Blue Mountains Mapping Zone Documentation:
Mapping Methods and Accuracy

1) Predictor Layer Preparation:

a) Image Standardization:

Standardization from DN values to at-sensor reflectance was performed by EROS Data Center as part of the MRLC Preprocessing Procedure. This procedure uses the method presented by Huang et. al (2001a) to transform Landsat 7 ETM+ imagery (Note, Landsat 5 used for some dates). The equation used for reflectance was as follows:

$$r_{BandN} = (\Pi((DN_{Band} * Gain_{Band} + Bias_{BandN}) * D2)/(E_{BandN}(E_{BandN}(SINq))$$

r_{BandN} = Reflectance for Band N

DN = Digital Number for Band N
D = Normalized Earth-Sun Distance
E_{BandN} = Solar Irradiance for Band N

q = Solar Elevation
$Gain_{BandN}$ = Provided within header file, and band specific
$Bias_{BandN}$ = Provided within header file, and band specific

b) Image Dates and Mosaic:

Images were mosaicked using ERDAS Imagine 8.7 Mosaic Tool with "no cutline" for *type*, and the "Overlay" option for *overlap function*.

Image dates and scenes were as follows:

Path/Row	Spring Sat	Spring Julian Date	Spring Calender	Spring Overlay Order	Summer Sat	Summer Julian Date	Summer Calender	Summer Overlay Order	Fall Sat	Fall Julian Date	Fall Calender	Fall Overlay Order
42029	7	01115	4/25/2001	4	7	99206	7/25/1999	4	7	00257	9/13/2000	4
43028	7	02093	4/3/2002	1	7	02205	7/24/2002	1	7	00280	10/6/2000	1
44028	7	01081	3/22/2001	5	7	02196	7/15/2002	5	7	01273	9/30/2001	5
43030	7	02157	6/6/2002	3	7	02221	8/9/2002	3	7	02285	10/12/2002	3
43029	7	00120	4/29/2000	2	7	02205	7/24/2002	2	7	00264	9/20/2000	2
44029	5	02124	5/4/2002	6	7	02164	6/13/2002	6	7	01273	9/30/2001	6
42028	7	01115	4/25/2001	7	7	02230	8/18/2002	7	7	01273	9/30/2001	7

c) Image Derived Datasets:
Normalized Difference Vegetation Index (NDVI): This dataset was provided by the EROS Data Center as part of the MRLC processing output.

Tasseled cap: Brightness, Greenness & Wetness band transformations were provided by the EROS Data Center as part of the MRLC processing output. This transformation followed the methods of Huang et. al 2001b.

Fractional Vegetation: The percent of ground covered by photosynthetic vegetation was estimated by the equation of Carlson and Ripley (1997). Reference values used in the equation were identified by examination of NDVI histograms and locating known sites of

bare soil and irrigated agricultural fields. The equation is N* = (SQRT ((NDVI - NDVImin)/(NDVImax - NDVImin))) * 100, where NDVImin is the NDVI value for bare soil pixels in the scene, and NDVImax is the NDVI value for fully vegetated pixels in the scene. Fractional vegetation is related to calculations of impervious surface (Imp. Surface = 100 - fractional vegetation). The output is an integer layer with values between 0 and 100.

bm_summer_fv = (SQRT((bm_ndvi_summer - 0.07) / (0.9 - 0.07))) * 100
Image Texture: The texture of the image is

d) DEM Derived Datasets:
Thirty-meter digital elevation models were obtained from the EROS Data Center, National Elevation Database (NED, 1999). DEMs were converted from floating point grids to integer grids and mosaicked for the region, then clipped to the mapping area.

Slope: A slope layer was created using the ARC/INFO SLOPE command. Values represent slope in degrees.

Topographic Relative Moisture Index: A TRMI grid (values ranging from 0-28) was created using an Arc/INFO AML obtained from the Southwest Regional GAP Project created by G. Manis (Manis et. al 2001). The TRMI model is based on the methods defined by Haplin, P. N. 1999, and Parker, A. J. 1982.

Landform: A 10 class landform grid was created using an Arc/INFO AML obtained from the Southwest Regional GAP Project created by G. Manis (Manis et. al 2001). This product was derived from the topographic relative moisture index.

For modeling purposes all arcinfo grids were converted to ERDAS Imagine .img files.

2) Samples:

a) Sample Collection Methods:

Samples were collected in a variety of ways. Originally, it was thought that most, if not all, of the sampling would be derived from field collected information polygons delineated over imagery in the field by field crews. Classification trees, however, require substantial amounts of training data so that additional information had to be acquired. All samples were assigned a label corresponding to an Ecological System (Comer et. al 2003). On the ground data was collected as polygons delineated over imagery in the field by ORNHIC and USGS field crews. A listing of the number of ground points by source is provided below. On screen digitizing was done using ETM imagery as a backdrop.

Source Sites

Map Region	Shrub Sample (2003)	Nevada ReGap	Bureau of Land Mangement			US Forest Service	Gifford Pinchot NF	Malheur NF	Mt Hood NF	Ochoco NF	Umatilla NF	Wallowa NF	Winema NF	Screen Digitized
	Training		Burns	Lakeview	Vale	Freemont NF	Gifford Pinchot NF	Malheur NF	Mt Hood NF	Ochoco NF	Umatilla NF	Wallowa NF	Winema NF	
Blue Mountains	47							8800			5824	10123		68

3) Cover Types:
a) Classification Tree Modeled Cover Types:

Twenty-two cover types were modeled in this zone. All forest and barren types were modeled and later reclassified into single classes in the overall Tree modeling.

The following cover types were modeled using the See5 Classification Tree:

Ecological System	Total	Code
Annual/Biannual Farmland	5	280
Columbia Basin Foothill and Canyon Dry Grassland	991	137
Columbia Plateau Scabland Shrubland	7	44
Columbia Plateau Low Sagebrush Steppe	2	154
Inter-Mountain Basins Big Sagebrush Shrubland - ssp. Wyoming	22	54
Inter-Mountain Basins Big Sagebrush Shrubland - ssp. tridintata	0	149
Inter-Mountain Basins Big Sagebrush Steppe - ssp. Wyoming	9	78
Inter-Mountain Basins Big Sagebrush Steppe - ssp. tridintata	0	150
Inter-Mountain Basins Juniper Savanna	107	40
Inter-Mountain Basins Montane Sagebrush Shrubland	1	71
Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	12	50
Inter-Mountain Basins Volcanic Rock and Cinder Lands	268	13
Invasive Annual Grass	8	308
North Pacific Dry--Mesic Silver Fir-Western Hemlock-Douglas-Fir Forest	10205	178
Northern Rocky Mountain Ponderosa Pine Woodland and Savanna	5222	146
Open Water	16	211
Recently Logged	5	310
Rocky Mountain Lodgepole Pine Forest	3466	31
Rocky Mountain Lower Montane Foothills Shrubland	650	47
Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	3751	28
Rocky Mountain Aspen Forest and Woodland	5	23
Temperate Pacific Montane Wet Meadow	42	103

b) Non Tree modeled Cover Types:
Screen digitized over ETM:

Agriculture-Irrigated and Non-Irrigated (280) were supplemented and modified using the Oregon Department of Agriculture's boundaries which represent all farms, orchards and pasture lands in the state. It was not possible to use the data tables supplied with the boundaries did not differentiate between different land use types. To determine actual agricultural fields from other land cover types the fractal dimension was calculated for all polygons in the data table. All values of less than 0.97 (simple geometric shapes) and with area greater 5 ha were burned into the tree model as agriculture.

Recently Burned (302) areas were developed from Fires in Western North America by visually comparing the shapefile to the ETM imagery and selecting those fire scars

visible on the imagery and those fires that occurred after the image date. Source of shapefile: USGS Snake River Field Station, 2004, http://sagemap.wr.usgs.gov

Interstates and highways were buffered by 30m to develop a transportation mask.

4) Summary of Predictor Layers Used:

a) Multi band predictors:
ETM bands 1-5 & 7 for fall
ETM bands 1-5 & 7 for spring
ETM bands 1-5 & 7 for summer

b) Single Band Predictors:

Fall Tasseled Cap brightness band
Fall Tasseled Cap greenness band
Fall Tasseled Cap wetness band
Summer Tasseled Cap brightness band
Summer Tasseled Cap greenness band
Summer Tasseled Cap wetness band
Spring Tasseled Cap brightness band
Spring Tasseled Cap greenness band
Spring Tasseled Cap wetness band

Continuous (integer) slope
Continuous (integer) elevation
Categorical 10 class landform (from DEM)
Topographic Relative Moisture Index (from DEM)
** Post Classification

5) Modeling Methods:

a) See5 Classification Tree Modeling:
Sub-Sampling: Pseudo-replication within each sample polygon was conducted in order to increase the number of samples used by the classification algorithm. While this use of non-independent data is not ideal for classification tree modeling, it has been found to improve classification accuracies, particularly when there are limited amounts of training data. 30 random points were placed within each polygon using the Arcview tool Hawth's Tool. The points were converted to pixels while ensuring that the resulting pixels (the new grid) aligned with the raster predictor layers. The resulting sub-sampled pixels would often be less than 30 per sample polygon, if random points fell within the same pixel.

Training and Validation Sites: Twenty percent of the all sample polygons were withheld for validation using the Feature Select extension in ArcView. With the remaining 80%,

30 sub-samples were randomly generated within each sample polygon. This was done by first randomly generating points within each polygon using the Random Points extension in Hawth's Tool for ArcVGIS, later converted to *.img files. Individual points were converted to *.img files (each to be considered a separate observation for the See5 classifier) and were `drilled' through predictor layers using the Sampling tool from CART Module for Imagine (EarthSatellite Corp. 2003), producing two important files for See5: the *.names and *.data files.

See5 Classification Tree: See5 (Release 2.01) data mining software (Rulequest 2004) was used for generating classification trees. Boosting was employed using 20 trials.

The following briefly describes the files used by the See5 Program (Rulequest 2004).

*.names file: Identifies the dependent variable *.img file and the predictor *.img files created from the CART Module Sampling tool. Required by See5 software.
*.data file: Contains the training cases from which See5 extracts rules. This is also produced from the CART Module Sampling tool, by `drilling' the dependent variable pixels through the specified predictor images. Required by See5 Software.
*.set file: Produced from See5 software. This file contains the settings for the classification tree run. For example the third value `15' indicates the number of boosts used for boosting.

*.tree file: Produced from the See5 software. This file contains the classification tree in `tree' format. This along with the *.data and *.names file are required by the CART Module Classifier tool to spatially apply the tree.

*.out file: Output file generated by See5 and displayed when See5 classification tree model has completed the final run. This file provides a visual representation of the classification tree that is somewhat easier to interpret than the *.tree file.

As a result of spatially applying the classification tree using the CART Module's Classifier an *.img file, which is the spatial application of the tree's rules was created.

Tree model Validation: The map comprised only 47 samples and did not offer an opportunity to complete validation following the guidelines in other map zones.

c) Post Classification, Recoding and Other Modeling Steps: Post-classification modeling was done to map classes where there were not enough training data to map using the full Tree model, to differentiate between ecological systems that have similar characteristics, use ancillary data sets, or correct problems found during qualitative review. The processes are listed in the order in which they were implemented.

Corrections
A visual examination of the 100% Tree model was done using ETM imagery.

Open Water
Water was over-mapped throughout the canyon regions of the mapping area. Primary areas of error were due to shadows in the deep canyons in which the available image dates did not penetrate the full spatial extent of the canyon wall. Canyon terrain models were used to remove shadow-water conflicts. Groups of misclassified pixels were replaced using a focal majority routine which used the surrounding pixels as representative values.

Forest
Conifer Forest (242) is a composite type composed of all individual ecological systems present within the map zone.

Recent Fires
The recent fires layer was applied as a mask over the landcover map superseding all underlying ecological systems with the Recently Burned (302) land cover type

Transportation
Following the lead of the SWGAP project, Interstates and Highways were applied to the final product. The transportation layer, recoded as Developed Medium (223), was applied as a mask over the landcover map using a conditional statement. See 3b for development

c) Generalizing to MMU and Map Completion:
Once the Tree model and the post-classification steps were employed, the map was generalized using the Clump tool in ERDAS Imagine 8.7. The parameter of 4 connected neighbors was used in the clumping process. Isolated pixels that fell under the specified 1-acre (5 pixels) minimum mapping unit (MMU) were removed using the Eliminate tool. Following the Clump & Eliminate step, the non-Tree modeled classes were then "burned in" to the final map using the Overlay function.

6) Accuracy Assessment

Accuracy statistics were not calculated using the 20% withheld samples on the preliminary Tree model.

7) Citations:

Carlson, T. N. and D. A. Ripley. 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sensing of Environment. 62:241-252.

Comer, P., D. Faber-Langendoen, R. Evans, S. Gawler, C. Josse, G. Kittel, S. Menard, M. Pyne, M. Reid, K. Schulz, K. Snow, and J. Teague. 2003. Ecological systems of the United States: A working classification of US terrestrial systems. NatureServe, Arlington, Virginia.

EarthSatellite. 2003. CART Software User's Guide, prepared by EarthSatellite

Corporation for the US Geological Survey in support of the National Land Cover Database (NLCD) 2000. EarthSatellite, January 2003.

Haplin, P. N. 1999. "GIS Analysis for Conservation Site Design:A short-course developed for the Nature Conservancy" Nicholas Schoolof the Environement--Landscape Ecology Lab, Duke University.

Huang, C.L., J. Vogelmann, B. Tolk, and Z. Zhu. 2003. Preliminary Landfire Vegetation Products in the Wasatch Range-Uinta Mountain Area of Utah. (http://ams.confex.com/ams/pdfpapers/65875.pdf)

Huang, C. L. Yang, C. Homer, B. Wylie, J. Vogelman and T. DeFelice, 2001, At-sensor reflectance: A first order normalization of Landsat 7 ETM+ Images. (http://landcover.usgs.gov/pdf/huang2.pdf)

Huang, C., B. Wylie, C. Homer, L. Yang, and G. Zylstra, 2001b, Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. International Journal of Remote Sensing, 8: 1741-1748.

Manis, G., J. Lowry and R. D. Ramsey, 2001, Preclassification: An ecologically predictive landform model. GAP Analysis Bulletin No. 10. USGS.

Rulequest, 2004. See5: An informal tutorial. WWW URL http://www.rulequest.com/see5-win.html (23 August, 2004).

Parker, A. J. 1982. The Topographic Relative Moisture Index: AnApproach to Soil-Moisture Assessment in Mountain Terrain. PhysicalGeography, 1982, 3,2, pp.160-168.