the contiguous United States and Alaska producing a forest type map for these areas. This forest type map will update and improve an older version of the forest type map made in 1992.

INTRODUCTION

A forest type map for the entire United States and including Puerto Rico was created in 1992 (Zhu and Evans 1994). This forest type map was created with Advanced Very High Resolution Radiometer (AVHRR) composite images with a spatial resolution of 1 km collected during the 1991 growing season. The USDA Forest Service Forest Inventory Analysis (FIA) unit is creating a new version of this forest type map hereafter referred to as the National Forest Type Map 2004. The National Forest Type Map 2004 will be created using primarily Moderate Resolution Imaging Spectroradiometer (MODIS) composite images with a spatial resolution of 250 m collected between 2002 and 2003. This forest type map will not only be at a higher spatial resolution, but will also have more classes than the 1992 version. The 1992 forest type map had 25 classes, while the National Forest Type Map 2004 will potentially have 138 classes. A list of the forest type classes is in Table 1.

The Forest Health Monitoring (FHM) and Forest Inventory and Analysis (FIA) programs both require more current geospatial forest type information for national-scale analysis. In the case of FHM, the forest type data will be a key input to the next version of the December 2002 report "Mapping Risk from Forest Insects and Diseases". This report was prepared using local knowledge, FIA plot data, other plot data, and forest type information from the 1992 RPA Forest Resource Assessment. Having consistent national scale forest type group information is important to model forest areas at risk of increased mortality due to insects and diseases. Areas at risk were defined as forested stands that are expected to have 25% or greater mortality above normal conditions within the next 15 years.

White/Red/Jack Pine Group
Jack Pine
Red Pine
Eastern White Pine
Eastern White Pine/Eastern Hemlock
Eastern Hemlock

Douglas-fir Group
Douglas-fir
Port-Orford-Cedar

Lodgepole Pine Group
Lodgepole Pine

Other Western Softwoods Group
Knobcone Pine
Southwest White Pine
Bishop Pine
Monterey Pine
Foxtail Pine/Bristlecone Pine
Limber Pine
Whitebark Pine
Misc. Western Softwoods
California Mixed Conifer Group
California Mixed Conifer

Maple/Beech/Birch Group
Sugar Maple/Beech/Yellow Birch
Black Cherry
Cherry/Ash/Yellow-poplar
Hard Maple/Basswood
Elm/Ash/Locust
Red Maple/Upland

Other Western Hardwoods Group
Pacific Madrone
Mesquite Woodland
Cercocarpus Woodland
Intermountain Maple Woodland
Misc. Western Hardwood Woodlands

Pinyon/Juniper Group
Rocky Mountain Juniper
Western Juniper
Juniper Woodland
Pinyon Juniper Woodland

Spruce/Fir Group
Balsam Fir
White Spruce
Red Spruce
Red Spruce/Balsam Fir
Black Spruce
Tamarack
Northern White-cedar

Ponderosa Pine Group
Ponderosa pine
Incense Cedar
Jeffrey Pine/Coulter Pine/Bigcone Douglas-fir
Sugar Pine

Exotic Softwoods Group
Scotch Pine
Australian Pine
Other Exotic Softwoods
Norway Spruce
Introduced Larch

Oak/Gum/Cypress Group
Swamp Chestnut Oak/Cherrybark Oak
Overcup Oak/Water Hickory
Atlantic White-cedar
Baldcypress/Water Tupelo
Sweetbay/Swamp Tupelo/Red Maple

Aspen/Birch Group
Aspen
Paper Birch
Balsam Poplar

Tropical Hardwoods Group
Stable Palm
Mangrove
Other Tropical

Tanoak/Laurel Group
Tanoak
California Laurel
Giant Chinkapin

Longleaf/Slash Pine Group
Longleaf Pine
Slash Pine

Western White Pine Group
Western White Pine

Fir/Spruce/Mountain Hemlock Group
White Fir
Red Fir
Noble Fir
Pacific Silver Fir
Engelmann Spruce
Engelmann Spruce/Subalpine Fir
Grand Fir
Subalpine Fir
Blue Spruce
Mountain Hemlock
Alaska-Yellow-Cedar

Elm/Ash/Cottonwood Group
River Birch/Sycamore
Cottonwood
Willow
Sycamore/Pecan/American Elm
Sugarberry/Hackberry/Elm/Green Ash
Red Maple/Lowland
Cottonwood/Willow
Oregon Ash

Alder/Maple Group
Red Alder
Bigleaf Maple

Exotic Hardwoods Group
Paulownia
Melaluca
Eucalyptus
Other Exotic Hardwoods

Western Larch Group
Western Larch

Redwood Group
Redwood
Giant Sequoia

Loblolly/Shortleaf Pine Group
Loblolly Pine
Shortleaf Pine
Virginia Pine
Sand Pine
Table Mountain Pine
Pond Pine
Pitch Pine
Spruce Pine
Eastern Redcedar

Hemlock/Sitka Spruce Group
Western Hemlock
Western Redcedar
Sitka Spruce

Oak/Pine Group
Eastern White Pine/Northern Red Oak/White Ash
Eastern Redcedar/Hardwood
Longleaf Pine/Oak
Shortleaf Pine/Oak
Virginia Pine/Southern Red Oak
Loblolly Pine/Hardwood
Slash Pine/Hardwood
Other Pine/Hardwood

Western Oak Group
Gray Pine
California Black Oak
Oregon White Oak
Blue Oak
Deciduous Oak Woodland
Coast Live Oak
Canyon Live Oak/Interior Live Oak

Oak/Hickory Group
Post Oak/Blackjack Oak
Chestnut Oak
Red Oak/White Oak/Hickory
White Oak
Northern Red Oak
Yellow-popular/White Oak/Northern Red Oak
Sassafras/Persimmon
Sweetgum/Yellow-poplar
Bur Oak
Scarlet Oak
Yellow Poplar
Black Walnut
Black Locust
Southern Scrub Oak
Chestnut Oak/Black Oak/Scarlet Oak
Red Maple/Oak
Mixed Upland Hardwoods

Table 1. Forest type groups and forest type classes used for the National Forest Type Map 2004.

In the case of FIA, the forest type data is expected to enhance regional and statewide analyses of forest inventory data by depicting current forest type distributions in a consistent fashion across the country. In addition, the data could be a valuable input to a second generation of forest characteristics maps. For instance, the FIA Remote Sensing Band (RSB) is currently developing a national geospatial dataset of forest biomass. This dataset is being produced using biomass that is estimated from FIA plot data and a host of spatially continuous predictor layers. Subsequently, FIA RSB hopes to use the forest type product to improve biomass mapping in a second-generation biomass product.

As part of this mapping effort, techniques and methodologies were developed for forest type mapping of Arizona, New Mexico, and Wyoming. This paper describes this effort. Some methods were successful while others were not. The methods that worked will be used in the forest type mapping of the nation.

METHODS

The forest type map is being created using regression-tree methodology (Breiman et al., 1984). Regression-trees recursively divides data into smaller groups on the basis of tests performed at each node in the tree. The tests used are learning algorithms developed within the pattern-recognition and machine-learning communities. Regression-tree classification procedures have several advantages over more traditional classification procedures such as supervised and unsupervised algorithms based upon maximum likelihood (Lillesand and Kiefer 2000). Regression-trees are non-parametric and, as such, do not require knowledge 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 multiple remote sensing and GIS data layers. With traditional classification methods, it is often desirable to use separability measures, such as principal component analysis, to understand the effects of different data layers on classifications. This process can be very time consuming and labor intensive. With regression-tree classifications, however, the hierarchal structure makes interactions between data layers easy to interpret. Regression-tree classifications are significantly less labor intensive than other classification techniques and can be used efficiently for large land cover classifications. With quality training data, the accuracies of regression-tree classifications are either similar to or better than supervised/unsupervised classification (Lawrence and Wright 2001, Friedl and Brodley 1997, Hansen et al. 1996).

The regression-tree software package used for the forest type mapping was See5 (www.rulequest.com). See5 is a fairly simple program with few options as seen in the See5 interface (Figure 1). One of these options, boosting, has been shown to improve the accuracy in land cover classification (Chan et al., 2001). The boosting option creates additional decision-trees by resampling with replacement from the initial data set. Each additional decision-tree tries to correct the predictions from the previous decision-tree by resampling observations that have been misclassified in the previous model with a higher probability of selection. The final model prediction is then decided by applying a plurality voting scheme to the multiple decision-tree predictions. Boosting with 10 trials was applied in the forest type mapping of Wyoming, Arizona, and New Mexico. In addition, See5's pruning option was used to prevent overfitting the model.

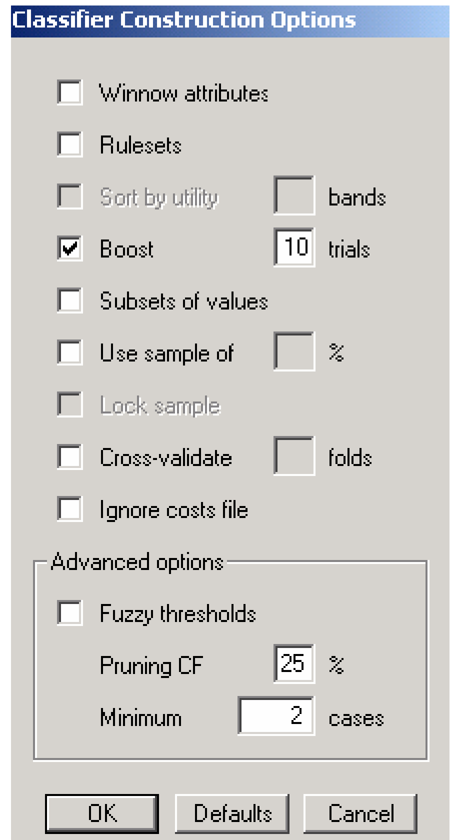

Figure 1. See5 interface.

For each data set analyzed, a random sample comprised of 50% of the data was set aside to be used for accuracy assessment purposes. See5 can apply a testing data set to the models built using the training data set and the results are shown in a confusion matrix. This is how the accuracy assessments were done.

See5 was designed to handle large data sets, but was not designed to handle remote sensing and GIS data layers. Tools within ERDAS Imagine were developed that converts remote sensing and GIS data layers to a See5 data format. Tools were also developed in ERDAS Imagine that applies the See5 models to create spatial data layers.

Predictor Variables

The predictor variable database currently consists of 269 remote sensing and GIS data layers. Table 2 provides a summary of these data layers. All of these predictor variables were resampled to 250 m and reprojected to Albers Equal Area projection. If the data was continuous, such as satellite imagery, a cubic convolution filter was used for resampling. If the data was thematic, such as aspect, a nearest neighbor filter was used for resampling. These data layers are freely available to the USDA Forest Service and can be obtained here: ftp://fsweb.rsac.fs.fed.us/fia_ftp/.

The primary remote sensing data used was MODIS. Three dates of eight-day MODIS composite imagery were collected. The start dates of the composites were 7 Apr 2002, 13 Jul 2002, and 17 Oct 2002. All of these composites included clouds; there were no eight-day MODIS composites that were cloud-free. Thus, additional 32-day MODIS composite imagery were collected. All the 32-day MODIS composites that were cloud-free during the

Table 2. Summary of Predictor Variables Used for Forest Type Mapping

- All STATSGO data layers available for the continental U.S. These include available water capacity, soil bulk density, soil permeability, soil ph, soil porosity, soil plasticity, soil depth to bedrock, rock volume, soil types, and soil texture.
- Ecological polygon layers such as the USGS Mapping zones, Bailey's Ecoregions, and Unified Ecoregions for Alaska
- Elevation, slope, and aspect
- MODIS Vegetation Indices layers such as EVI, NDVI.
- MODIS Vegetation Continuous Fields including percent tree cover, percent herbaceous cover, and percent bare ground
- MODIS fire points for 2001 and 2002 developed from the MODIS Active Fire Maps
- Several dates of MODIS 8-day composite imagery including imagery from the spring, summer, and fall.
- All dates of MODIS 32-day composite imagery that were cloud-free between the years 2001-2003.
- USGS NLCD layers
- Temperature and Precipitation PRISM layers including minimums, maximums, and averages.

years 2001-2003 were used. The start dates for these composites were: 7 Apr 2001, 12 Jul 2001, 9 May 2002, 13 Jul 2002, 14 Sep 2002, 17 Oct 2002, and 10 Jun 2003.

Additional MODIS products such as NDVI, EVI, and VCF were used as predictor variables. The MODIS NDVI and EVI products were 16-day composites obtained for the same starting dates as the 8-day MODIS composites. NDVI stands for normalized difference vegetation index and is a vegetation enhancement index commonly used in remote sensing. EVI stands for enhanced vegetation index. EVI reduces atmospheric influences, which allows for increase sensitivity to high biomass regions. VCF stands for vegetation continuous fields. VCF currently comes in three layers: percent tree cover, percent herbaceous cover, and percent bare ground. These layers are MODIS derived layers developed from a supervised regression tree algorithm (Hansen et al., 2003).

All the soil layers available for the contiguous U.S. were obtained from www.essc.psu.edu. These soil layers were developed by the National Resources Conservation Service (NRCS). The layers included available water capacity, soil bulk density, soil permeability, soil ph, soil porosity, soil plasticity, depth of soil to bedrock, rock volume, soil types, and soil texture.

Elevation was obtained from the USGS National Elevation Dataset (NED) (gisdata.usgs.net/NED). The spatial resolution of NED is 30 m. The dataset was resampled to 90 m using a cubic convolution filter. Using a 3x3 window, a mean elevation was calculated resulting in a 270 m mean elevation data layer. This data layer was resampled to 250 m. Aspect and slope were derived from this elevation data layer.

Eight classes of the USGS 1990 National Land Cover Data (NLCD) (landcover.usgs.gov) were used as predictor variables. These classes include deciduous forest (41), developed (21-23), evergreen forest (42), mixed forest (43), shrubland (51), and woody wetland (91); the numbers in parentheses are the NLCD class codes. The NLCD is a 30 m spatial resolution product. To convert the NLCD to a 250 m product, each code was converted to a single binary layer. A 9x9 window was used to calculate focal sums on each layer resulting in 270 m layers. The pixels were divided by 81 to obtain a percentage and the layers were resampled to 250 m.

Additional layers included in the predictor data base include monthly and average precipitation data layers (www.ocs.orst.edu/prism), maximum, minimum, and average temperature data layers (ww.ocs.orst.edu/prism), MODIS fire points for 2001 and 2002 developed from the MODIS Active Fire Maps, and Bailey's ecoregions.

Response Variables

FIA plot data with actual coordinates for Arizona, New Mexico, and Wyoming were used for the development of the regression-tree training and testing data sets. The FIA data were collected between 1984 and 2003 with the majority of the data (80%) collected between 2000 and 2003. The primary attributes used for modeling were analytical forest type, analytical forest type group, field forest type, and field forest type group. The field crews determine the field forest type at the time the plot is sampled. The analytical forest type is calculated after the field work is completed and is based upon a complex stocking algorithm. In addition, total basal area, basal area by species, crown cover, net growth, net volume, stand age, quadratic mean diameter, stand density index, trees per acre, and tree height were also modeled as potential predictors in forest type models. Analytical and field forest types were also collapsed to more general analytical and field forest type groups, respectively (Foresty Inventory and Analysis National Core Field Guide, 2004 fia.fs.fed.us).

RESULTS AND DISCUSSION

Several modeling trials were conducted. One of the modeling trials looked at the issue of scale. Specifically, at what scale should be the modeling be conducted: regional scale (e.g. the interior west), state scale, USGS mapping zone scale (Homer and Gallant, 2001), or Bailey's ecoregions scale? Because the number of plots available for each Bailey's ecoregion was insufficient for the modeling process, Bailey's ecoregions were not used. However, analytical forest type was modeled at the regional, statewide and mapping zone scales, with overall accuracies are shown in Table 3. The conclusion is that, on average, the modeling is more accurate at the USGS mapping zone scale than at any of the other scales. Therefore, all further modeling was done at the USGS mapping zone scale.

64%	Region (Arizona, New Mexico, and Wyoming combined)
68%	Arizona
65%	New Mexico
52%	Wyoming
70%	USGS Zone 14 - Arizona
64%	USGS Zone 15 - Arizona, New Mexico
50%	USGS Zone 21 - Wyoming
61%	USGS Zone 22 - Wyoming
80%	USGS Zone 24 - Arizona, New Mexico
72%	USGS Zone 25 - Arizona, New Mexico
70%	USGS Zone 27 - New Mexico
76%	USGS Zone 29 - Wyoming

Table 3. Overall accuracies of analytical forest type modeled at different scales.

Moisen et al. (2003) analyzed forest type at a spatial resolution of 30 m for USGS mapping zone 16 (Utah). They suggest an alternative approach to modeling forest type by first modeling basal area by species or species group, then assigning forest type based on basal area majority. For this project, basal area by species was modeled using Cubist, which is a regression-tree program very similar to See5 except Cubist is used for modeling continuous variables whereas See5 is used for modeling discrete variables. Initially, basal area by species was modeled by mapping zones. However, sample sizes were low and accuracies were also low. To attempt to improve the results, basal area by species was modeled at the regional scale. The accuracies were also low at this scale. Because of the low accuracies, the basal area by species approach to forest type mapping was discontinued.

Other FIA variables, which were mentioned previously, were modeled to see contribution they would make if used as predictors in the forest type models.. All of these variables are continuous variables and, thus, were modeled using Cubist. The accuracies for these variables varied quite a bit. Some variables had good accuracy while others had terrible accuracy. Despite these variables results, analytical forest type was modeled using these modeled FIA variables instead of the predictor layers. The analytical forest type accuracy was slightly lower than the analytical forest type accuracy obtained by modeling with the predictor layers. Thus, the use of these other FIA variables was discontinued.

Since the analytical forest type attribute is difficult to understand how it is derived, a new forest type attribute was created. This new forest type attribute was determined by the dominate species basal area. The accuracy for this new variable was considerably lower than the accuracy for the analytical forest type. This method was discontinued.

Analytical forest type and field forest type were modeled using just the predictor data set and the original FIA data. Compared to all of the other trials mentioned previously, this gave the most accurate results. Field forest type and field forest type group were usually more accurate than the analytical forest type and the analytical forest type group. Also, the accuracies of the forest type groups for both analytical forest type and field forest type were usually higher than the accuracies of the forest types. These results are shown in Table 4 and Figures (1-3).

	Overall Accuracies			
	Analytical Forest Type	Field Forest Type	Analytical Forest Type Group	Field Forest Type Group
Zone 14	70%	74%	93%	86%
Zone 15	64%	67%	75%	78%
Zone 21	51%	55%	56%	56%
Zone 22	61%	73%	77%	75%
Zone 24	76%	80%	89%	91%
Zone 25	66%	69%	73%	75%
Zone 27	70%	80%	89%	91%
Zone 29	60%	60%	62%	62%

Table 4. Comparison of overall accuracies for modeling the different USGS mapping Zones.

CONCLUSION AND FUTURE WORK

The Remote Sensing Applications Center is proceeding with the modeling of forest type for the contiguous U.S. including Alaska. Lessons learned from the modeling trials presented above will be used for this effort. Forest type mapping will be conducted at the USGS mapping zone scale. No additional FIA variables will be modeled or used in the modeling process. The primary variables modeled will include field forest type and field forest type group. Analytical forest type and analytical forest type group will be modeled if available.

REFERENCES

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

Chan, J.C.-W., Huang, C., and DeFries, R.S. (2001). Enhanced algorithm performance for land cover classification using bagging and boosting. *IEEE Transactions on Geoscience and Remote Sensing* 39(3): 693-695.

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

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

Hansen, M.C., DeFries, R.S., Townshend, J.R.G, Carroll, M., Dimiceli, C., and Sohlberg, R.A. (2003). Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, Volume 7, Paper Number 10.

Homer, C.G., and Gallant A. (2001) "Partitioning the Conterminous United States into Mapping Zones for Landsat TM Land Cover Mapping", USGS Draft White Paper.

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

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

Moisen, G.G., Frescino, T.S., Huang, C., Vogelmann, J., Zhu, Z. (2003). Predictive modeling of forest cover type and tree canopy height in the central rocky mountains of Utah. *Proceedings of the 2003 meetings of the American Society for Photogrammetry and Remote Sensing, Anchorage, Alaska*.

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

Figure 1. Comparisons of analytical forest type, analytical forest type group, field forest type, and field forest type group of Arizona.

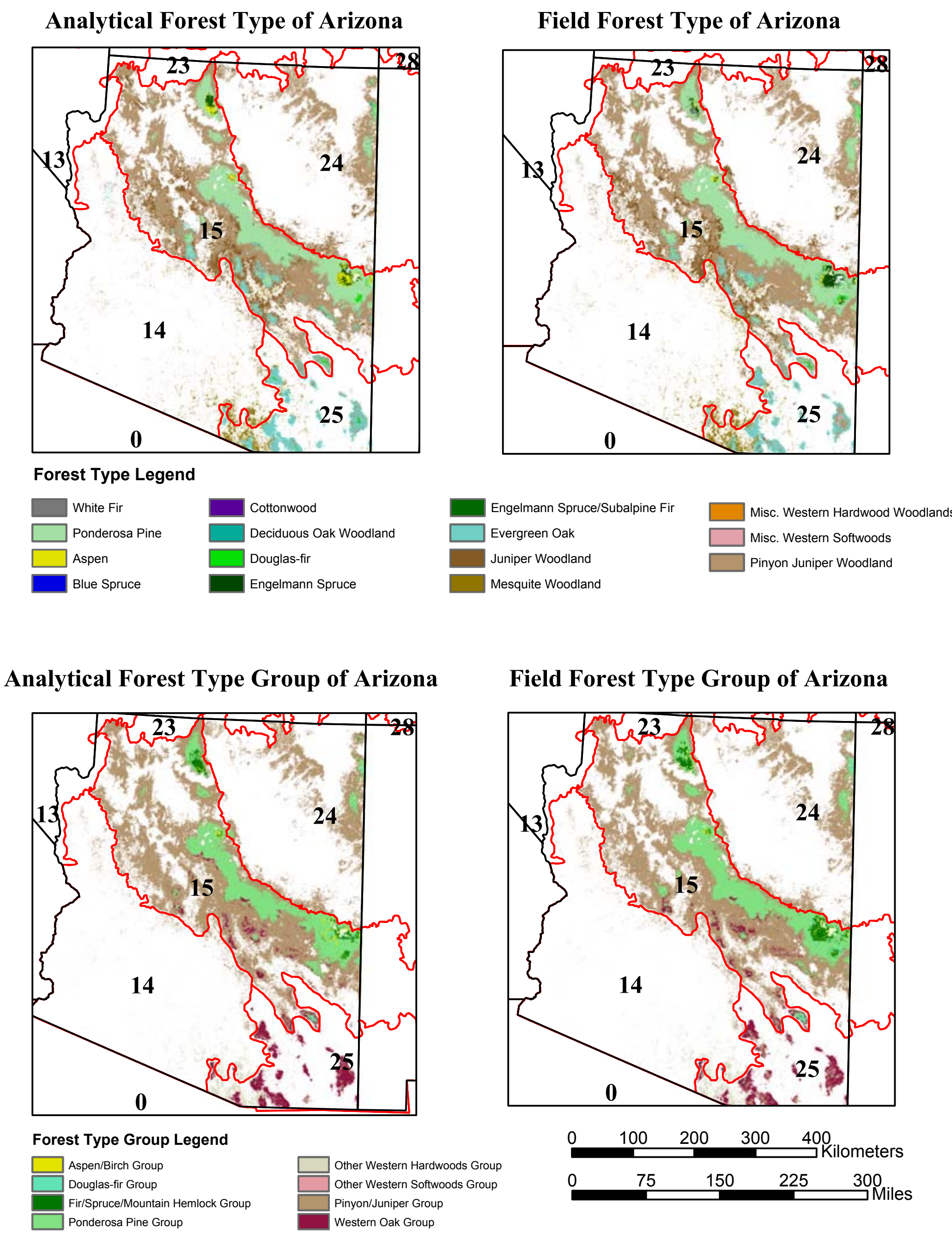

Figure 2. Comparisons of analytical forest type, analytical forest type group, field forest type, and field forest type group of New Mexico.

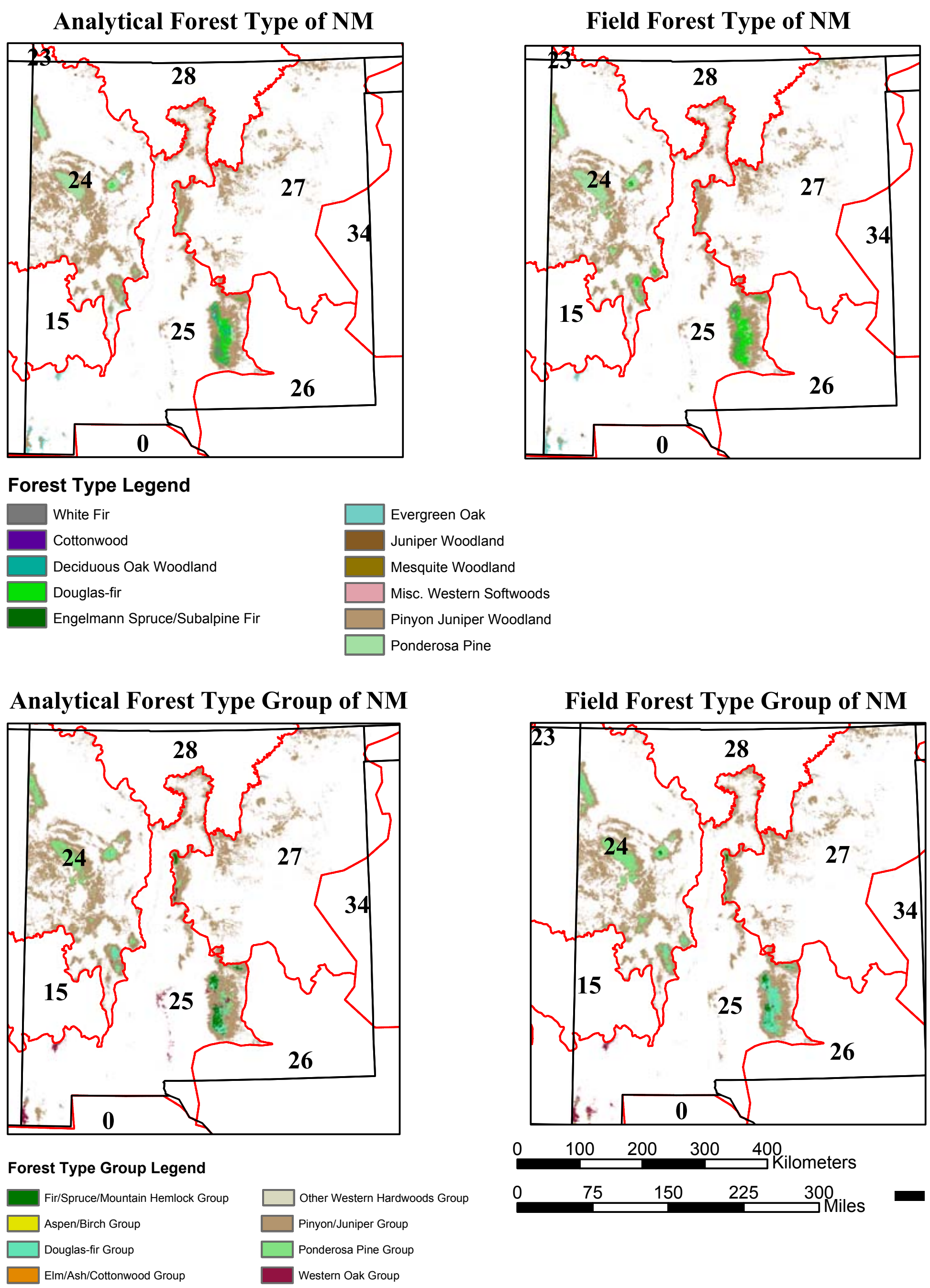

Figure 3. Comparisons of analytical forest type, analytical forest type group, field forest type, and field forest type group of Wyoming.