J10D.5 PRELIMINARY LANDFIRE VEGETATION PRODUCTS IN THE WASATCH RANGE-UINTA MOUNTAIN AREA OF UTAH

Chengquan Huang, James Vogelmann and Brian Tolk, SAIC/EROS Data Center, Sioux Falls, SD 57198
Zhiliang Zhu, EROS Data Center, Sioux Falls, SD 57198

1. ABSTRACT

Geospatial data providing detailed information on vegetation composition and structure are critical for fire risks assessment, post-fire rehabilitation, and a range of ecological studies. In this study, we describe the development of vegetation products including cover type, canopy density and height for natural vegetation communities including forests, shrubs and grasses at the 30 m spatial resolution in the Wasatch Range-Uinta Mountains region of central Utah. Source data included Landsat ETM+ images and 30 m digital elevation data and derived slope, aspect and a position index. Ground reference data collected from over 6000 plots by different agencies were used. Vegetation cover type was modeled and predicted using a classification tree technique, while height and canopy density were modeled and predicted using both classification and regression tree techniques. Model results were evaluated through cross-validation. The derived preliminary vegetation products were evaluated using an independent reference data set collected in southern Utah.

2. LANDFIRE PROTOTYPE STUDY

Reliable and update-to-date vegetation products with adequate spatial and thematic details are critical inputs to fire risks and fuel models, post-fire rehabilitation strategy development, and a wide range of ecological studies and management applications. Developing such vegetation products at the national scope requires a mapping strategy that is repeatable and produces consistently reliable products. The LANDFIRE prototype study is being conducted to develop, evaluate and implement such a strategy. Two test areas – the Wasatch Range and Uinta Mountains of central Utah and the Northern Rocky Mountains of Montana, were selected for the prototype study (figure 1). In this work, we report the preliminary results developed in the Utah area.

Figure 1. Test areas for the LANDFIRE prototype study

3. MAPPING METHODS

3.1 *Relevant vegetation attributes*

Vegetation attributes required by the LANDFIRE project include vegetation cover type and two variables describing vegetation structure – canopy density and average canopy height, for three major vegetation life forms – forest, shrub, and grass. Cover type is defined at the alliance (community with multiple dominant species) or association (community with a single dominant species) levels as specified in the National Vegetation Classification Standard developed by the United States Federal Geographic Data Committee (Grossman, Faber-Langendoen, and others, 1998). The actual cover types mapped in the central Utah area are listed in table 1.

Table 1. LANDFIRE vegetation cover types mapped in the Utah study area

Forest	**Shrub**
Ponderosa Pine	Mt. Big Sagebrush Complex
Lodgepole Pine	Wyoming/Basin Big Sagebrush
Timberline Pine	Dwarf Sagebrush Complex
Douglas-fir	Blackbrush
Grand-fir/White-fir	Other Evergreen Shrubs
Spruce-fir	Dry Deciduous Shrubs
Pinyon-Juniper	Mountain Deciduous Shrubs
Juniper	Riparian Deciduous Shrubs
Riparian Hardwood	Salt Desert Shrub
Aspen-Birch	Rabbitbrush
Deciduous Oaks	**Grass**
Other Hardwood	Warm Season Perennial
Other	Cool Season Perennial
Open Water	Wetland Herbaceous
Urban/Developed	Alpine
Barren	Annual Grasses
Agriculture	Native Forbs
Snow/Ice	

Physically, canopy cover and average height are continuous variables. While ideally they should be modeled as continuous variables, in some applications they can be used as categorical variables. In this study, we evaluated the feasibility of modeling these two attributes both ways, i.e., as continuous variables and as categorical variables.

3.2 *Predictor geospatial data*

Major predictor variables for modeling the above vegetation attributes included Landsat images and a number of terrain variables including elevation, slope, aspect, and a topographic position index. In addition, a large number of ecophysiological variables describing environmental gradients were developed, based on which the potential vegetation type (PVT), or climax vegetation type, was modeled for each grid cell. All predictor variables had a spatial resolution of 30 m. The preliminary products reported in this work were developed using Landsat images and the terrain variables. PVT and the ecophysiological variables were not available when the preliminary products were developed.

3.3 *Machine learning algorithms*

Many algorithms have been developed for deriving vegetation information using satellite images (Hall, Townshend, and Engman, 1995; Zhu and Evans, 1994). Supervised classification tree and regression tree (CART) techniques were the main methods for developing the LANDFIRE vegetation products. The CART techniques were selected because they have a number of advantages over other methods. Classification tree, for example, is a non-parametric method and therefore is independent of the distribution of class signature and can handle both continuous and nominal input variables. It is fast to train and yet is often as accurate as, sometimes more accurate than, many other classifiers (Hansen, Dubayah, and DeFries, 1996; Huang, Davis, and Townshend, 2002). Several global land cover products have been or are being developed using classification tree methods (e.g. Friedl, Zhang, and others, 2002; Hansen, DeFries, and others, 2000), while regression trees were found robust for mapping continuous variables such as tree canopy density and percent imperviousness over large areas (e.g. Huang, Yang, and others, 2001; Yang, Huang, and others, 2003)

3.4 *Reference data and validation strategy*

The effectiveness of the CART algorithms relies on the availability of adequate, high quality reference data. For the central Utah area, reference data included both existing and newly collected field data. The preliminary products were developed using a total of 6210 field plots where detailed vegetation information was collected through intensive field work.

Two approaches were employed to validate the preliminary vegetation products in the central Utah area. One was cross validation, i.e., dividing the training samples into *N* equal sized subsets and using each subset to evaluate the model developed using the remaining subsets. Because each evaluation subset was not used in developing the model to be evaluated, each evaluation process can be considered "independent" of the corresponding model development. Therefore the average results of the *N* cross validations represent an "independent" assessment of the overall model developed using all training samples. Such an assessment, however, could be biased when training data does not represent the actual distribution of vegetation in a given study area. In addition, spatial auto-correlations among training samples often lead to overestimation of accuracy values by cross validation.

To evaluate the "biasness" of cross validation results in the Utah area, an independent reference data set was collected in the Dixie National Park in southern Utah, an area of about a quarter the size of the entire study area. Field plots were selected using a stratified random sampling design adjusted to avoid accessibility problems. A total of 138 field plots were visited in early summer of 2003, including 80 plots on forested land, 44 on shrub land, 12 on grassland, and 2 on barren. Information on all required vegetation attributes was collected for each plot.

4. COVER TYPE

According to the reference data sets, there are over 30 natural vegetation cover types in the Utah study area. After lumping the cover types with sample sizes that were too small for mapping purpose, 12 forest types, 10 shrub types and 6 grass types in addition to 5 other cover types (water, urban, barren, agriculture, and snow/ice) were mapped. Cover types of the three major life forms were modeled separately using the classification tree method. Cross validation shows that the cover types of each life form could be separated with overall accuracies of around 60% (table 2). The model for each life form was applied to all pixels within the study area to create three classifications, one for forest types, one for shrub types, and one for grass types. The three classifications were then integrated to create a single LANDFIRE vegetation cover type map using a more general land cover classification for the same area developed through the MRLC project (Homer, Huang, and others, 2002) (figure 2).

Table 2. Overall accuracy of cover types assessed by life-forms using 1) cross-validation (for the entire study area) and 2) independent assessment (in southern Utah only)

Life form	**Cross Validation**		**Independent Assessment**	
	Accuracy (%)	Number of cover types	Accuracy (%)	Number of cover types
Forest	59	12	75	6
Shrub	60	10	62	6
Grass	59	6	50	3

The developed cover type map was evaluated in southern Utah using the independently collected field reference data set. Because the sample size was too small for many cover types in the independently collected reference data set, only 6 forest types, 6 shrub types, and 3 grass types were assessed (table 2).

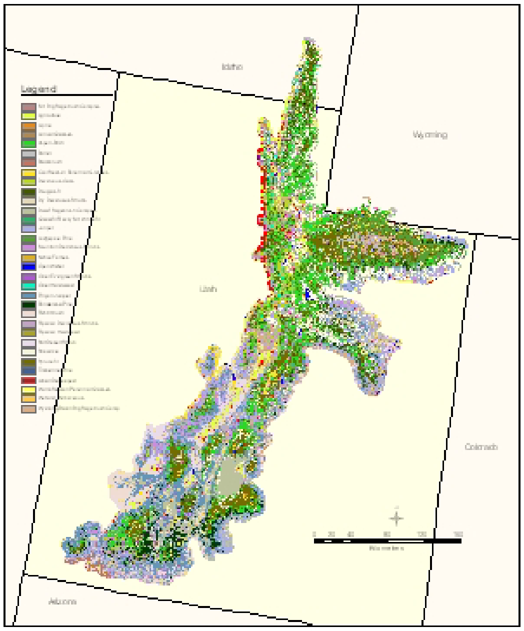

Figure 2. Preliminary LANDFIRE vegetation cover type map for the Utah study area

5. CANOPY DENSITY AND HEIGHT

Because physically both canopy density and height are continuous variables, we first used the regression tree technique to model them as continuous variables for each life form. Model performance was evaluated using two measures – average prediction error and correlation coefficient (*r*) between predicted and actual values. Cross validation estimates of the two measures are reported in table 3. As expected, this table shows that relationships between predicted and actual values were generally better for cover density than for height. Correlation coefficients of 0.5 or higher were achieved for cover density of all three life forms and for forest and shrub height. However, modeled grass height was poorly correlated with the actual value, suggesting that modeling this attribute as a continuous variable could be difficult.

Because in many applications, canopy density and height could also be used as categorical variables, we also modeled each attribute as a two-category (high and low) variable for each life form. Table 4 lists the overall accuracies estimated through cross validation for the entire study area. The independent field reference data set collected in southern Utah allowed independent assessments of these attributes (table 4). Results from both cross validation and independent assessment indicate that except for grass canopy density, high and low values of both canopy density and height could be separated with overall accuracy values ranging between 60% and 90%.

Table 3. Cross validation estimates of average prediction error and correlation coefficient (*r*) between predicted and actual values for canopy density and height of the three major life forms

Life form	Cover density		Height	
	r	Average error (%)	*r*	Average error (ft)
Forest	0.88	9.9	0.73	9.80
Shrub	0.63	12.2	0.50	1.45
Grass	0.55	11.5	0.20	0.35

Table 4. Overall accuracy (%) of high and low classifications of canopy density and height assessed by life-forms using 1) cross-validation (for the entire study area) and 2) independent assessment (in southern Utah only)

Life form	Canopy density		Height	
	Cross-validation	Inde-pendent	Cross-validation	Inde-pendent
Forest	92.3*	-*	80.0	85
Shrub	73.7	76.7	82.4	62.8
Grass	71.2	54.5	65.0	63.6

* The independently derived accuracy value for this attribute is not available because we haven't finished processing the field data for this attribute. The cross validation accuracy for this attribute could be overestimated, because significant spatial auto-correlation existed among the training samples for deriving tree canopy density.

The high and low classification models for both canopy density and height were applied to all pixels of the entire study area. The resultant classifications of canopy density and height for the three life forms were combined using the LANDFIRE cover type map (figure 2) to produce the LANDFIRE vegetation structure map (figure 3).

6. DISCUSSION

Tables 2 and 4 shows that for several vegetation attributes, cross validation estimates of the overall accuracy were close to those derived using the independent reference data set collected in southern Utah, demonstrating that under certain circumstances cross validation results can be reasonably unbiased. For other attributes, however, cross validation accuracy estimates were quite different from the independently derived values. Several factors should be considered in comparing cross validation accuracy estimates to the independently derived values listed in the two tables. First, the cross validation values were based on samples collected throughout the entire study area, while the independent test samples were collected from the southern part of the study area (roughly a quarter of the entire study area). Second, the independent test samples were collected using a stratified random sampling design, while the reference data used in model development and cross validation were collected using different sampling strategies depending on the attribute of concern and the source from which an attribute was derived. Third, the sample sizes for shrub and herbaceous in the independently collected data set in southern Utah were relatively small (40 and 12 for shrub and grass).

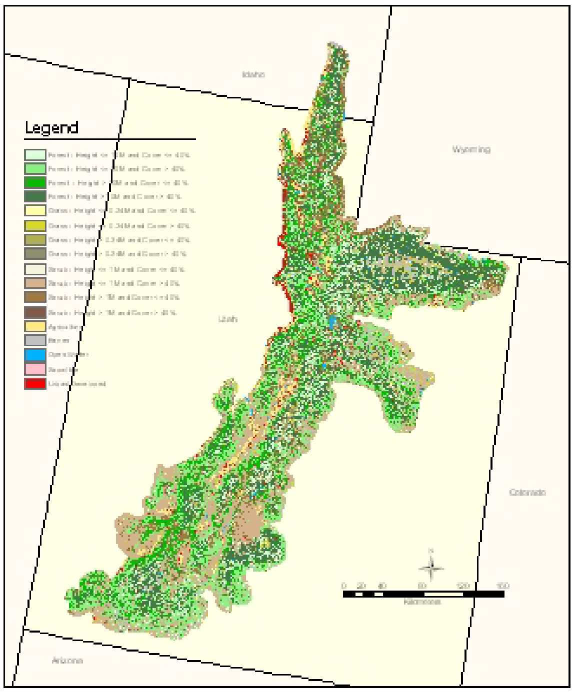

Figure 3. Preliminary LANDFIRE vegetation structure map for the Utah study area

Research activities are being conducted to improve the preliminary LANDFIRE vegetation products in a number of ways. First, the preliminary products were derived without using the potential vegetation type (PVT) and the large number of ecophysiological, or biophysical setting, data layers. We have seen some improvements in vegetation cover type mapping by using some of these data layers. Research is being conducted to identify the appropriate ways to use these data layers. Second, we had very limited number of training samples on shrub and grass vegetation on the eastern part of the mapping zone in producing the preliminary products. Additional training samples were collected this summer in that area. Use of these samples likely will improve the derived vegetation products. Third, we are investigating alternative techniques for modeling percent cover and height of shrub and grass vegetation. One problem in modeling grass structural attributes is that they are highly variable from season to season, making it difficult to match field measurements to satellite observations. We plan to incorporate annual measurements derived using MODIS data to minimize uncertainties arising from phenological differences.

7. REFERENCES

Friedl, M.A., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C., McIver, D.K., and Hodges, J.C.F., 2002, Global land cover mapping from MODIS: Algorithms and early results: Remote Sensing of Environment, v. 83, no. 1-2, p. 287-302.

Grossman, D.H., Faber-Langendoen, D., Weakley, A.S., Anderson, M., Bourgeron, P., Crawford, R., Goodin, K., Landaal, S., Metzler, K., Patterson, K., Pyne, M., Reid, M., and Sneddon, L., 1998, International classification of ecological communities: terrestrial vegetation of the United States. Volume I. The national vegetation classification system: development, status, and applications: Arlington, VA, The Nature Conservancy, 127 p.

Hall, F.G., Townshend, J.R., and Engman, E.T., 1995, Status of remote sensing algorithms for estimation of land surface state parameters: Remote Sensing of Environment, v. 51, p. 138-156.

Hansen, M., DeFries, R.S., Townshend, J.R.G., and Sohlberg, R., 2000, Global land cover classification at 1 km spatial resulution using a classification tree approach: International Journal of Remote Sensing, v. 21, no. 6/7, p. 1331-1364.

Hansen, M., Dubayah, R., and DeFries, R., 1996, Classification trees: an alternative to traditional land cover classifiers: International Journal of Remote Sensing, v. 17, no. 5, p. 1075-1081.

Homer, C., Huang, C., Yang, L., and Wylie, B., 2002, Development of a circa 2000 land cover database for the United States, *in* 2002 ASPRS-ACSM Annual Conference and FIG XXII Congress, Washington, D.C., The American Society for Photogrammetry & Remote Sensing, CD ROM, 1 disk.

Huang, C., Davis, L.S., and Townshend, J.R.G., 2002, An assessment of support vector machines for land cover classification: International Journal of Remote Sensing, v. 23, no. 4, p. 725-749.

Huang, C., Yang, L., Wylie, B., and Homer, C., 2001, A strategy for estimating tree canopy density using Landsat 7 ETM+ and high resolution images over large areas, *in* Third International Conference on Geospatial Information in Agriculture and Forestry, Denver, Colorado, CD-ROM.

Yang, L., Huang, C., Homer, C.G., Wylie, B.K., and Coan, M.J., 2003, An approach for mapping large-area impervious surfaces: Synergistic use of Landsat 7 ETM+ and high spatial resolution imagery: Canadian Journal of Remote Sensing, v. 29, no. 2, p. 230-240.

Zhu, Z., and Evans, D.L., 1994, US forest types and predicted percent forest cover from AVHRR data: Photogrammetric Engineering & Remote Sensing, v. 60, no. 5, p. 525-531.