Chapter 7

Mapping Potential Vegetation Type for the LANDFIRE Prototype Project

Tracey S. Frescino and Matthew G. Rollins

Introduction

Mapped potential vegetation functioned as a key component in the Landscape Fire and Resource Management Planning Tools Prototype Project (LANDFIRE Prototype Project). Disturbance regimes, vegetation response and succession, and wildland fuel dynamics across landscapes are controlled by patterns of the environmental factors (biophysical settings) that entrain the physiology and distribution of vegetation. These biophysical characteristics of landscapes are linked to stable vegetation communities that occur in the absence of disturbance (Arno and others 1985; Cooper and others 1991; Ferguson 1989; Pfister and Arno 1980; Pfister and others 1977). In the LANDFIRE Prototype Project, these stable vegetation community types were referred to as potential vegetation types (PVTs). Further, the concept of potential vegetation was used as a basis for developing biophysical map units that were critical for developing the LANDFIRE wildland fuel and fire regime products. In the LANDFIRE Prototype Project, maps of potential vegetation facilitated linkage of the ecological process of succession to simulation landscapes used as input the LANDSUMv4 landscape fire succession model for modeling historical vegetation reference conditions and historical fire regimes (Long and others, Ch. 9). In addition, maps of PVT were used to guide the parameterization and calibration of the landscape fire succession model LANDSUMv4 (Pratt and others, Ch. 10) and to stratify vegetation communities for mapping current vegetation and wildland fuel mapping (Zhu and others, Ch, 8; Keane and others, Ch. 12).

Analysis of the biophysical characteristics of landscapes is commonly used to quantify distributions of vegetation along biophysical gradients (Bray and Curtis 1957; Gleason 1926; Whittaker 1967). Previous research has employed cluster analysis and ordination techniques to delineate biophysical gradients and link them to corresponding potential vegetation (Galiván and others 1998). Other research has used supervised classification methods or predictive vegetation mapping techniques (Franklin 1995) to link potential natural vegetation with biophysical gradients (Keane and others 2000; Keane and others 2001; Lenihan and Neilson 1993; Rollins and others 2004) and gradients of climate, topography, and soils (Brzeziecki and others 1993; Jensen and others 2000).

We developed PVT map unit classifications based on species' shade tolerance and moisture tolerance to link LANDFIRE reference plot data to unique environmental conditions or biophysical settings. Here, we define biophysical setting as the suite of biotic and abiotic factors that affect the composition, structure, and function of vegetation. Our main assumption was that the shade tolerant species would serve as unique indicators of biophysical conditions (Daubenmire 1967). Because of dynamic climate and ecosystem complexities, we did not assume that a stable climax community would exist without the influence of disturbance (Keane and Rollins, Chapter 3).

In: Rollins, M.G.; Frame, C.K., tech. eds. 2006. The LANDFIRE Prototype Project: nationally consistent and locally relevant geospatial data for wildland fire management. Gen. Tech. Rep. RMRS-GTR-175. Fort Collins: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Initially, we investigated an unsupervised clustering approach to stratify the landscape using a series of indirect biophysical gradients (Hargrove and Luxmore 1998; Hessburg and others 2000a, Hessburg and others 2000b). This approach successfully delineated unique biophysical settings, but the categories were not significantly correlated to patterns of vegetation. Alternatively, we used a supervised predictive modeling approach based on ground-referenced data to explicitly link biophysical gradients to potential vegetation. This approach provided an objective and repeatable method that could be linked directly to vegetation patterns identifiable in the field. This chapter describes the process used for mapping potential vegetation for the LANDFIRE Prototype Project and provides recommendations for generating maps of potential vegetation for the national implementation of LANDFIRE.

Methods

The LANDFIRE Prototype Project involved many sequential steps, intermediate products, and interdependent processes. Please see appendix 2-A in Rollins and others, Ch. 2 for a detailed outline of the procedures followed to create the entire suite of LANDFIRE Prototype products. This chapter focuses specifically on the procedure followed in developing the potential vegetation maps, which served as spatial templates for nearly all mapping tasks in the LANDFIRE Prototype Project.

Field-referenced Data

Comprehensive field-based reference data are critical for implementing a supervised mapping application, and these "training data" must be a statistically robust sample of the population. The LANDFIRE reference database (LFRDB) was designed to meet these criteria and provided an excellent source of consistent, comprehensive reference data from which to develop training sites for our predictive landscape models (Caratti, Chapter 4). Georeferenced field locations were obtained from the LFRDB and assigned PVTs based on hierarchical, floristic keys organized along gradients of shade tolerance and moisture tolerance developed a priori (Long and others, Chapter 6). The development of the keys began with existing national classifications (Kuchler 1975) and was then revised by regional (Quigley and others 1996) and local (Pfister and others 1977) classifications. The keys were further revised using the LFRDB, an extensive literature review, and review by regional ecological experts. To qualify as a separate class, individual PVTs had to fit the criteria of being identifiable in the field, scalable, mappable, and model-able (See Keane and Rollins, Ch. 3 and Long and others, Ch. 6).

The keys divided PVTs into three physiological life forms, forest, shrub, and herbaceous, with forest PVTs following a shade tolerance gradient and shrub and herbaceous PVTs following moisture gradients. Initially, Zone 16 had 13 classes of forest PVTs, 10 classes of shrub PVTs, and 3 classes of herbaceous PVTs. Distinguishing between classes requires a sufficient number of training plots for each class. We grouped classes having fewer than 20 training plots with other classes, resulting in 10 forest classes, 8 shrub classes, and 3 herbaceous classes (table 1). To minimize the number of classes in Zone 19 and in an effort to increase overall map accuracy, we implemented the classification key for this zone under the criterion that a minimum of 30 training plots were necessary for a PVT to form a unique class. Table 2 shows Zone 19 PVT classes and the number of training plots from the database assigned to each class.

Spatial Data

The biophysical gradient layers included variables created using WXFIRE, an ecosystem simulation model developed by R.E. Keane at the USDA Forest Service, Rocky Mountain Research Station, Missoula Fire Sciences Laboratory in Missoula, Montana (Keane and others 2006; Keane and Rollins, Ch. 3) and variables from the National Elevation Database (http://ned.usgs.gov). The WXFIRE model integrates DAYMET climate data (Running and Thornton 1996; Thornton and others 1997; Thornton and others 2000) with landscape data and site specific parameters (for example, soils and topography) and generates spatially explicit maps of climate and ecosystem variables that integrate landscape-weather interactions (See Holsinger and others, Ch. 5 for details about these variables and how they were derived). For topographic gradients, we used variables from the National Elevation Database, including elevation, derivatives of slope, aspect, a classified landform variable, and a topographic position index. This process resulted in a total of 38 biophysical gradients available for use as independent variables in our predictive landscape models of PVT. We reviewed correlation matrices and principle component analyses to reduce (winnow) this list of variables used in the modeling process. For Zone 16, we used 21 variables (table 3) and for Zone 19, 22 variables (table 4).

Table 1—Zone 16 codes, life forms, names, and the number of training sites and test sites by PVT. Life form categories include F (forest), S (shrub), and H (herbaceous).

Code	Life form	Name	Number of training sites	Number of test sites
1	F	Spruce - Fir / Blue Spruce	157	13
2	F	Spruce - Fir / Spruce - Fir	1188	92
3	F	Grand fir - White Fir	439	40
4	F	Douglas-fir / Lodgepole Pine - Timberline Pine	65	3
5	F	Douglas-fir / Douglas-fir	263	19
6	F	Lodgepole Pine - Timberline Pine	104	10
7	F	Ponderosa Pine	205	16
8	F	Pinyon - Juniper / Mountain Big Sagebrush	433	31
9	F	Pinyon - Juniper / Wyoming - Basin Big Sagebrush	1052	95
10	F	Riparian Hardwood	126	11
11	S	Riparian Shrub	33	4
12	S	Blackbrush - Chaparral - Dry Deciduous Shrub	22	3
13	S	Dwarf Sagebrush	99	14
14	S	Salt Desert Shrub	35	2
15	S	Mountain Mahogany	66	4
16	S	Gambel Oak	172	14
17	S	Wyoming - Basin Big Sagebrush	118	12
18	S	Mountain Big Sagebrush	171	17
19	H	Wetland Herbaceous	57	6
20	H	Alpine	47	6
21	H	Herbaceous	109	9

Table 2—Zone 19 codes, life forms, names, and the number of training sites and test sites by PVT. Life form categories include F (forest), S (shrub), and H (herbaceous).

Code	Life form	Name	Number of training sites	Number of test sites
1	F	Western Redcedar	176	23
2	F	Grand Fir - White Fir	194	33
3	F	Spruce - Fir / Montane	1418	235
4	F	Spruce - Fir / Timberline	951	133
5	F	Spruce - Fir / Subalpine	1165	171
6	F	Douglas-fir / Ponderosa Pine	363	56
7	F	Douglas-fir / Lodgepole Pine	546	88
8	F	Douglas-fir / Timberline Pine	161	26
9	F	Douglas-fir / Douglas-fir	947	125
10	F	Lodgepole Pine	460	55
11	F	Ponderosa Pine	76	8
12	F	Timberline Pine / Limber Pine	51	7
13	F	Timberline Pine / Whitebark Pine	40	6
14	F	Rocky Mountain Juniper	33	3
15	F	Riparian Hardwood	28	2
16	S	Riparian Shrub	94	5
17	S	Mountain Mahogany	32	3
18	S	Dry Shrub	51	4
19	S	Dwarf Sagebrush Complex	68	10
20	S	Mountain Big Sagebrush Complex	249	43
21	S	Threetip Sagebrush	187	26
22	S	Wyoming - Basin Big Sagebrush Complex	514	75
23	H	Wetland Herbaceous	112	9
24	H	Alpine	30	3
25	H	Fescue Grasslands	174	22
26	H	Bluebunch Wheatgrass	144	23

Table 3—Zone 16 PVT predictor layers. See Holsinger and others, Ch. 5, table 6 for biological significance of each layer.

Code	Units	Description
aet	kg H_20 yr^{-1}	Actual evapotranspiration
dsr	days	Days since last rain
dss	days	Days since last snow
gsws	-MPa	Growing season water stress
mc1	%	NFDRS – 1-hr wood moisture content
outflow	kg H_20 m^{-2} day^{-1}	Soil water lost to runoff and ground
pet	kg H_20 yr^{-1}	Potential evapotranspiration
ppt	cm	Precipitation
psi	-MPa	Water potential of soil and leaves
psi.max	-MPa	Maximum annual leaf water potential
rh	%	Relative humidity
srad.tg	kJ m^{-2} day^{-1}	Total solar radiation
tmin	°C	Minimum daily temperature
vmc	Scalar	Volumetric water content
sdepth	cm	Soil depth
elev	m	Elevation
aspect	8 classes	Aspect class*
slope	%	Slope
lndfrm	10 classes	Landform**
trmi	Index (0-1)	Topographic relative moisture index
posidx	Index (0-1)	Topographic position index

*Aspect classes – 0:Level; 1:North; 2:North-East; 3:East; 4:South-East; 5:South; 6: South-West; 7:West; 8:North-West
**Landform classes – 1:Vally flats; 2:Toe slopes; 3:Gently sloping ridges and hills; 4: Nearly level plateaus and hills; 5:Very moist steep slopes; 6:Moderately moist steep slopes; 7:Moderately dry slopes; 8:Very dry steep slopes; 9:Cool aspect cliffs, canyons; 10:Hot aspect cliffs, canyons.

Table 4—Zone 19 PVT predictor layers. See Holsinger and others, Ch. 5, table 6 for biological significance of each layer.

Code	Units	Description
aet	kg H_20 yr^{-1}	Actual evapotranspiration
dday	°C	Degree-days
dss	days	Days since last snow
evap	kg H_20 m^{-2} day^{-1}	Evaporation
g.sh	M sec^{-1}	Leaf-scale stomatal conductance
gsws	-MPa	Growing season water stress
outflow	kg H_20 m^{-2} day^{-1}	Soil water lost to runoff and ground
pet	kg H_20 yr^{-1}	Potential evapotranspiration
ppfd	Umol m^{-2}	Photon flux density
ppt	cm	Precipitation
psi	-MPa	Water potential of soil and leaves
snowfall	kg H_20 m^{-2} day^{-1}	Snowfall
srad.fg	KW m^{-2} day^{-1}	Solar radiation flux to the ground
tmax	°C	Maximum daily temperature
tmin	°C	Minimum daily temperature
tnight	°C	Nighttime daily temperature
trans	kg H_20 m^{-2} day^{-1}	Soil water transpired by canopy
vmc	Scalar	Volumetric water content
sdepth	cm	Soil depth
elev	m	Elevation
posidx	index (0-1)	Topographic position index
slope	%	Slope

*Aspect classes – 0:Level; 1:North; 2:North-East; 3:East; 4:South-East; 5:South; 6:South-West; 7:West; 8:North-West

Modeling and Mapping Process

Classification trees, also known as decision trees, have been widely applied in landscape mapping applications (Brown de Colstoun and others 2003; Friedl and Brodley 1997; Hansen and others 2000; Joy and others 2003; Moisen and others 2003, Moore and others 1991; Rollins and others 2004). Classification trees were originally developed for artificial intelligence research to identify patterns and recognize these patterns in similar situations using a hierarchical structure of rules (Quinlan 1986). The rules are constructed from available training data where observations are delineated into smaller subsets of more homogenous classes. Specifically, the classification tree algorithm considers each predictor variable and examines all n-1 ways to split the data into two clusters. For every possible split of each predictor variable, the within-cluster impurity is calculated. The first split in the tree is that which yields the smallest overall within-cluster impurity. This process is repeated for each branch defined by the previous split (Breiman and others 1984).

Classification trees are well-suited to vegetation mapping because they accommodate common conceptions that vegetation has a nonlinear, non-normal response to environmental gradients (Austin and others 1984). In addition, they are nonparametric models, meaning they make no underlying assumptions about the distribution of the data, and they are adaptable for nonlinear relationships between the predictors and the response (Friedl and Brodley 1997). Classification trees are also valuable because they are robust, are able to incorporate both categorical and continuous variables, and are relatively insensitive to outliers (Breiman and others 1984). Furthermore, for a large project such as LANDFIRE, classification trees offer the advantage that models are generated and executed quickly.

The classification trees for modeling PVTs were generated using the commercially available See5 machine-learning algorithm (Quinlan 1986, 1993; Rulequest Research 2004) and were applied within an ERDAS Imagine (ERDAS, Inc. 2001) interface. See5 uses a classification and regression tree (CART) approach for constructing a tree, generating a tree with high complexity, and pruning it back to a more simple tree by merging classes (Breiman and others 1984). This pruning process was found to improve the efficiency of the model and minimize the classification error (Breiman and others 1984). We used the boosting feature of See5 to improve the accuracy of the model (Friedl and others 1999; Quinlan 1986). In the boosting procedure, multiple trees are built in an iterative process and, each tree "learns" from the misclassification errors of the previously built tree (Bauer and Kohavi 1999). The final tree is selected from all the trees based on a weighted vote of the predictions. We also employed other features of See5 including winnowing, which excludes variables that are not relevant in the model, and differential misclassification cost weighting, which assigns more weight to classes with more costly classification errors.

Although not fully automated, the process for mapping PVTs was simplified using a suite of tools developed by Earth Satellite Corporation (2003) in support of the National Land Cover Database (NLCD 2000). These tools were developed to integrate the Rulequest See5/C5.0 software package with the ERDAS Imagine image-processing software. For mapping PVTs, we used the sampling tool to set up See5 input files and the classifier tool to generate the final map and a coinciding map of error or confidence. The sampling tool allows a user to input a spatially explicit layer of field-referenced training data as the dependent variable and multiple spatially explicit gradient layers as the independent variables and then outputs the input files needed to run See5. The classifier tool applies the output tree model from See5 over the specified spatial extent or a specified masked extent.

To meet the input requirements of See5 and to improve the efficiency of the model-making process, we followed three pre-processing rules: (1) all layers must be ERDAS Imagine images, (2) all layers must have the same number of rows and columns, and (3) all layers must be size 16-bit or smaller, with positive values. A few data preparation steps were necessary to follow these rules. The biophysical gradient layers are output from WXFIRE as Arc/Info grids with float data values. We ran an Arc/Info AML (Arc Macro Language) to translate and dilate or "stretch" the grids to an unsigned, 16-bit integer format; converted the grids to ERDAS Imagine images using a batch setup in ArcGis 8.0 (ESRI Inc. 2001), and masked the images in Imagine using a buffered mask of the zone region (the zone boundaries). Through the entire LANDFIRE process, we used a 3-km buffer around the zone boundary. This buffer facilitated edge matching and reduced the edge effects in modeling historical fire regimes (Pratt and others, Ch. 10) The topographic and soil gradient layers were also converted to images and masked with the buffered zone region. We generated the spatially explicit dependent layer within ArcGis 8.0 using the spatial analyst tool to convert a data table to an image and set the extent to match the gradient images. Prior to creating this layer, we performed exploratory data analyses, both spatial and non-spatial, to look for

and remove any major outliers or unusual patterns in the data. The output from the sampling tool includes a "data file," which contains values from the model response and the corresponding value of the model predictor layers for each georeferenced training site, and a "names file" identifying the model input names and data types.

For each prototype mapping zone, we built three different See5 classification trees and generated three different maps. The first classification tree was generated using a binary response variable describing forest and non-forest PVTs. The resulting map was used to stratify the zones to improve the performance of the PVT models. The other two classification trees were generated and applied to forest PVTs within the predicted forested areas and non-forest PVTs within the predicted non-forested areas. The final map was a combined product of the forest PVT predictions and the non-forest PVT predictions from each zone. For Zone 16, classes of agriculture, barren, open water, and urban/developed were masked from the Zone 16 cover type map and were considered non-forest types. For Zone 19, we masked only classes of barren, open water, and snow/ice following the assumption there is a potential for vegetation to grow on agricultural and urban lands. These classes had not been mapped for Zone 19 at this stage of the mapping process and were masked after the final PVT map was generated.

For the forested and non-forested stratification map, all training plots classified as forest PVTs were grouped into one class and the training plots classified as shrub or herbaceous PVTs into another class. There were a total of 4,032 training sites for Zone 16 with 4,032 forested plots and 929 non-forested plots (table 1). For Zone 19, there were a total of 8,264 training sites, 6,609 forested plots and 1,655 non-forested plots (table 2). Multiple models were executed exploring the different features of See5, including winnowing, boosting, and analyzing differential misclassification costs. We selected the model having the lowest error. The final PVT maps for each zone were created using the classifier tool and represented an integration of the forest/non-forest models defined by the masking strategy described above.

Accuracy Assessment

We used a 10-fold cross-validation routine performed by See5 to assess the accuracy of the binary forested and non-forested stratification map and used an independent test set to assess the accuracy of the forest and non-forest PVT predictions. We determined that a 10-fold cross-validation measure would be sufficient for assessing the accuracy of the stratification map and would maximize the number of plots used for developing the model. The independent test set would, in turn, assess the accuracy of the final map product. To perform the 10-fold cross-validation routine, the training data set was divided into 10 blocks of approximately the same size and class distribution. A classification tree was built ten times, and each time, one block was withheld for testing purposes. The error rate was averaged from the total number of errors and the total number of training sites. See5 output an error matrix generated from the sum of all errors and calculated the percent of the predictions that were correctly classified.

From the LANDFIRE reference database, we randomly reserved ten percent of the training sites. These sites were withheld from the modeling process and were used to independently evaluate the accuracy of the final map. There were a total of 421 test sites for Zone 16 and 1194 test sites for Zone 19 (tables 1 and 2). See5 automatically tested the model predictions at these sites and output an error matrix and a percentage measure of PVTs that were correctly classified. We brought the error matrix results into R statistical software (Ihaka and Gentleman 1996) and calculated user and producer accuracy measures and a kappa statistic to see if the model could achieve above-random accuracy (Cohen 1960; Congalton and Green 1999).

Error matrices provide a global summary of the accuracy of the map but do not show the range and variability of the accuracies across the map (Congalton 1988). The classifier tool provides the ability to generate a coinciding map of confidence. This map displays the prediction errors and thereby presents a spatial, visual representation of map accuracy. We generated a map of confidence for Zone 19 to examine this feature.

Results

The forest and non-forest stratification maps for zones 16 and 19 are displayed in figure 1. The classification model selected for Zone 16 used 12 boosting trials and a misclassification cost of 2, meaning the cost of misclassifying a non-forested plot as forested was doubled. This weighting compensated for the potential inaccuracies resulting from the fewer non-forested shrub and herbaceous training sites relative to the forested training sites. No variables were excluded from the model using the winnowing feature. The percent of plots correctly classified, according to the 10-fold cross-validation routine performed by See5, was 82.5 percent. For Zone 19, we also selected a classification tree using 12 boosting routines with a misclassification cost of 2. The 10-fold validation procedure identified the accuracy at 91.6 percent.

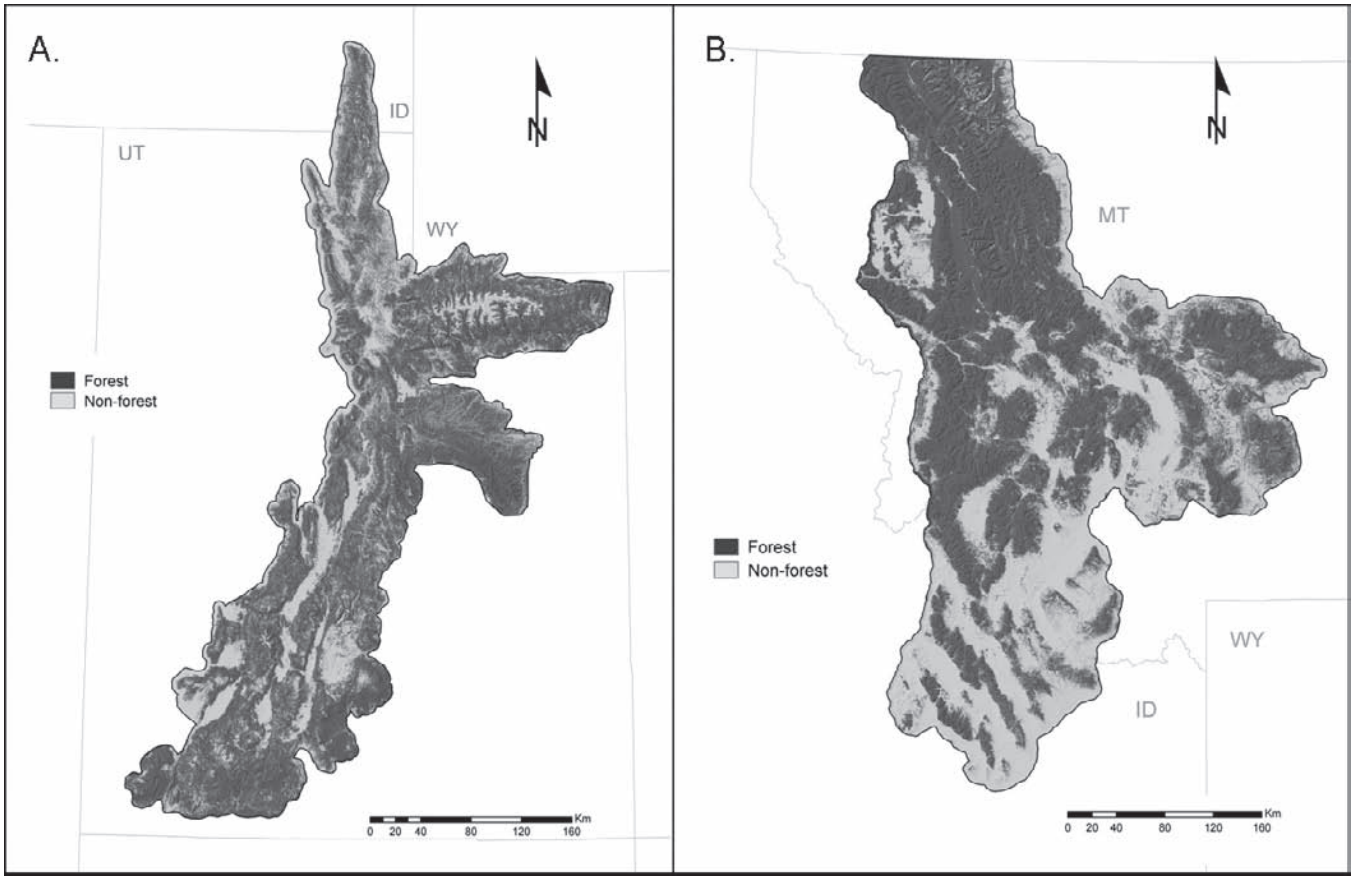

Figure 1—Forest and non-forest stratification maps. A, Zone 16; B, Zone 19.

The classification tree selected for Zone 16 forest PVTs used 10 boosting trials, and ten variables (gsws, outflow, pet, psi, psi.max, vmc, sdepth, aspect, slope, posidx) were winnowed from the model (table 3). The non-forest PVT classification tree for Zone 16 also used 10 boosting trials, and eleven variables (dsr, dss, gsws, mc1, outflow, posidx, psi.max, vmc, sdepth, aspect, lnd-frm) were winnowed (table 3). The classification tree we selected for Zone 19 forest PVTs used 14 boosting trials and used all the variables in the model. The non-forest PVT classification tree for Zone 19 used 16 boosting trials with two variables (tnight, srad.fg) winnowed (table 4). The variables that explain the most variance in the models are usually at the top of the classification tree, defining the initial breaks. For the Zone 19 forest classification tree, no variables were winnowed, and the variables that most often appeared at the top of the trees were snowfall, gl.sh, dday, dss, evap, pet, and tmin (table 4). For the Zone 19 nonforest classification tree, tnight and srad.fg were winnowed, and the prominent variables were gl.sh, ppt, pet, aspect, and dday (table 4).

The total percent of plots correctly classified for Zone 16 was 61 percent, with a kappa coefficient of 0.55 (table 7). For Zone 19, the total percent of plots correctly classified was 58 percent with a kappa coefficient of 0.54 (table 7). The error matrices for forest and non-forest PVTs in Zone 16 are shown in tables 5 and 6, respectively. The number of plots correctly classified is represented by the diagonal values in bold font. The total percent of plots correctly classified for the forested lands was 65 percent with a kappa coefficient of 0.55 (table 7). The percent of plots correctly classified for shrub and herbaceous lands was 48 percent, with a kappa coefficient of 4.0 (table 7). The user and producer accuracies for each class in Zone 16 is provided in table 8. User accuracies range from 0 percent for the Douglas-fir / Lodgepole Pine - Timberline Pine type to 89 percent for the Spruce – Fir / Spruce – Fir type. Producer accuracies range from 0 percent for the Douglas-fir / Timberline Pine type to 100 percent for the Blackbrush and Salt Desert Shrub types. Zero percent values are the result of having no test sites occurring within a particular class. Most of the lower user accuracies are within PVT subgroups. The Spruce – Fir / Blue – Spruce PVT has a user accuracy of only 15 percent (table 8). From the error matrix, we can see that 54 percent of

Table 5—Error matrix for Zone 16 forest PVTs. PVT codes are listed in table 1. The number of test sites correctly classified is shown in bold.

PVT	PVT Code									
Code	1	2	3	4	5	6	7	8	9	10
1	**2**	7	1	0	2	0	1	0	0	0
2	1	**82**	2	0	2	1	2	0	0	2
3	1	7	**23**	0	3	0	2	1	1	2
4	0	1	1	**0**	0	1	0	0	0	0
5	0	8	3	0	**3**	0	2	0	2	1
6	0	4	3	0	0	**1**	0	1	0	1
7	0	1	1	0	1	0	**5**	1	7	0
8	0	0	4	0	0	0	1	**6**	19	1
9	0	2	1	0	0	2	2	3	**84**	1
10	0	1	2	0	2	0	0	0	0	**6**

Table 6—Error matrix for Zone 16 non-forest PVTs. PVT codes are listed in table 1. The number of test sites correctly classified is shown in bold.

PVT	PVT Code										
Code	11	12	13	14	15	16	17	18	19	20	21
11	**2**	0	0	0	0	0	0	0	0	0	2
12	0	**2**	0	0	0	0	0	1	0	0	0
13	0	0	**8**	0	0	0	3	3	0	0	0
14	0	0	0	**1**	0	1	0	0	0	0	0
15	0	0	0	0	**3**	1	0	0	0	0	0
16	0	0	0	0	1	**8**	1	2	0	1	1
17	0	0	0	0	0	2	**5**	5	0	0	0
18	0	0	1	0	0	3	2	**9**	0	1	1
19	1	0	0	0	0	0	0	2	**1**	0	2
20	0	0	0	0	0	0	0	1	2	**1**	2
21	1	0	0	0	0	2	0	3	0	0	**3**

Table 7—Overall accuracies and kappa coefficients for Zone 16 a nd Zone 19.

Zone	Category	Overall accuracy	Kappa
16	Total	61.2	0.55
	Forest	64.8	0.55
	Shrub and herbaceous	47.8	0.40
19	Total	58.4	0.54
	Forest	56.5	0.49
	Shrub and herbaceous	66.4	0.58

Table 8—Zone 16 user and producer accuracy measures.

PVT code	PVT name	User accuracy	Producer accuracy
1	Spruce – Fir / Blue Spruce	15.4	50.0
2	Spruce – Fir / Spruce - Fir	89.1	72.6
3	Grand Fir - White Fir	57.5	60.5
4	Douglas-fir / Lodgepole Pine - Timberline Pine	0.0	0.0
5	Douglas-fir / Douglas-fir	15.8	23.1
6	Lodgepole Pine - Timberline Pine	14.3	20.0
7	Ponderosa Pine	31.3	33.3
8	Pinyon - Juniper / Mountain Big Sagebrush	19.4	50.0
9	Pinyon - Juniper / Wyoming - Basin Big Sagebrush	88.4	74.3
10	Riparian Hardwood	54.6	42.9
11	Riparian Shrub	50.0	50.0
12	Blackbrush - Chaparral - Dry Deciduous Shrub	66.7	100.0
13	Dwarf Sagebrush	57.1	88.9
14	Salt Desert Shrub	50.0	100.0
15	Mountain Mahogany	75.0	75.0
16	Gambel Oak	61.5	47.1
17	Wyoming - Basin Big Sagebrush	41.7	45.5
18	Mountain Big Sagebrush	52.9	34.6
19	Wetland Herbaceous	16.7	33.3
20	Alpine	16.7	50.0
21	Herbaceous	33.3	27.3

the test sites classified as Spruce – Fir / Blue Spruce were predicted as Spruce – Fir / Spruce – Fir (table 5). The Pinyon – Juniper / Mountain Big Sagebrush type had similar results. The user accuracy was 19 percent, but 61 percent of the Pinyon – Juniper / Mountain Big Sagebrush test sites were predicted as Pinyon – Juniper / Wyoming – Basin Big Sagebrush (table 6).

Error matrices for Zone 19 forest and non-forest PVTs are presented in tables 9 and 10, respectively. The total percent of plots correctly classified for forest PVTs was 57 percent, with a kappa coefficient of 0.49 and 66 percent for shrub and herbaceous PVTs, with a kappa coefficient of 0.58 (table 7). Table 11 shows the user and producer accuracies for each class in Zone 19. For Zone 19, the

Table 9—Error matrix for Zone 19 forest PVTs. PVT codes are listed in table 2. The number of test sites correctly classified is shown in bold.

PVT	PVT Code														
Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	**17**	2	3	0	1	0	0	0	0	0	0	0	0	0	0
2	3	**19**	7	0	2	2	0	0	0	0	0	0	0	0	0
3	7	3	**147**	17	41	5	6	1	6	1	0	1	0	0	0
4	0	0	7	**107**	16	0	0	0	0	1	0	0	2	0	0
5	1	3	42	35	**75**	4	6	0	0	5	0	0	0	0	0
6	1	3	5	0	1	**27**	3	1	12	0	2	1	0	0	0
7	0	0	20	4	3	2	**29**	1	22	7	0	0	0	0	0
8	0	0	1	3	2	0	1	**8**	10	1	0	0	0	0	0
9	0	2	11	1	4	10	13	3	**74**	5	0	1	0	0	1
10	0	0	2	3	6	0	4	0	6	**33**	1	0	0	0	0
11	0	0	0	0	0	2	0	0	1	0	**5**	0	0	0	0
12	0	0	1	0	0	0	0	0	3	0	0	**3**	0	0	0
13	0	0	0	2	0	0	0	0	0	2	0	0	**2**	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0
15	0	0	0	0	2	0	0	0	0	0	0	0	0	0	**0**

Table 10—Error matrix for Zone 19 non-forest PVTs. PVT codes are listed in table 2. The number of test sites correctly classified is shown in bold.

PVT	PVT Code										
Code	16	17	18	19	20	21	22	23	24	25	26
16	**2**	0	0	0	2	0	0	1	0	0	0
17	0	**1**	0	0	0	0	2	0	0	0	0
18	0	0	**3**	0	0	0	0	0	0	1	0
19	0	0	0	**1**	1	0	8	0	0	0	0
20	0	1	0	0	**28**	5	3	1	0	3	2
21	0	0	0	0	1	**18**	6	0	0	1	0
22	0	0	0	2	7	5	**61**	0	0	0	0
23	1	0	0	0	1	0	0	**6**	0	1	0
24	0	0	0	0	0	0	0	0	**1**	2	0
25	0	1	0	0	1	2	1	1	1	**15**	0
26	0	0	2	0	1	1	6	0	0	1	**12**

Table 11—Zone 19 user and producer accuracy measures.

PVT Code	PVT name	User accuracy	Producer accuracy
1	Western Redcedar	73.9	58.6
2	Grand Fir / White Fir	57.6	59.4
3	Spruce - Fir / Montane	62.6	59.8
4	Spruce - fir / Timberline	80.5	62.2
5	Spruce - Fir / Subalpine	43.9	49.0
6	Douglas-fir / Ponderosa Pine	48.2	51.9
7	Douglas-fir / Lodgepole Pine	33.0	46.8
8	Douglas-fir / Timberline Pine	30.8	57.1
9	Douglas-fir / Douglas-fir	59.2	55.2
10	Lodgepole Pine	60.0	60.0
11	Ponderosa Pine	62.5	62.5
12	Timberline Pine / Limber Pine	42.9	50.0
13	Timberline Pine / Whitebark Pine	33.3	50.0
14	Rocky Mountain Juniper	100.0	100.0
15	Riparian Hardwood	0.0	0.0
16	Riparian Shrub	40.0	66.7
17	Mountain Mahogany	33.3	33.3
18	Dry Shrub	75.0	60.0
19	Dwarf Sagebrush Complex	10.0	33.3
20	Mountain Big Sagebrush Complex	65.1	66.7
21	Threetip Sagebrush	69.2	58.1
22	Wyoming - Basin Big Sagebrush Complex	81.3	70.1
23	Wetland Herbaceous	66.7	66.7
24	Alpine	33.3	50.0
25	Fescue Grasslands	68.2	62.5
26	Bluebunch Wheatgrass	52.2	85.7

user accuracies range from 0 percent for the Riparian Hardwood type to 100 percent for the Rocky Mountain Juniper type. Again, we see similar patterns in the error matrices of within-subgroup inaccuracies. Forty-five percent of the Spruce – Fir / Subalpine test sites were misclassified as Spruce – Fir / Montane (Western Larch or Douglas-fir) or Spruce – Fir / Timberline, and 25 percent of the Spruce – Fir / Montane (Western Larch or Douglas-fir) test sites were misclassified as Spruce – Fir / Timberline or Spruce – Fir / Subalpine (table 9). Similarly, 19 percent of the Mountain Big Sagebrush test sites were misclassified as Threetip Sagebrush or Wyoming – Basin Big Sagebrush Complex, and 16 percent of the Wyoming – Basin Big Sagebrush Complex test sites were misclassified as Mountain Big Sagebrush or Threetip Sagebrush types (table 10). Thirty-six Douglas-fir sites were misclassified as Spruce – Fir / Montane (Western Larch or Douglas-fir) (table 10). The final PVT maps for zones 16 and 19 are presented in figure 2. The spatial estimate confidence for Zone 19 is shown in figure 3.

Discussion

The LANDFIRE PVT Mapping Approach

The LANDFIRE PVT mapping process represents an innovative framework for linking vegetation dynamics, such as post-disturbance recovery and succession, to landscape patterns represented by the biophysical variables compiled from the National Elevation Database and modeled using the WXFIRE model. The variables that were most important (defined by the first few splits of the tree) for the successful mapping of forest PVTs included: actual and potential evapotranspiration, days since snow, degree days, evaporation, relative humidity, leaf resistance to sensible heat, and minimum temperature. For the non-forest PVTs, the most important variables included: actual and potential evapotranspiration, precipitation, degree days, relative humidity, and minimum temperature. These gradients are associated with plant-water interactions and explain the influence of water and temperature derivatives in determining the distribution of vegetation across landscapes.

Classification and Regression Trees

Classification tree modeling proved an efficient means for identifying relationships between PVTs and biophysical variables across broad landscapes. With a comprehensive set of training data, classification trees can serve as strong predictors of these relationships. This predictive power extends across scales and is fully repeatable in time and space. Classification trees have proven successful in modeling and mapping vegetation at regional (Moisen and others 2003), national (Vogelmann and others 2001; Zhu and others, Chapter 8), and global scales (Hansen and others 2000). Although some research has found the predictive accuracy of classification trees to be inferior to other predictive modeling tools (Moisen and Frescino 2002; Pal and Mather 2003), the statistical flexibility, speed, and objectivity of the trees justify their use for large-scale mapping efforts such as LANDFIRE. See5 software adds efficiency to classification tree modeling by providing automated procedures, flexibility in terms of changing modeling functions, and by built-in accuracy measures.

Accuracy Assessment

There are several possible sources of the generally low accuracies found in the PVT maps created during the LANDFIRE Prototype Project. First, the performance of mapping models depends greatly on the quality of input data. The training databases for PVT mapping were collected and compiled from the LANDFIRE reference database (LFRDB), a database that comprised existing agency and non-agency field-referenced data sets and contained inventory, monitoring, and analysis data that originate from a variety of sampling objectives, sizes, and designs. (see Caratti, Ch. 4 for details). Data inaccuracies, major outliers, and unbalanced or insufficient numbers of training sites can have significant negative effects on the quality of mapping models (Friedman 2001). While the LFRDB was a large, comprehensive database that was compiled quickly and economically, the disparate sampling objectives, designs, and procedures certainly affected the final accuracies of the PVT maps.

A second possible explanation for the low accuracies is related to the model building characteristics of classification trees. As See5 builds classification trees, map units are divided using hard breaks, making it difficult to discriminate between vegetation types that have similar responses to the biophysical predictor variables. Most of the lower accuracies found during the LANDFIRE Prototype Project were within groups of similar PVTs, suggesting that these PVTs occur on overlapping biophysical settings, as represented by the predictor variables. The distributions of the three spruce-fir PVTs and the four Douglas-fir PVTs over a gradient of potential evapotranspiration are quite similar in Zone 19 (fig. 4). The error matrices reflect these similarities as well, indicating that the See5 classification tree algorithms had difficulty in discriminating these PVT subgroups. In any mapping application, this overlap between classes

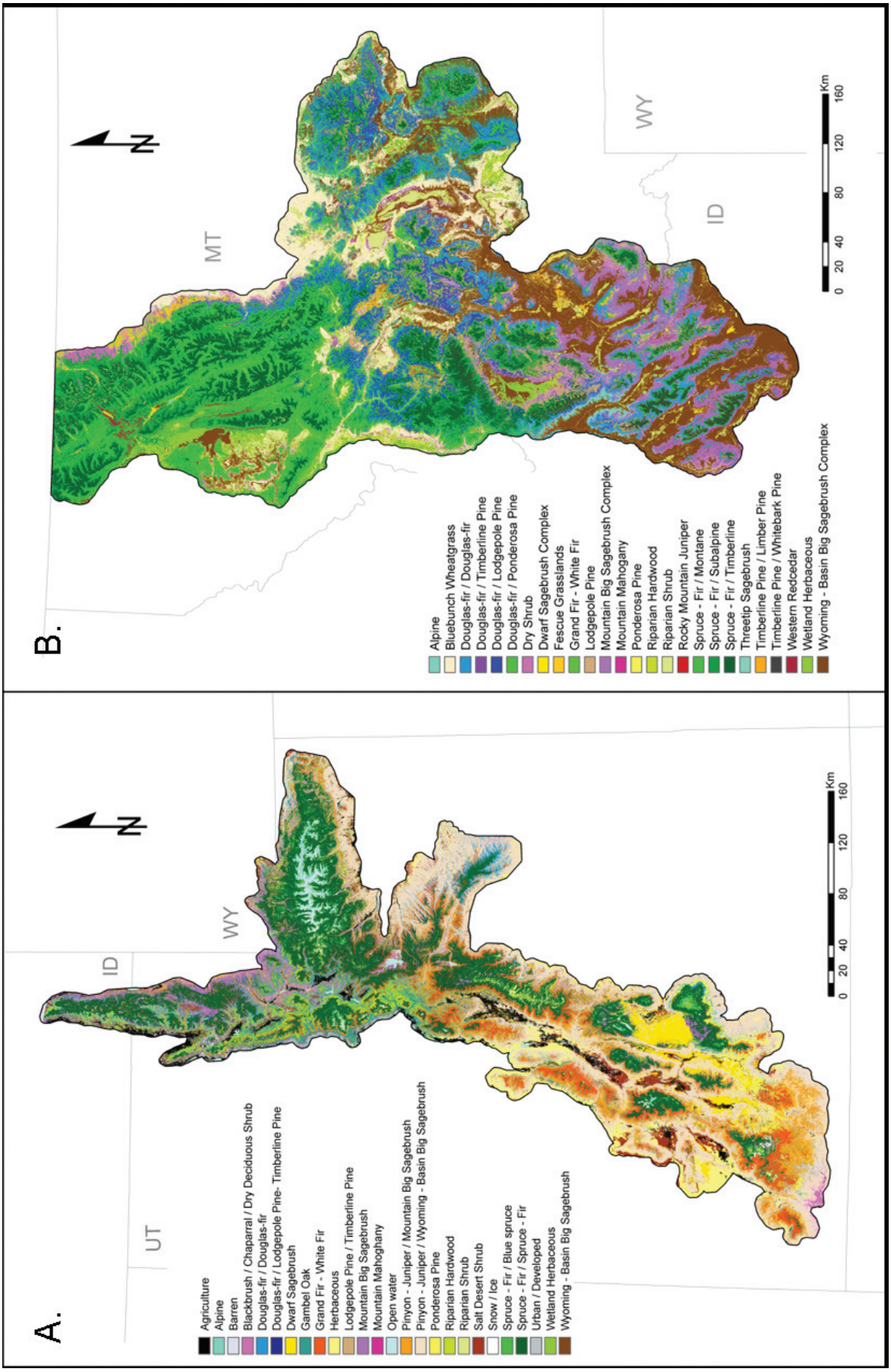

Figure 2—Final potential vegetation type (PVT) maps. A, Zone 16; B, Zone 19.

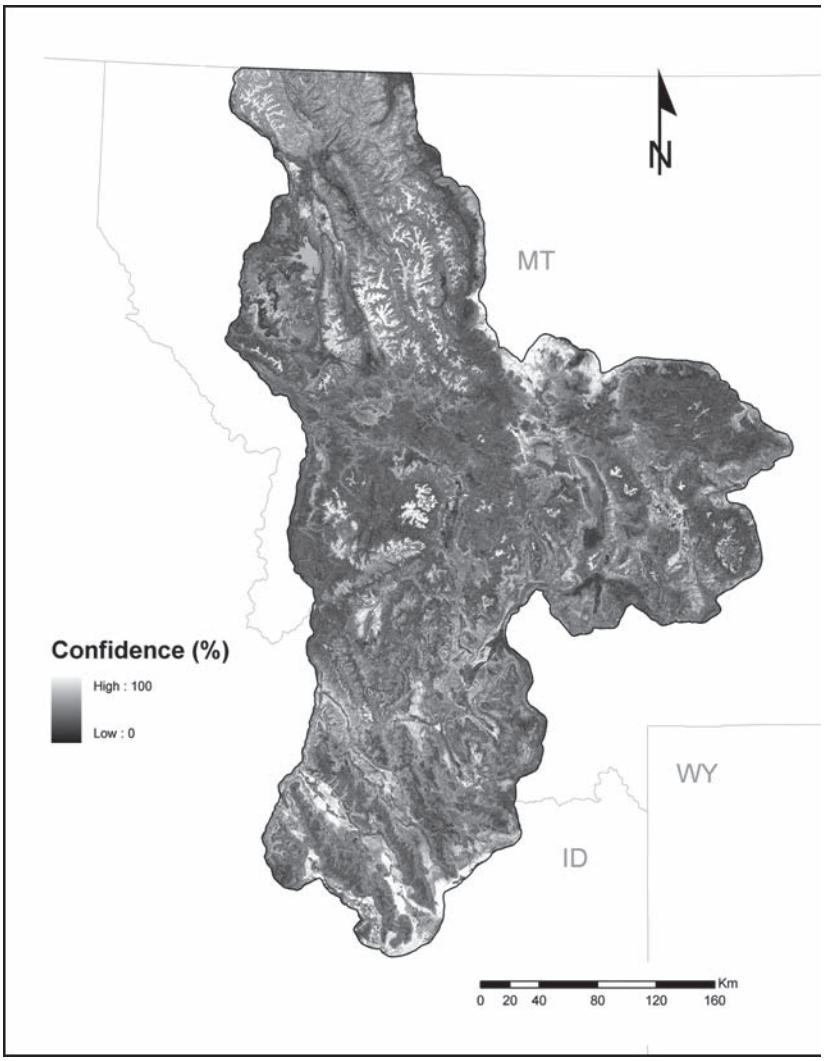

Figure 3—Potential Vegetation Type (PVT) Confidence map for Zone 19.

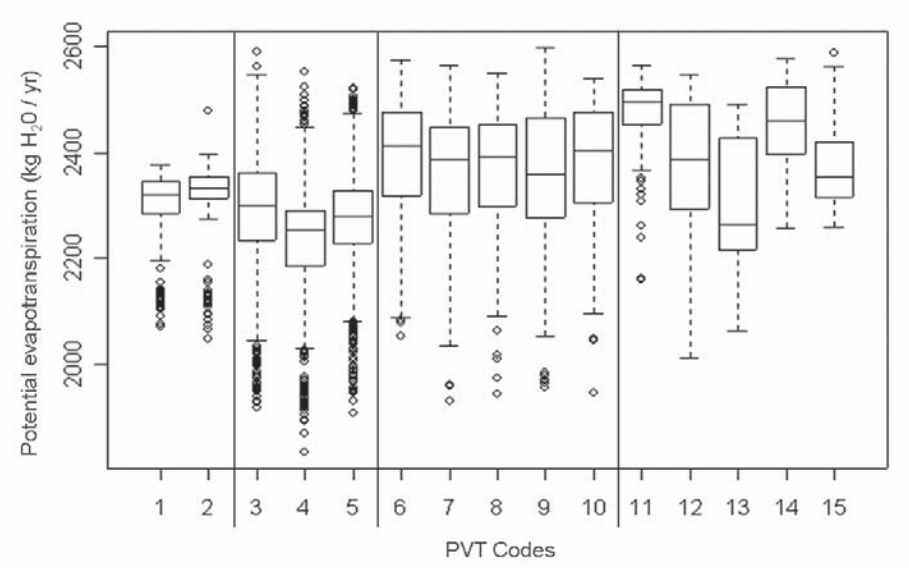

Figure 4—Boxplot distributions of potential vegetation types (PVTs) by potential evapotranspiration gradient. See table 1 for code descriptions. Codes 3 to 5 are Douglas-fir PVT variants and codes 6 to 10 are spruce – fir PVT variants.

negatively affects overall accuracy. Although accuracies may have been higher if we had grouped these PVT subgroups into single map units, we determined that the resulting loss in resolution in fire regime modeling would limit the utility of the final LANDFIRE fuel and fire regime products.

A third possible reason for overall low PVT map accuracies relates to the limited set of predictor variables used in PVT mapping. We did not include Landsat imagery in the mapping process because we did not want current land patterns influencing the final PVT maps; we relied completely on the affinity of individual PVT map units to specific distributions and combinations of biophysical variables. Further, because of technical difficulties, output from the LANDFIRE Biogeochemical Cycles model (LFBGC) (Holsinger and others, Ch. 5), which spatially represents the rates of the hydrologic, carbon, and nitrogen cycles, was not available in time to be used in the LANDFIRE Prototype Project and was therefore not included in the final mapping models. These ecophysiological gradients have proven to be highly useful in discriminating between potential vegetation types in other research (Keane and others 2001; Rollins and others 2004).

A fourth potential reason for low accuracies in the PVT maps lies in the possibility that the validation procedure we used did not represent true accuracy. The validation procedure used in the LANDFIRE Prototype Project included a cross-validation routine and a test set comparison using a randomly selected set of data withheld from classification tree building. Although, in both cases, the test sites were randomly selected from a probability sample, sampling was conducted at different intensities within different sub-populations. Therefore, more test sites are drawn from heavily sampled areas and fewer from less intensively sampled areas. Other possible sources of error include positional inaccuracies in the LANDFIRE reference database and errors imbedded in the biophysical predictor variables. It should be noted that quality control and assurance measures and methods for generating the biophysical gradient layers have been refined for national implementation (See Holsinger, Ch. 5 for details).

Recommendations for National Implementation

For mapping PVT at the national scale, we recommend employing the approach and methods described in this chapter. The efficiency and nonparametric flexibility of classification trees make them the optimal method for implementing LANDFIRE nationally, and the ease of implementation of the mapping models created using See5 software in ERDAS Imagine facilitate the broad-scale implementation of classification trees. We suggest conducting more structured quality control and assurance in the LANDFIRE reference database. In addition, we recommend detailed exploration of the relationships between response and predictors in the mapping database using correlation matrices and principle component analyses to reduce the number of gradient predictors and to remove major outliers or unusual patterns in the training data.

In addition, alternative validation sampling schemes should be considered for national implementation to ensure that the test sites are independent and representative of the population. For example, accuracy assessment sites developed solely from the systematically sampled Forest Inventory and Analysis data would ensure independent and representative test sites and therefore be a possible alternative as an equal probability sampling design. A similar procedure would be needed for shrub and herbaceous lands.

To compensate for positional errors in the training data set, we suggest employing alternative methods for calculating map accuracy when implementing LANDFIRE nationally. The agreement between each test site and its neighborhood of pixels (for example, 3 by 3) should be assessed. If the test site class matches any of the pixels, it correctly classifies the prediction. This kind of assessment is appropriate for plots in the LANDFIRE reference database that were not measured specifically for 30-meter pixel accuracy assessments.

Conclusion

In conclusion, maps of potential vegetation were valuable for supporting the broad-scale mapping of wildland fuel and also as a foundation for modeling fire regimes. The LANDFIRE process of generating biophysical gradients from topographic information and from the WXFIRE model served as an innovative framework for linking vegetation dynamics, such as post-disturbance recovery and succession, to landscape patterns represented by maps of potential vegetation. Although we found that the quality of field data for use as training data and of input spatial data layers can be limiting to the process of potential vegetation mapping, the LANDFIRE Prototype Project illustrated that the added effort involved in developing maps of potential vegetation results in higher quality data products representing fuel and fire regime characteristics.

For further project information, please visit the LANDFIRE website at www.landfire.gov.

The Authors

Tracey S. Frescino is a Forester with the USDA Forest Service, Rocky Mountain Research Station (RMRS), Interior West Forest Inventory and Analysis (FIA) Program. Frescino received a B.S. degree in Environmental Studies from SUNY's Environmental Science and Forestry program in 1991 and an M.S. degree in Fisheries and Wildlife from Utah State University in 1998. She has been with FIA since 1992 working as a field technician and a reporting analyst, and she is currently serving as a specialist in the FIA's techniques group. From spring of 2003 to spring of 2005, Frescino worked as an FIA collaborator for the LANDFIRE Prototype Project at the RMRS Missoula Fire Sciences Laboratory in Missoula, Montana. She was responsible for mapping potential vegetation and canopy fuel, in addition to facilitating access to and interpretation of FIA data for the LANDFIRE effort.

Matthew G. Rollins is a Landscape Fire Ecologist at the USDA Forest Service, Rocky Mountain Research Station, Missoula Fire Sciences Laboratory (MFSL). Rollins' research emphases have included assessing changes in fire and landscape patterns under different wildland fire management scenarios in large western wilderness areas, relating fire regimes to landscape-scale biophysical gradients and climate variability, and developing predictive landscape models of fire frequency, fire effects, and fuel characteristics. Rollins is currently the lead scientist of the LANDFIRE Project, a national interagency fire ecology and fuel assessment being conducted at MFSL and the USGS Center for Earth Resources Observation and Science (EROS) in Sioux Falls, South Dakota. He earned a B.S. degree in Wildlife Biology in 1993 and an M.S. degree in Forestry in 1995 from the University of Montana in Missoula, Montana. His Ph.D. was awarded by the University of Arizona in 2000, where he worked at the Laboratory of Tree-Ring Research.

Acknowledgments

Special thanks to Alisa Keyser and Gretchen Moisen of the USDA Forest Service, Rocky Mountain Research Station, Forest Inventory and Analysis for map design and technical review, respectively, and to John Caratti of Systems for Environmental Management for technical review.

References

Arno, S.F.; Simmerman, D.G.; Keane, R.E. 1985. Characterizing succession within a forest habitat type-an approach designed for resource managers. Res. Note INT-357. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 8 p.

Austin, M.P.; Cunningham, R.B.; Fleming, P.M. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. Vegetatio. 55:11-27.

Bauer, E.; Kohavi, R. 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Machine Learning. 36:105-142.

Bray, J.R.; Curtis, J.T. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecological Monographs. 27:325-349.

Breiman, L.; Frieman, J.H.; Olshen, R.A.; Stone, C.J. 1984. Classification and regression trees. Belmont, California: Wadsworth,.

Brown de Costoun, E.C.; Story, M.H.; Thompson, C.; Commisso, K.; Smith, T.G.; Irons, J.R. 2003. National Park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. Remote Sensing of Environment. 85:316-327.

Brzeziecki, B.; Kienast, F.; Wildi, O. 1993. A simulated map of the potential natural forest vegetation of Switzerland. Journal of Vegetation Science. 4:499-508.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 20(1):37-46.

Congalton, R.G. 1988. Using spatial autocorrelation effect upon the accuracy of supervised classification of land cover. Photogrammetric Engineering and Remote Sensing. 63:403-414.

Congalton, R.G.; Green, K. 1999. Assessing the accuracy of remotely sensed data: principles and practices. Boca Raton, FL: CRC/Lewis Press. 137 p.

Cooper, S.V.; Neiman, K.E.; Roberts, D.W.. 1991. Forest habitat types of northern Idaho: a second approximation. Gen. Tech. Rep. INT-236. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 143p.

Ferguson, D.E.; Morgan, P.; Johnson, F.D., eds. 1989. Proceedings – land classifications based on vegetation: Applications for resource management. Gen. Tech. Rep. INT-257. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Franklin, J. 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography. 19(4):474-499.

Friedl, M.A.; Brodley, C.E. 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment. 61:399-409.

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

Friedman, J. 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics. 29:1189-1232.

Galiván, R.G.; Fernández-González, F.; Blasi, C. 1998. Climatic classification and ordination of the Spanish Sistema Central: relationships with potential vegetation. Plant Ecology. 139:1-11.

Gleason, H.A. 1926. The individualistic concept of the plant association. Torrey Botanical Club Bulletin. 53:7-26.

Hansen, M.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R. 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing. 21:1331-1364.

Hargrove, W. W.; Luxmore, R.J. 1998. A New High-Resolution National Map of Vegetation Ecoregions Produced Empirically Using Multivariate Spatial Clustering. In: Proceedings of the Environmental Systems Research Institute Annual Users conference, Redlands CA: Environmental Systems Research Institute.

Hessburg, P.F.; Salter, R.B.; Richmond, M.B.; Smith, B.G. 2000a. Ecological subregions of the Interior Columbia River Basin, USA. Applied Vegetation Science. 3:163-180.

Hessburg, P.F.; Smith, B.G.; Salter, R.B.; Ottmar, R.D.; Alvarado, E. 2000b. Recent changes (1930's-1990's) in spatial patterns of interior northwest forests, USA. Forest Ecology and Management. 136:53-83.

Ihaka, R.; Gentleman, R. 1996. A language for data analysis and graphics. Journal of Computational and Graphical Statistics. 5(3):299-314.

Jensen, M.E.; Redmond, R.L.; Dibenedetto, J.P.; Bourgeron, P.S.; Goodman, I.A. 2000. Application of ecological classification and predictive vegetation modeling to broad-level assessments of ecosystem health. Environmental Monitoring and Assessment. 64:197-212.

Joy, S.M.; Reich, R.M.; Reynolds, R.T. 2003. A non-parametric, supervised classification of vegetation types on the Kaibab National Forest using decision trees. International Journal of Remote Sensing. 24(9):1835-1852.

Keane, R.E.; Mincemoyer, S.A.; Schmidt, K.M.; Long, D.G.; Garner, J.L. 2000. Mapping vegetation and fuels for fire management on the Gila National Forest complex, New Mexico. Gen. Tech. Rep. RMRS-GTR-46-CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Keane, R.E.; Burgan, R.E.; Wagtendonk, J. 2001. Mapping wildland fuels for fire management across multiple scales: integrating remote sensing, GIS, and biophysical modeling. International Journal of Wildland Fire. 10:301-319.

Keane, R.E.; Holsinger, L.M.; Pratt, S.D. 2006. Simulating historical landscape dynamics using the landscape fire succession model LANDSUM version 4.0. Gen. Tech. Rep. RMRS-GTR-171CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 73 p. [Online]. Available: http://www.fs.fed.us/rm/pubs/rmrs_gtr171.html [May 18, 2006].

Kuchler, A.W. 1975. Potential natural vegetation of the conterminous United States: Manual and Map, 2nd edition. New York, NY: American Geological Society.

Lenihan, J.M.; Neilson, R.P. 1993. A rule-based vegetation formation model for Canada. Journal of Biogeography. 20:615-628.

Moisen, G.G; Frescino, T.S. 2002. Comparing five modeling techniques for predicting forest characteristics. Ecological Modeling. 157:209-225.

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 meeting of the American Society for Photogrammetry and Remote Sensing, Anchorage Alaska. Washington D.C.: American Society for Photogrammetry and Remote Sensing.

Moore, D.M.; Lees, B.G.; Davey, S.M. 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographical information system. Environmental Management. 15(1):59-71.

Pal, M.; Mather, P.M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment. 86:554-565.

Pfister, R.D.; Kovalchik, B.L.; Arno, S.F.; Presby, R.C. 1977. Forest habitat types of Montana. Gen. Tech. Rep. INT-34. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Pfister, R.D.; Arno, S.F. 1980. Classifying forest habitat types based on potential climax vegetation. Forest Service. 26:52-70.

Quigley, T.M.; Graham, R.T.; Haynes, R.W. 1996. An integrated scientific assessment for ecosystem management in the Interior Columbia River Basin and portions of the Klamath and Great Basins. Gen. Tech. Rep. PNW-GTR-382. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 303p.

Quinlan, J.R. 1986. Induction of decision trees. Machine Learning. 1:81-106.

Quinlan, J.R. 1993. C4.5: programs for machine learning. San Mateo, CA: Morgan Kaufman Publishers.

Rollins, M.G.; Keane, R.E.; Parsons, R.A. 2004. Mapping fuels and fire regimes using remote sensing, ecosystem simulation, and gradient modeling. Ecological Applications. 14(1): 75-95.

Running, S.W.; Thornton, P.E. 1996. Generating daily surfaces of temperature and precipitation over complex topography. In: Goodchild, M.F.; Steyaert, L.T.; Parks, B.O., eds. GIS and environmental modeling: progress and research issues. Fort Collins, CO: GIS World, Inc.: 93-99.

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

Thornton, P.E.; Hasenauer, H.; White, M.A. 2000. Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. Agriculture and Forest Meteorology. 104:225-271.

Vogelmann, J.E.; Howard, S.M.; Yang, L., Larson, C.R.; Wylie, B.K.; Driel, N.V. 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering and Remote Sensing. 67(6):650-662.