

Northern Rockies 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 outline" for *type*, and the "Overlay" option for *overlap function*.

Image dates and scenes were as follows:

Path/Row	Spring Sat	Spring Julian Date	Spring Calendar	Spring Overlay Order	Summer Sat	Summer Julian Date	Summer Calendar	Summer Overlay Order	Fall Sat	Fall Julian Date	Fall Calendar	Fall Overlay Order
42027	7	01115	4/25/2001	4	7	42027	02230	8/18/2002	7	00257	9/13/2000	4
43027	7	00216	8/3/2000	3	7	43027	02221	8/9/2002	7	99277	10/4/1999	3
44026	5	03159	6/8/2003	7	7	44026	02196	7/15/2002	7	01273	9/30/2001	7
44027	7	02164	6/13/2002	1	7	44027	02196	7/15/2002	7	00271	9/27/2000	1
45026	5	01080	3/21/2001	6	7	45026	00198	7/16/2000	7	99291	10/18/1999	6
45027	7	00086	3/26/2000	2	7	45027	00198	7/16/2000	7	00278	10/4/2000	2
43026	7	03144	5/24/2003	5	7	43026	02205	7/24/2002	7	99277	10/4/1999	5

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.

Shrub Cover: Overall percent of shrub cover was described following Jennings et. al 2004. Overlap in the top three occurring shrubs strata where addressed by the following:

$$C_i = \left(1 - \prod_{j=1}^n (1 - \% \text{cov } j / 100) \right) * 100$$

where C_i is the percent cover of stratum i for species or growth form j in stratum i .

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 $N^* = (\text{SQRT}((\text{NDVI} - \text{NDVI}_{\text{min}})/(\text{NDVI}_{\text{max}} - \text{NDVI}_{\text{min}}))) * 100$, where NDVI_{min} is the NDVI value for bare soil pixels in the scene, and NDVI_{max} 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.

$\text{nr_summer_fv} = (\text{SQRT}((\text{nr_ndvi_summer} - 0.07) / (0.9 - 0.07))) * 100$

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								Screen Digitized
	Training		Burns	Lakeview	Vale	Freemont NF	Gifford Pinchot NF	Malheur NF	Mt Hood NF	Ochoco NF	Umatilla NF	Wallowa NF	Winema NF	
Northern Rockies	131													355

3) Cover Types:

a) Classification Tree Modeled Cover Types:

Nineteen 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 Systems	Total	Code
Agriculture-Irrigated and Non-Irrigated	210	280
Columbia Basin Foothill and Canyon Dry Grassland	11	137
Columbia Plateau Scabland Shrubland	7	44
Inter-Mountain Basins Big Sagebrush Shrubland - ssp. wyoming	11	54
Inter-Mountain Basins Big Sagebrush Steppe - ssp. wyoming	32	78
Inter-Mountain Basins Montane Sagebrush Shrubland	34	71
Inter-Mountain Basins Semi-Desert Shrub-Steppe	1	79
Invasive Annual Grassland	11	308
Invasive Perennial Forbland	5	307
Northern Rocky Mountain Ponderosa Pine Woodland and Savanna	18	146
Open Water	145	211
Temperate Pacific Freshwater Emergent Marsh	1	199

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.

c) *Cover Types Modified with a Post-Classification model:*
 (see section 5c for details)

Ecological Systems	Action
Agriculture-Irrigated and Non-Irrigated	Modified with agricultural boundaries
Inter-Mountain Basins Big Sagebrush Shrubland - ssp. wyoming	Shrub Cover applied to split for steppe vs shrubland
Inter-Mountain Basins Big Sagebrush Steppe - ssp. wyoming	Shrub Cover applied to split for steppe vs shrubland
Inter-Mountain Basins Montane Sagebrush Shrubland	Shrub Cover applied to split for steppe vs shrubland
Inter-Mountain Basins Semi-Desert Shrub-Steppe	Shrub Cover applied to split for steppe vs shrubland

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

Tree model Shrub Cover Percentage (integer) **

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.

b) Shrub Cover: Sampling protocols followed by project teams (section 2a) required the collection of a visual estimate of percent coverage of individual shrub strata. Following similar methodology used in trial regions of SW-ReGap (Huang et. al 2003, Jennings et. al 2004) a overall percent shrub cover was estimated for each training site (80%/20% training/validation). The total percent coverage is represented as a continuous surface at each site, and was reclassified to five categorical types following guidelines suggested by LandFire.

Category	Range %
Very High	> 45%
High	36-45%
Moderate	26-35%
Steppe	11-25%
Grassland	<10%

Tree model Validation: Twenty percent of the sample polygons were randomly selected and withheld from Tree modeling. The preliminary Tree models were run as described in section 5a using the remaining 80% of the training site data. The 20% withheld samples were used to assess the predictive capability of the Tree modeled map via the kappa_stats.avx extension for ArcView by intersecting the reference polygons with the Tree modeled land cover map. This extension considers the site correctly mapped when the majority of pixels within the reference polygon agree with the reference label. Output from kappa_stat.avx includes the kappa statistic and an error matrix indicating errors of omission and commission.

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.

Post-Classification Tree modeling

Shrub Coverage: Unique combinations of all shrub dominated ecological systems and the independent shrub cover were modified to represent a Shrub -> Steppe -> Grassland type based upon the ecoregion of occurrence.

Ecological System	Shrub Cover				
	Very High	High	Moderate	Steppe	Grassland
Inter-Mountain Basins Big Sagebrush Shrubland - ssp. wyoming	Shrubland	Shrubland	Shrubland	Steppe	Steppe
Inter-Mountain Basins Big Sagebrush Steppe - ssp. wyoming	Shrubland	Shrubland	Steppe	Steppe	Grassland
Inter-Mountain Basins Montane Sagebrush Shrubland	Shrubland	Shrubland	Shrubland	Steppe	Steppe
Inter-Mountain Basins Semi-Desert Shrub-Steppe	Shrubland	Shrubland	Steppe	Steppe	Grassland

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 calculated using the 20% withheld samples on the preliminary Tree model. These statistics include an error matrix, kappa statistics and breakdown of user's/producer's accuracy and error.

Error Matrix

ERROR MATRIX: Reference Data in Columns/Classification Data in Rows	Columbia Plateau Scabland Shrubland	Exotics	Northern Rocky Mountain Ponderosa Pine Woodland and Savanna	Inter-Mountain Basins Semi-Desert Shrub-Steppe	Inter-Mountain Basins Semi-Desert Grassland	Inter-Mountain Basins Big Sagebrush Shrubland-ssp. wyoming	Columbia Basin Palouse Prairie	Inter-Mountain Basins Montane Sagebrush Shrubland	Total	User Accuracy
Columbia Plateau Scabland Shrubland	3	0	0	0	0	0	0	0	3	100%
Exotics	0	0	0	0	0	0	0	0	0	Null
Northern Rocky Mountain Ponderosa Pine Woodland and Savanna	0	0	4	0	0	0	0	1	5	80%
Inter-Mountain Basins Semi-Desert Shrub-Steppe	0	0	0	1	0	0	0	0	1	100%
Inter-Mountain Basins Semi-Desert Grassland	0	0	0	0	0	0	0	0	1	0%
Inter-Mountain Basins Big Sagebrush Shrubland-ssp. wyoming	1	3	0	0	1	11	1	2	20	55%
Columbia Basin Palouse Prairie	0	0	0	0	0	0	0	0	0	Null
Inter-Mountain Basins Montane Sagebrush Shrubland	0	0	0	0	0	0	0	2	2	100%
Total	4	3	4	1	1	11	1	5	32	
Producer Accuracy	75%	0%	100%	100%	0%	100%	0%	40%		66%

User's/Producer's Accuracy and Omission/Commission Error

Ecological Systems	User Accuracy	Secificity	Predictive Power	Omission Error	Commissi on Error
Null	Null	1	0.9375	1	0
Columbia Plateau Scabland Shrubland	1	1	0.965517	0.25	0
Exotics	Null	1	0.90625	1	0
Northern Rocky Mountain Ponderosa Pine Woodland and Savanna	0.8	0.964286	1	0	0.035714
Inter-Mountain Basins Semi-Desert Shrub-Steppe	1	1	1	0	0
Inter-Mountain Basins Semi-Desert Grassland	0	0.967742	0.967742	1	0.032258
Inter-Mountain Basins Big Sagebrush Shrubland- ssp. wyoming	0.55	0.571429	1	0	0.428571
Columbia Basin Palouse Prairie	Null	1	0.96875	1	0
Inter-Mountain Basins Montane Sagebrush Shrubland	1	1	0.9	0.6	0

Overall Statistics:

Overall Accuracy: $(21 / 32) = 0.656250000$

Overall Misclassification Rate: $(11 / 32) = 0.343750000$

Overall Sensitivity: 0.656250000

Overall Specificity: 0.957031250

Overall Omission Error: 0.343750000

Overall Commission Error: 0.042968750

Kappa Statistics

KHAT = 0.53684

VARIANCE = 0.01205230

Z = 4.89

P = 0.00000000

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 School of the Environment--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. Vogelmann 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.

Michael Jennings M., D. Faber-Langendoen, R. Peet, O. Loucks, D. Glenn-Lewin, A. Damman, M. Barbour, R. Pfister, D. Grossman, D. Roberts, D. Tart, M. Walker, S. Talbot, J. Walker, G. Hartshorn, G. Waggoner, M. Abrams, A. Hill, M. Rejmanek, 2004. *Guidelines for describing associations and alliances of the U.S. National Vegetation Classification, V4*. The Ecological Society of America, Vegetation Classification Panel.

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: An Approach to Soil-Moisture Assessment in Mountain Terrain. *Physical Geography*, 1982, 3,2, pp.160-168.