Predictive Mapping of Forest Attributes on the Fishlake National Forest

Tracey S. Frescino and Gretchen G. Moisen[1]

Abstract.—Forest land managers increasingly need maps of forest characteristics to aid in planning and management. A set of 30-m resolution maps was prepared for the Fishlake National Forest by modeling FIA plot variables as nonparametric functions of ancillary digital data. The set includes maps of volume, biomass, growth, stand age, size, crown cover, and various aspen characteristics. Ancillary data layers included pre-classified TM data, raw TM bands, and topographic variables. Predictive models were built using automated multivariate adaptive regression splines (MARS), and refined using local knowledge and digital orthoquads (DOQs). Validation and application issues are discussed.

National forest planners must frequently make decisions using existing information (Campbell and O'Brien 2004). There is rarely time or resources to collect new data specific to each question encountered. Tabular summaries and analytical reports prepared by the Forest Inventory and Analysis program (FIA) have proven useful for past assessments, but there is an increasing need for spatially explicit delineations of forest data. For example, maps are needed to assess suitable wildlife habitat, marketable harvest areas, desired future conditions, and historical distributions of forest cover types.

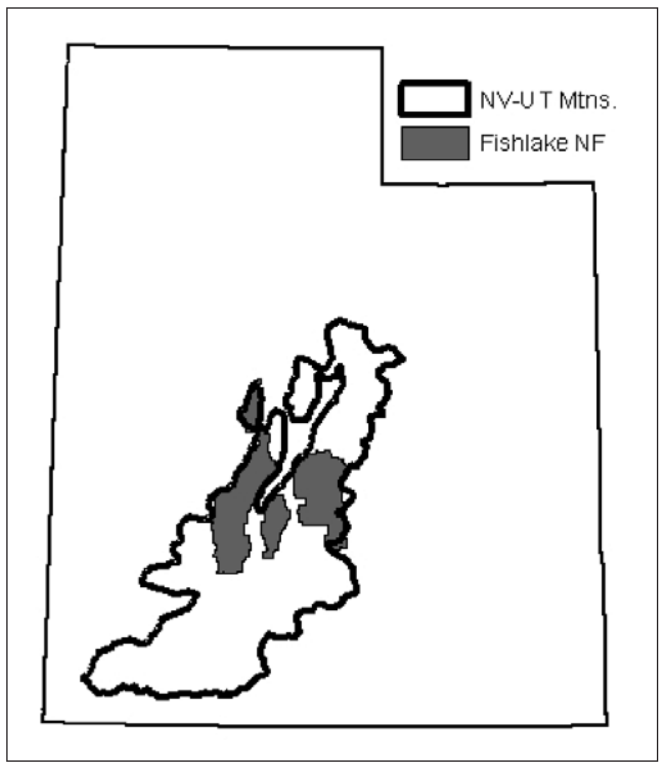
Our study demonstrates a method for generating spatially explicit maps of various forest attributes for use on national forests. The overall objective was to generate a series of maps to facilitate national forest management planning and to assist with a wildlife modeling study of cavity-nesting birds in aspen stands (Schultz 2002; Schultz *et al.* 2004; Edwards *et al.* 2002, 2004). Specifically, our objectives were to (1) build predictive models integrating FIA plot data with 30-m resolution digital data using multivariate adaptive regression splines (MARS) and geographical information systems (GIS) techniques; (2) refine and validate the models with statistical and visual error estimates; and (3) generate 30-m resolution maps of various FIA variables.

Methods

Study Area

The Fishlake National Forest comprises approximately 1,434,500 acres of land located in central Utah (fig. 1). It is a diverse forest with elevations ranging from less than 5,000 feet to over 12,000 feet. The forest supports a variety of vegetative cover types and forest resources. Pinyon-juniper cover types occur at low elevations and provide valuable habitat for deer, elk, and various small mammals and songbirds. Ponderosa pine and aspen cover types appear at higher elevations. Ponderosa pine provides valuable wildlife cover and is a valuable com-

Figure 1.—*Training data extents: Fishlake National Forest boundary and the Nevada-Utah Mountain ecoregion boundary.*



[1] Interior West Forest Inventory and Analysis, Rocky Mountain Research Station, U.S. Department of Agriculture, Forest Service, 507 25th Street, Ogden, UT 84401.

Table 1.—*FIA forest attributes, including units and description*

Forest Attribute (Alias)	Units	Description
Tree basal area (BALIVE)	Sq. ft./acre	Basal area of live trees 1 inch diameter and greater
Tree volume (NVOLTOT)	Cu. ft./acre	Net volume of live trees 5 inches diameter and greater
Tree biomass (BIOMASS)	Tons/acre	Woody biomass per acre of live trees 1 inch diameter and greater
Tree crown cover (CRCOV)	%	Crown cover of live trees 1 inch diameter and greater
Trees per acre (TPA)	# Trees	Trees per acre of live trees 1 inch diameter and greater
Stand age (STAGE)	Years	Weighted average age of the stand
Quadratic mean diameter (QMD)	Inches	Tree diameter based on the weighted average basal area of live timber trees 1 inch diameter and greater and live woodland trees 3 inches diameter and greater
Net annual growth (NGRWCF)	Cu. ft./acre	Annual net volume growth per acre of live growing-stock timber trees 5 inches diameter and greater and woodland trees 3 inches diameter and greater
Aspen presence (ASP)	Yes/no	Presence of aspen trees 1 inch diameter or greater
Aspen basal area (ASPBA)	Sq. ft./acre	Basal area of live aspen trees 1 inch diameter and greater
Percent aspen basal area	%	Percent basal area of live aspen trees 1 inch diameter and greater
Average tree height (TRHTAVG)	Feet	Average height of dominant or codominant trees
Snag density (SNAGNUM)	# Snags	Snags per acre of standing dead trees 5 inches diameter and greater
Aspen rot presence (ASPROT)	Yes/no	Presence of aspen disease

mercial tree species. Aspen is widely known to be prime wildlife habitat, affording beneficial cover, water, and food for a variety of wildlife species. Aspen cover types are being threatened by successional climax species, such as subalpine fir, white fir, and spruce, which are crowding out the aspen and diminishing the benefits to wildlife. Spruce-fir cover types occur at the higher elevations.

The Forest falls almost entirely within Bailey's (1980) Nevada-Utah Mountain ecoregion province, revised by Homer *et al.* (1997) (F-1). FIA data were available throughout this ecoregion. Because this ecoregion is ecologically similar to the Fishlake National Forest, we considered modeling forest characteristics for both areas.

Data

There were 836 forested locations within the Nevada-Utah Mountain ecoregion and 231 forested FIA locations within the Fishlake National Forest. We identified a set of eight FIA forest attributes to assist with management planning (tree basal area, tree volume, tree biomass, tree crown cover, trees per acre, quadratic mean diameter, stand age, and net annual growth), and a set of six additional variables needed for modeling aspen habitats for cavity-nesting birds (aspen presence, aspen basal area, percent aspen basal area, average tree height, snag density, and aspen rot presence). We used data collected on the FIA plots to compile individual tree measurements and combined them with stand variables to produce location-level summaries of all variables (table 1).

Data extraction and mining routines were performed within a GIS environment. We acquired a set of twelve 30-m resolution digital layers that would be appropriate for predicting forest attributes (table 2). Seven of these layers were based on 30-m resolution Enhanced Thematic Mapper (ETM) satellite data obtained through the Multi-Resolution Land Characteristics (MRLC) consortium. Three were raw spectral bands, one was a normalized difference vegetation index (NDVI) derived from the raw spectral bands, and the remaining three were classified ETM products generated by the Land Cover Characterization (LCC) program of the U.S. Geological Survey (USGS) Earth Resources Observation Systems (EROS) Data

Center (EDC) (Huang *et al.*, in press). The other five predictor variables were derived from 30-m resolution National Elevation Dataset (NED) digital elevation models (DEMs), including elevation, aspect, slope, hillshade, and topographic class. Elevation was extracted directly from the DEMs while aspect, slope, and hillshade were derived from the DEM using functions from the GRID module in ArcInfo GIS (ESRI Inc., Redlands, CA). The topographic class variable was derived from the DEM using a customized arc macro language AML (Zimmerman, unpublished data). The aspect variable was transformed from degrees to a symmetric radiation wetness index, calculated using the following formula (Roberts and Cooper 1989):

$$\text{Aspect} = \frac{1 - \cos(aspect - 30)}{2}$$

This transformation assigns the highest values to land oriented in a north-northeast direction, the coolest and wettest orientation in Utah. The hillshade variable was derived using an illumination angle of 225 degrees.

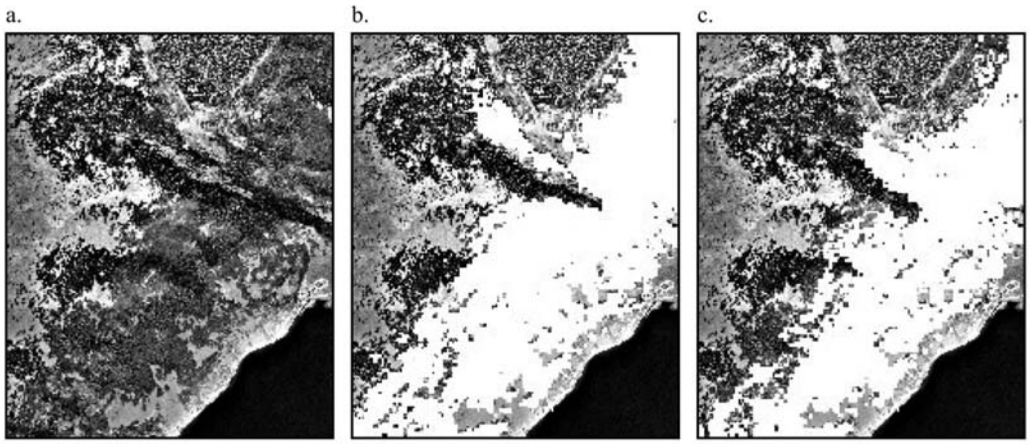
Models

Predictive models of various forest attributes were generated using Multivariate Adaptive Regression Splines (MARS) (Friedman 1991, Prasad and Iverson 2002, Steinberg *et al.* 1999). MARS is a flexible, nonparametric regression modeling tool that automatically finds the complex relationships between a response variable and a set of continuous and discrete predictors. MARS builds models by fitting numerous piecewise linear regressions, and approximates nonlinearity by allowing the slope of the regression lines to change over different intervals of the predictor space. These intervals are defined by *basis functions*, which are the building blocks of a MARS model. MARS starts by building a large and overly complex model with many basis functions. An optimal model is then found by deleting basis functions in order of least contribution to model performance. This prevents over-fitting and ensures that the mode will stand up to new data for prediction applications such as mapping. Features of MARS that make it particularly well suited to mapping forest attributes are that it handles both categorical and continuous variables, selects the relevant predictor

Table 2.—*Ancillary data predictor variables, including units and description*

Forest Attribute (Alias)	Units	Description
ETM Band 3 (ETMB3)	Brightness value (0-255)	Red (0.63 - 0.69 micrometers); June 2000–leaf on
ETM Band 4 (ETMB4)	Brightness value (0-255)	Near-infrared (0.76 - 0.90 micrometers); June 2000–leaf on
ETM Band 5 (ETMB5)	Brightness value (0-255)	Mid-infrared (1.55 – 1.75 micrometers); June 2000–leaf on
ETM NDVI (ETMVI)	0.0 – 1.0	Normalized Difference Vegetation Index; June 2000–leaf on
Classified ETM (LCC10)	10 classes	2:Nonforest; 10:Pinyon/juniper;15:Douglas-fir; 20:Ponderosa pine; 30:Spruce/fir; 35:Lodgepole; 50:Other western softwoods; 75:Aspen/birch; 85:Western oak; 90:Other western hardwoods (based on June 2000–leaf on ETM)
Classified ETM (LCC4)	4 classes	2:Nonforest; 41:Deciduous; 42:Evergreen; 43:Mixed (based on June 2000–leaf on ETM)
Classified ETM (LCC2)	2 classes	1:Forest; 2:Nonforest (based on June 2000–leaf on ETM)
Elevation (ELEV)	Meters	Elevation from mean sea level
Aspect (TRASP)	0 to 1	Transformed index representing radiation and wetness
Slope (SLP)	%	The rate of change from one cell to the next
Hillshade (HLSHD)	Brightness value (0-255)	Shaded relief considering shadows and an illumination angle of 225 degrees
Topoclass (TOPOCL)	4 classes	Classified to identify topographic features (1:Ridge;, 2:Slope; 3:Toe slope; 4:Valley bottom)

Figure 2.—*Visual assessment of aspen presence predictions compared to a DOQ. a. DOQ without predictions. b. Predictions based on a model built using the Nevada-Utah Mountain ecoregion data set; c. Predictions based on a model built using the Fishlake National Forest boundary data set. White represents the predicted aspen presence.*



variables and specifies their relationship with the response automatically, determines the level and nature of interactions as well as transformations, handles missing values, protects against over-fitting, and is fast and efficient for large data sets.

We examined the effect of using training data from the two different data extents, the Fishlake National Forest and the Nevada-Utah Mountain ecoregion. In all cases, we ran the models with a maximum of 100 basis functions, specified ten-fold cross-validation to select model degrees of freedom and prevent over-fitting, and allowed second-order interactions between predictor variables. Model performance was evaluated and refined by looking at R^2 and mean square error measures generated by MARS on a subset of the forest variables. We also visually assessed the model predictions using parallel screen displays of digital orthoquads (DOQs) and the output maps.

Maps

Maps were generated within a GIS environment. MARS output was converted to Arc Macro Language (AML) using Visual Basic (J. Nelson, unpublished data) and then run in Arc GRID. Thirty-meter resolution, spatially explicit maps were output for each FIA attribute. The nonforest class from the LCD classified-ETM product was used to mask the nonforest areas on the ground. Alternative approaches to applying MARS models to large geographic areas using Iterative Data Language are discussed in Terletzky and Frescino (2004).

Results

The models using the 836 training locations within the Nevada-Utah Mountain ecoregion performed better than the models using 231 training locations within the Fishlake National Forest in most cases. Table 3 shows the R^2 and MSE results from MARS for eight different FIA attributes, comparing the models built using different training data sets. The numbers in bold represent higher R^2 and lower MSE values, indicating better model fits. For five of the eight attributes, the R^2 values were higher when using the Nevada-Utah Mountain ecoregion data set. MSE values were lower when using the Nevada-Utah Mountain ecoregion data set for seven of the eight attributes.

Figure 2 shows an example of the visual assessment for models predicting aspen presence, comparing prediction results with what is displayed from a DOQ. The visual assessments of the predictions from the models built from the Nevada-Utah Mountain ecoregion data set appeared better than the predictions from the models built using the Fishlake National Forest data set.

Discussion and Conclusions

Predictive modeling is not an exact science. Many factors influence model performance. One is the extent of the training data set. With several examples and evaluation procedures, we determined that the models performed "better" when using a larger data set. Although many of the data were outside the area of interest, the data were ecologically similar and significantly helped to establish functional relationships between the forest attributes and the ancillary data products.

What does "better" mean? Although we were able to objectively compare model performance using R^2 and MSE measures, further analysis is needed to validate the model predictions using independent data and true tenfold cross-validation procedures. More importantly, global measures of map accuracy often cannot capture what is obvious to a forest land manager wanting to use predictive maps in real-world applications. Further investigation is needed to build measures of utility into the picture.

Table 3.—*R^2 and MSE results from MARS for models built using the Fishlake National Forest (Fnf) data set and Nevada-Utah Mountain ecoregion (Uteco) data set (numbers in bold represent the best-fit model)*

Attribute	Training data set	R^2	MSE
BALIVE	Fnf	0.053	4,575.91
	Uteco	**0.284**	**3,012.26**
NVOLTOT	Fnf	0.419	1,255,907.10
	Uteco	**0.525**	**1,110,762.60**
BIOMASS	Fnf	0.348	372.43
	Uteco	**0.476**	**330.91**
CRCOV	Fnf	**0.438**	**290.15**
	Uteco	0.385	295.51
TPA	Fnf	0.335	110,635.50
	Uteco	**0.349**	**96,821.37**
Stage	Fnf	0.046	4,419.27
	Uteco	**0.114**	**3,335.20**
Trhtavg	Fnf	**0.739**	308.74
	Uteco	0.554	**195.20**
Aspba	Fnf	**0.478**	2,110.66
	Uteco	0.396	**1,956.01**

As mentioned previously, one of the features of MARS is that it selects the relevant predictor variables. This allows us to see which predictor variables most influence the occurrence of the different forest attributes. Table 4 shows the predictor variables that were used to build the final models for eight forest attributes. The order of the variables corresponds to the relative importance of each in the model, or the amount of variance reduced by each. In general, the variables that seemed to have the most influence were the ETM raw spectral bands 5 and 3, elevation, the classified-ETM 10 and 4 classes, and the topographic class. This makes sense since band 5 (mid-infrared) characteristically indicates vegetation moisture and band 3 (red) responds to chlorophyll absorption. Elevation is a surrogate for temperature and moisture as well as the topographic class that distinguishes ridges from slopes from valley bottoms. The classified-ETM products would help distinguish differences between different forest classes, removing shadows and other features that the raw imagery may confuse.

Modeling forest attributes is an attempt to delineate characteristics in the landscape using available field data and ancillary resources, such as satellite imagery and topographic data. We assume there are significant relationships between these attributes and ancillary resources. Further research is needed to refine these relationships and obtain new ancillary products to build more accurate models.

Table 4.—*Relevant variables contributing to model variance reduction*

Forest attribute	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6	Variable 7
BALIVE	ETMB5	ELEV	LCD10	LCD2	TOPOCL		
NVOLTOT	ELEV	ETMb3	ETMB5	LCD4	TOPOCL	LCD10	ETMB4
BIOMASS	ETMB5	ELEV	LCD10	TOPOCL			
CRCOV	ETMB3	LCD4					
TPA	LCD4	ELEV	ETMB5				
STAGE	LCD4	ELEV	ETMB5				
TRHTAVGN	LCD10	ETMB3	ELEV	TOPOCL			
ASPBA	LCD4	LCD10	ETMb3	ETMb5	ELEV	SLP	

National forest planners are enthusiastic about incorporating these spatially explicit products into their planning procedures and integrating them with other digital data to help understand the spatial diversity in the landscape and make decisions related to wildlife habitat, marketable harvest areas, desired future conditions, and so on. Wildlife modelers are also enthusiastic about adding spatially explicit maps of specific attributes into their models. These maps will provide valuable information about structural components of the forest and allow predictions of wildlife species, spatially across the landscape.

Literature Cited

Bailey, R.G. 1980. Description of the ecoregions of the United States. Misc. Publ. 1391. Washington, DC: U.S. Department of Agriculture.

Campbell, R.C., Jr.; O'Brien, R.A. 2004. Spatial information needs on the Fishlake National Forest: Can FIA help? In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory 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.

Edwards, T.C., Jr.; Moisen, G.G.; Frescino, T.S.; Lawler, J.J. 2002. Modeling multiple ecological scales to link landscape theory to wildlife conservation. In: Bissonette, J.A.; Storch, I., eds. Landscape ecology and resource management: making the linkages. Covelo, CA: Island Press.

Edwards, T.C., Jr.; Moisen, G.G.; Frescino, T.S; Schultz, R.J. 2004. The application of FIA-based, data to wildlife habitat modeling: a comparative study. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory 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.

Friedman, J. H. 1991. Multivariate adaptive regression splines. Annals of Statistics. 19: 1–141.

Homer, C.; Ramsey R.D.; Edwards, T.C., Jr.; Falconer, A. 1997. Landscape cover-type modeling using multi-scene TM mosaic. Photogrammetric Engineering and Remote Sensing. 63: 59–67.

Huang, C.; *et al*. 2002. Synergistic use of FIA plot data and Landsat 7 ETM+ images for large area forest mapping. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; Moser, John W., eds. Proceedings, 3rd annual forest inventory and analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MI: U.S. Department of Agriculture, Forest Service, North Central Research Station: 50–55.

Omernik, J.M. 1987. Map supplement: ecoregions of the conterminous United States. Annals of the Association of American Geographers. 77: 118–125 (map).

Prasad, A.M.; Iverson, L.R. 2002. Predictive vegetation mapping using a custom built model-chooser: comparison of regression tree analysis and multivariate adaptive regression splines. In: Proceedings of the 4th International conference on integrating GIS and environmental modeling: problems, prospects and research needs; 2000 September 2–8; Banff, Alberta, Canada.

Roberts, D.W.; Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. In: Ferguson, D.; Morgan, P.; Johnson, F.D., eds. Land classification based on vegetation: applications for resource management. Gen. Tech. Rep. INT-257. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station: 90–96.

Schultz, R.J., Jr. 2002. Scale dependency in habitat selection by cavity nesting birds in Fishlake National Forest. Logan, UT: Utah State University. 100 p. M.S. thesis.

Schultz, R.J., Jr.; Edwards, T.C.; Moisen, G.G.; Frescino, T.S. 2004. Development and validation of spatially explicit habitat models for cavity-nesting birds in Fishlake National Forest, Utah. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory 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.

Steinberg, D.; Colla, P.L.; Martin, K. 1999. MARS users guide. San Diego, CA: Salford Systems.

Terletzky, P.; Frescino, T.S. 2004. Integrating spatial components into FIA models of forest resources: some technical aspects. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory 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.