Sawtooth 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. The equation used for reflectance was as follows:

$$\rho_{BandN} = (\Pi((DN_{BandN} * Gain_{BandN} + Bias_{BandN}) * D^2) / (E_{BandN}(SIN\theta))$$

Where,
ρ_{BandN} = Reflectance for Band N
DN_{bandN} = Digital Number for Band N
D = Normalized Earth-Sun Distance
E_{bandN} = Solar Irradiance for Band N
θ = Solar Elevation
$Gain_{bandN}$ = band specific, provided in the header file sceneid.h1
$Bias_{bandN}$ = band specific, provided in the header file sceneid.h1

b) Image Dates and Mosaic:

Images were mosaicked using ERDAS Imagine 8.5 Mosaic Tool with "no cutline" for *type*, and the "Overlay" option for *overlap function*.

Image dates and scenes were as follows:

Path/Row	Spring (Year - Julian Day)	Summer (Year - Julian Day)	Fall (Year - Julian Day)	Spring Overlay Order	Summer Overlay Order	Fall Overlay Order
40/29	2001 - 141	2001 - 213	1999 - 256	2	2	2
40/30	2001 - 149	2000 - 227	1999 - 256	1	1	1
41/28	2000 - 122	2002 - 191	2000 - 282	6	6	6
41/29	2000 - 154	2002 - 191	2000 - 282	7	7	7
41/30	2001 - 124	2002 - 191	2000 - 282	4	4	4
42/28	2001 - 115	2002 - 230	1999 - 254	5	5	5
42/29	2001 - 115	1999 - 206	2000 - 257	3	3	3

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.

sawt_summer_fv = (SQRT((sawt_ndvi_summer - 0.07) / (0.9 - 0.07))) * 100

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 USGS Snake River Field Station field crews. A listing of the number of ground points by source is provided below. On screen digitizing was done using 2004 NAIP or ETM imagery as a backdrop.

Source	Sites
USGS Snake River Field Station	609
Idaho Conservation Data Center	429
Landfire (Utah State University)	39
Pacific Northwest National Laboratory	62
On Screen Digitized	1139
Region 1 Forest Service Ecoplot Data	333
Total	2611

3) Cover Types:

a) Classification Tree Modeled Cover Types:

Thirteen cover types were modeled in this zone. All forest, riparian, juniper, and barren types were modeled as single classes in the overall CT modeling.

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

Ecological System	Value	Sites
Grassland	1	215
Forest	2	1235
Riparian	3	121
North American Alpine Ice Field	4	32
North American Arid Emergent Marsh	7	15
Columbia Plateau Scabland Shrubland	44	22
Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	50	19
Inter-Mountain Basins Big Sagebrush Shrubland	54	97
Inter-Mountain Basins Montane Sagebrush Steppe	71	496
Inter-Mountain Basins Big Sagebrush Shrub-Steppe	78	73
Northern Rocky Mountain Lower Montane Mesic Deciduous Shrubland	151	86
Columbia Basin Low Sagebrush Steppe	154	19
Barren	330	181
Total		2611

b) Non CT Modeled Cover Types:

Screen digitized over ETM and NAIP imagery at a scale between 1:24,000 and 1:100,000:

Updated the 2004 Farm Service Agency Common Land Unit GIS Shapefiles. Source of GIS Shapefile: Idaho Farm Service Agency, digitized and certified 2004.

Updated the Digital Database of Mining-Related Features at Selected Historic and Active Phosphate Mines, Bannock, Bear Lake, Bingham, and Caribou Counties, Idaho to include mines in other portions of Idaho. Source of GIS Shapefile: Idaho State University, 2001

Updated the Idaho City Limits shapefile using the 2004 Color NAIP imagery. Source of shapefile: Idaho State Tax Commission. This shapefile was then used to develop the 4 developed classes.

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. Sources: Idaho Department of Lands Interstates and Highways and Oregon BLM Highways.

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

Ecological System	Value
Rocky Mountain Alpine Bedrock and Scree	2
Rocky Mountain Cliff and Canyon	6
Inter Mountain Basins Cliff and Canyon	9
Rocky Mountain Aspen Forest and Woodland	23
Great Basin Pinyon-Juniper	40
Inter Mountain Basins Juniper Savanna	75
Rocky Mountain Subalpine Mesic Meadow	83
Inter-Mountain Basins Semi-Desert Grassland	90
Northern Rocky Mountain Montane Grassland	133
Columbia Plateau Steppe and Grassland	135
Columbia Basin Foothill and Canyon Grassland	137
Open Water	211
Developed - Open Space/Parks	221
Developed - Low	222
Developed - Medium	223
Developed - High	224
Conifer Forest	242
Agriculture - Irrigated and NonIrrigated	280
Recently Burned	302
Recently Mined or Quarried	303

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

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. 20 random points were placed within each polygon using an Arcview Avenue script. 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 20 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%, 20 sub-samples were randomly selected for each sample polygon. This was done by first randomly generating points within each polygon using the Random Points extension in ArcView and then converting the points to a raster *.img file. Pixels in the *.img (each to be considered a separate observation for the See5 classifier) 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 1.8) 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 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.

CT Model Validation: Twenty percent of the sample polygons were randomly selected and withheld from CT modeling. The preliminary CT 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 CT modeled map via the kappa.avx extension for ArcView by intersecting the reference polygons through the CT 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.avx includes the kappa statistic and an error matrix indicating errors of omission and commission.

Final CT Modeled Map: The objective of this project was to produce the best map possible. With this objective in mind, the next step was to generate a final decision tree model using 100% of the available sample data. Following the same procedures as in section 5a a final CT model was developed using the same model inputs as the 80% model except the additional of the 20% withheld data that was used for the validation.

b) 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 CT 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 CT Modeling

Overmapping Corrections

A visual examination of the 100% CT Model was done using 1 meter color NAIP imagery. Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland and Columbia Plateau Scabland Shrubland were overmapped throughout the mapping area. An image was created with those pixels representing the overmapped classes having a value of 1 and all others as 0 using a conditional statement in the Imagine Model Builder. A new 100% CT model (FinalCT_no5044) was created without including Inter-Mountain Basins Mountain Mahogany Woodland or Shrubland and Columbia Plateau Scabland Shrubland in the model and using the mask to only map those areas that had been mapped as Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland and Columbia Plateau Scabland Shrubland. An additional mask was created delineating those areas where Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland and Columbia Plateau Scabland Shrubland occurred within the mapping area. Using conditional statements, pixels that mapped Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland and Columbia Plateau Scabland Shrubland in the 100% CT Model but where not actually Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland and Columbia Plateau Scabland Shrubland were replaced with values from FinalCT_no5044.

Forest

Conifer and Rocky Mountain Aspen Forest and Woodland were differentiated using a separate CT model. This was done due to the limited number of training points and the focus of the primary CT classification being on the shrubland classes. The total number of sites used to model the general forest class was revised to include only those that could be classified definitively as one of the output classes. This step was necessary due to large number of digitized training data points that were potentially a mixed forest type. The following is a breakdown training data numbers:

Ecological System	Value	Sites
Conifer Forest	242	1045
Rocky Mountain Aspen Forest and Woodland	23	24

Grassland

The grassland types were differentiated using a separate CT model in order to maximize the small number of training points. The following is a breakdown training data numbers:

Ecological System	Value	Sites
Rocky Mountain Subalpine Mesic Meadow	83	16
Northern Rocky Mountain Montane Grassland	133	118
Columbia Plateau Steppe and Grassland	135	25
Columbia Basin Foothill and Canyon Grassland	137	49

Post Classification non-CT Modeled Types

Inter Mountain Basins Cliff and Canyon, Rocky Mountain Cliff and Canyon and Rocky Mountain Alpine Bedrock and Scree
A post classification model was used to model Inter Mountain Basins Cliff and Canyon and split the high elevation barren classification into Rocky Mountain Cliff and Canyon and Rocky Mountain Alpine Bedrock and Scree because training data was very limited. Conditional statements were used to reclassify pixels as Inter Mountain Basins Cliff and Canyon and split the barren class.

The detail of the conditional statement follows:
EITHER 9 IF (sawt_lf_theme >= 9 AND sawt _fv_summer < 20 AND (sawt Final CT Map NE 330) OR sawt Final CT Map OTHERWISE

EITHER 6 IF (sawt mz_lf_theme >= 9 AND sawt mz_fv_summer < 20 AND sawt Final CT Map imbcc EQ 330) OR sawt Final CT Map imbcc OTHERWISE

EITHER 2 IF (sawt Final CT Map imbcc rmcc EQ 330) OR sawt Final CT Map imbcc OTHERWISE

Where:
sawt mz_lf_theme = The 10-class landcover model for the mapzone
sawt mz_fv_summer = The summer fractional vegetation for the mapzone
sawt Final CT Map = CT output image with 100 % samples
sawt Final CT Map imbcc = CT output image with 100 % samples with the addition of Inter Mountain Basins Cliff and Canyon
sawt Final CT Map imbcc rmcc = CT output image with 100 % samples with the addition of Inter Mountain Basins and Rocky Mountain Cliff and Canyon Systems

Developed – Open Space, Low, Medium, and High
The developed classes were developed using a digitized boundary file and fractional vegetation. The digitized developed areas boundaries were used to clip the fractional vegetation layer to the developed areas. Then the clipped fractional vegetation imagine was recoded using the following values:

Open Space: Fractional Vegetation = 80 - 100
Low Intensity: Fractional Vegetation = 50 - 79
Medium Intensity: Fractional Vegetation = 20 - 49
High Intensity: Fractional Vegetation = 0 - 19

Agriculture
The agriculture layer was applied as a mask over the landcover map using a conditional statement. See 3b for development.

Water
The Open Water cover type was developed using an unsupervised classification on the summer Landsat imagery. Each cluster was assigned a water or non-water class through visual inspection. The water classes were then combined using a conditional statement (EITHER 211 IF (openwater_unsup = water) OR 0 OTHERWISE). Through a visual examination of the output compared to ETM and 2004 NAIP imagery any necessary corrections to the classes were done using a spatial recode. The result was applied to the final CT model as a mask.

Recent Fires
The recent fires layer was applied as a mask over the landcover map using a conditional statement. See 3b for development.

Recent Mining
The recent mines layer was applied as a mask over the landcover map using a conditional statement. See 3b for development.

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 CT model and the post-classification steps were employed, the map was generalized using the Clump tool in ERDAS Imagine 8.5. 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-CT modeled classes were then "burned in" to the final map using the Overlay function.

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

Error Matrix

Ecological System	1	2	3	4	7	44	50	54	71	78	151	154	330	Total	User's Accuracy
Grassland	26	3	0	0	0	0	1	1	5	1	0	0	0	37	70%
Forest	2	163	2	0	2	0	1	0	7	0	1	0	0	178	92%
Riparian	1	2	22	0	1	0	0	0	0	0	0	0	0	26	85%
North American Alpine Ice Field	0	0	0	3	0	0	0	0	0	0	0	0	0	3	100%
North American Arid Emergent Marsh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Columbia Plateau Scabland Shrubland	2	0	0	0	0	5	0	0	0	0	0	0	0	7	71%
Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Inter-Mountain Basins Big Sagebrush Shrubland	0	0	0	0	0	0	0	12	2	4	0	1	0	20	60%
Inter-Mountain Basins Montane Sagebrush Steppe	6	6	0	0	0	0	1	3	86	3	7	2	1	115	75%
Inter-Mountain Basins Big Sagebrush Shrub-Steppe	0	0	0	0	0	0	0	5	0	4	0	0	0	9	44%
Northern Rocky Mountain Lower Montane Mesic Deciduous Shrubland	1	3	0	0	0	0	0	0	0	0	10	0	0	14	71%
Columbia Basin Low Sagebrush Steppe	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0%
Barren	5	0	0	0	0	0	0	0	0	1	0	1	31	38	82%
Total	43	177	24	3	3	5	3	21	100	14	18	4	32	448	
Producer's Accuracy	60%	92%	92%	100%	0%	100%	0%	57%	86%	29%	56%	0%	97%		81%

Kappa Statistics

Kappa: 0.748874
Standard error of kappa: 0.0228072
Z-Score for kappa: 32.835

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

Ecological Systems	Absolute Producers Accuracy	Absolute Errors of Omission	Absolute Users Accuracy	Absolute Errors of Commission	Number of Field Points
Grassland	60.5%	39.5%	70.3%	29.7%	43
Forest	92.1%	7.9%	91.6%	8.4%	177
Riparian	91.7%	8.3%	84.6%	15.4%	24
Snow and Ice	100.0%	0.0%	100.0%	0.0%	3
Wetland	0.0%	100.0%	0.0%	100.0%	3
Columbia Plateau Scabland Shrubland	100.0%	0.0%	71.4%	28.6%	5
Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	0.0%	100.0%	0.0%	100.0%	3
Inter-Mountain Basins Big Sagebrush Shrubland	57.1%	42.9%	60.0%	40.0%	21
Inter-Mountain Basins Montane Sagebrush Steppe	86.0%	14.0%	74.8%	25.2%	100
Inter-Mountain Basins Big Sagebrush Shrub-Steppe	28.6%	71.4%	44.4%	55.6%	14
Northern Rocky Mountain Lower Montane Mesic Deciduous Shrubland	55.6%	44.4%	71.4%	28.6%	18
Columbia Basin Low Sagebrush Steppe	0.0%	100.0%	0.0%	100.0%	4
Barren	96.9%	3.1%	81.6%	18.4%	32

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. 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.