SWReGAP Land Cover Mapping Methods Documentation

Functional Unit or Mapping Zone: CO-4 (MMZ9)

(Colorado's Eastern Plains.) Occasionally referred to as Mega-mapzone 9 (MMZ9).

Organization: Colorado Division of Wildlife, Denver, CO

Person Preparing Document: Eric Waller

Date of Preparation: 23 July, 2004

1) Predictor layer preparation:

a) Image standardization:

Standardization from DN values to at-sensor reflectance was performed on Landsat 7 ETM+ imagery using methods presented by Huang et. al (2001a). The resulting reflectance values were multiplied by 400, to make use of the range of byte data. 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
[Colorado specifically avoided performing any dark object subtraction. This is consistent with the U.S.G.S. EROS Data Center (EDC), which does not perform the dark object subtraction on their imagery. It is thought that this technique may not offer much improvement, accounting for path radiance but not accounting for atmospheric transmission. It can also be difficult to ascertain appropriate dark object values, especially in images with variable atmospheric conditions or regions with variable atmospheric depth (such as mountainous areas).]

b) Image dates and mosaicking:

Images were mosaicked using ERDAS Imagine 8.6 Mosaic Tool with "no cutline" for *type*, and the "Overlay" option for *overlap function*.
Image dates and scenes were as follows (due to the slightly varying orientation of the scene paths, a small piece of scene 34/32 was only needed for the fall composite):

ETM+ Scene (path/row)	Spring (yr-Julian date	Summer (yr-Julian date	Fall (yr-Julian date
31/33	2000-100	2000-196	2000-292
31/34	2000-132	2001-166	2000-292
32/32	2000-123	2000-219	2000-283
32/33	2000-107	2001-157	2000-283
32/34	2000-107	2001-173	2000-251
33/32	2001-132	2000-210	2000-290
33/33	2000-113	2000-226	1999-191
33/34	2000-130	1999-191	1999-287
34/32	Not needed	Not needed	1999-310

Images showing overlap arrangement (precedence given to lower numbers) of the scenes and their dates for each of the mosaics are available at:

/CO/ARCHIVE/CO4_MMZ9/MOSAICS/mz9seqsp.img
/CO/ARCHIVE/CO4_MMZ9/MOSAICS/mz9seqsu.img
/CO/ARCHIVE/CO4_MMZ9/MOSAICS/mz9seqfa.img

These images are highly useful in modeling. Without them, the model would have no information on which particular image the spectral information is derived from. This information is critical, since image standardization does not account for all of the differences between images (e.g. atmospheric and seasonality differences).

The fall scene sequence image above used in modeling appeared to have a strange problem depicting the small piece of 34/32 in the top left corner. A "corrected" image fixed that problem (but has other minor strange problems, and is a slightly different size than the others???):

/CO/ARCHIVE/CO4_MMZ9/MOSAICS/mz9seqfa_corrected.img

Six band ETM mosaics for each season can also be found at:

/CO/ARCHIVE/CO4_MMZ9/MOSAICS/mmz9ref_spr.img
/CO/ARCHIVE/CO4_MMZ9/MOSAICS/mmz9ref_sum.img
/CO/ARCHIVE/CO4_MMZ9/MOSAICS/mmz9ref_fall.img

c) Image derived datasets:

Tasseled cap: Brightness, Greenness & Wetness band transformations were created using coefficients derived for the Landsat 7 ETM+ sensor, by Huang, et. al (2001b). The .gmd file can be found at:
/CO/ARCHIVE/CO4_MMZ9/IMG_FILES/MODELS/eros_tcap_mod.gmd

Normalized Difference Vegetation Index (NDVI): Initially calculated as float data (possible data ranging between -1 and 1), with a model **(CO/ARCHIVE/CO4_MMZ9/IMG_FILES/MODELS/my_ndvi_float.gmd)** that performed the normalized band ratio (band4 - band3)/ (band4 + band3). Later, when it was determined that byte data were required, another model **CO/ARCHIVE/CO4_MMZ9/IMG_FILES/MODELS/nd2nd_byte.gmd)** was used to make all values positive with a possible range from 0 to 200 (by adding one and then multiplying by 100).

Non-Photosynthetic Vegetation Index (NPV): A newly developed index for this project, similar to NDVI in that it is a normalized ratio that ranges from -1 to +1, this ratio highlights ETM+ band 5 information that was not very evident in the new Tasseled Cap transforms. In earlier versions, the Tasseled Cap "Wetness" was somewhat inversely related to non-photosynthetic vegetation (and band 5). Non-photosynthetic vegetation

(and band 5) information can help differentiate many cover types. The lack of NPV information in the new transform made this ratio useful. The ratio is (band5 – band4 – band3)/ (band5 + band4 + band3). One is added and the result multiplied by 100 to scale the ratio between 0 and 200. The .gmd file can be found at:
/CO/ARCHIVE/CO4_MMZ9/IMG_FILES/MODELS/npv_ratio_c.gmd

All imagery derived predictor layers can be found at:
/CO/ARCHIVE/CO4_MMZ9/IMG_FILES/.

d) DEM derived datasets:

Thirty meter digital elevation models were obtained from the EROS Data Center, National Elevation Database (NED). The date for these data was October 1999. DEMs were converted from floating point grids to integer grids and mosaicked for the region, then clipped to the mapping area.
Slope: This is expressed in degrees, as calculated by ERDAS Imagine's slope routine.
Aspect: A nine class aspect grid was created. Values 1=N, 2, NE, 3=E, 4=SE, 5=S, 6=SW, 7=W, 8=NW, 9 = FLAT.

Landform: A 10 class landform grid was created from a topographic relative moisture (values ranging from 0-28) index grid (Manis et. al 2001).

For modeling purposes all arcinfo grids were converted to ERDAS Imagine .img files and can be found at:

/CO/archive/CO4_MMZ9/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. Each type of data can be distinguished by the SITEID field in the polygon coverage:

SWReGAP data – polygons collected in the field by Colorado field crew personnel. The SITEID contains CO for Colorado, followed by the date (mmddyy), the collector's initials, and a two digit number to
distinguish the site from others collected on the same day.

"Basinwide" data - points collected in the field by another recent Colorado vegetation mapping effort: the Colorado Vegetation Classification Project (http://ndis.nrel.colostate.edu). As points, they had to be buffered (45 meters) to provide polygons (approximately 3 pixels by 3 pixels) that could be used in our modeling. They usually contain a field person's initials and a date etc., but they might be best distinguished by not following the naming convention of the other sources.

"Augmentation" data – these are polygons screen digitized by image analysts guided by field information, aerial photography (http://terraserver-usa.com), satellite imagery and/or context. These were often generated in areas that were persistently and obviously misclassified in modeling, and to replace imprecise field data, such as a Basinwide point located on a road. The SITEID begins with an "AD" for "Augmentation Data", followed by the date (mmddyy), the analyst's initials, and a two digit number to distinguish the polygon from others collected on the same day.

In CO_4_MMZ9, extensive use was made of a riparian map based on air photo interpretation that was nearly complete for this portion of the state. (http://ndis1.nrel.colostate.edu/riparian/riparian.htm) These polygons were considered to be augmentation data as well, but can be differentiated by the initials "CV" and having a system call of either S095, S120, or D04. Many of the original polygons were not used because of difficulty in determining the appropriate system label. Polygons less than 8100 square meters (9 pixels) were also removed to prevent the data set from having too strong of an impact on the classification.

b) Summary of samples:

12,950 samples were available for modeling. Two polygon coverages containing all samples are located at: **CO/archive/CO4_MMZ9/train_data/.**

The two coverages are a pre- and post- "shrunk" coverage, relating to a step necessary for modeling described in section 5.a).

The table below shows the classes that were mapped in this map zone, and the number of sample polygons for each class. These classes make up the hybrid/alliance classification. Simple recoding of the alliances to systems for the systems classification is described in section 5.b).

(Technically, a few more cover types not shown were actually modeled and then aggregated for the final map; this recoding is described in section 5.b and in **CO/ARCHIVE/CO4_MMZ9/POST_MODEL/recodes.doc**; *the numbers below reflect the aggregation into the final classes.)*

The absence of any samples for S034 relates to the fact that this mesic version of S032 was difficult to identify in the field (and from other data sources) and was therefore only mapped with post classification tree modeling as described in section 5.b).

CODE	**# SAMPLES**	**DESCRIPTION**
S006	7	ROCKY MOUNTAIN CLIFF AND CANYON COMPLEX
S008	29	WESTERN GREAT PLAINS CLIFF AND OUTCROP
S014	24	INTER-MOUNTAIN BASINS GREASEWOOD WASH
S023	2	ROCKY MOUNTAIN ASPEN FOREST AND WOODLAND
S032	39	ROCKY MOUNTAINS MONTANE DRY-MESIC MIXED CONIFER FOREST AND WOODLAND
S034	0	ROCKY MOUNTAINS MONTANE MESIC MIXED CONIFER FOREST AND WOODLAND
S036	197	ROCKY MOUNTAINS PONDEROSA PINE WOODLAND
S038	208	ROCKY MOUNTAIN PINYON-JUNIPER WOODLAND
S046	410	ROCKY MOUNTAINS GAMBEL OAK - MIXED MONTANE SHRUBLAND
S047	91	ROCKY MOUNTAINS LOWER MONTANE-FOOTHILL SHRUBLAND
S048	723	WESTERN GREAT PLAINS SAND SAGEBRUSH SHRUBLAND
S065	99	INTER-MOUNTAIN BASINS MIXED SALT DESERT SCRUB
S074	105	SOUTHERN ROCKY MOUNTAIN JUNIPER WOODLAND AND SAVANNA
S079	49	INTER-MOUNTAIN BASINS SEMI-DESERT SHRUB STEPPE
S086	510	ROCKY MOUNTAINS FOOTHILL GRASSLAND
S087	41	CENTRAL MIXED GRASS PRAIRIE
S088	2242	WESTERN GREAT PLAINS SHORTGRASS PRAIRIE
S089	6	WESTERN GREAT PLAINS SAND PRAIRIE
S093	5	ROCKY MOUNTAINS LOWER MONTANE RIPARIAN WOODLAND AND SHRUBLAND COMPLEX
S095	809	WESTERN GREAT PLAINS RIPARIAN WOODLAND AND SHRUBLAND
S096	80	INTER-MOUNTAIN BASINS GREASEWOOD FLAT COMPLEX
S100	25	NORTH AMERICAN ARID WEST EMERGENT MARSH
S120	413	WESTERN GREAT PLAINS FLOODPLAIN
S125	12	ROCKY MOUNTAIN FOOTHILL LIMBER PINE-JUNIPER WOODLAND
S132	5	WESTERN GREAT PLAINS TALLGRASS PAIRIE
N11	444	OPEN WATER
N21	154	DEVELOPED, OPEN SPACE-LOW INTENSITY
N22	94	DEVELOPED, MEDIUM - HIGH INTENSITY
N81	164	PASTURE/HAY
N82	4019	CULTIVATED CROPS AND IRRIGATED AGRICULTURE
D02	20	RECENTLY BURNED
D03	16	RECENTLY MINED OR QUARRIED

	230	INVASIVE SOUTHWEST RIPARIAN WOODLAND AND
D04		SHRUBLAND
D06	287	INVASIVE PERENNIAL GRASSLAND
D08	71	INVASIVE ANNUAL GRASSLAND
D09	115	INVASIVE ANNUAL AND BIENNIAL FORBLAND
D11	5	RECENTLY CHAINED PINYON-JUNIPER AREAS
D15	715	CONSERVATION RESERVE PROGRAM (CRP) LANDS
A1540	166	YUCCA GLAUCA SHRUB HERBACEOUS ALLIANCE

3) Cover types:

a) Classification Tree (CT) modeled cover types:

See above table.

b) Non CT modeled cover types:

NONE. Colorado relied on classification tree (CT) modeling, with post-classification modeling and recoding as necessary, to map all cover types. Some ancillary coverages were used in this process, but they were never directly "burned in".

4) Summary of predictor layers used:

a) Multi band predictors:

NONE. In particular, raw ETM+ bands were not used. The Tasseled Cap Transforms and the ratios captured most of the spectral variability in the imagery in a way that might be more efficiently utilized by the classification tree algorithm, and in a way that might be more meaningfully interpreted. (As a test, we conducted a Kappa evaluation of a map created with raw bands and a map created with the transforms plus ratios; the results were essentially identical.)

b) Single band predictors:

In addition to the scene sequence images, the model used the spring, summer, and fall Tasseled Cap products (Brightness, Greenness, and Wetness), the spring, summer, and fall NDVI and NPV ratio, and elevation. Slope, aspect, and landform were also used in this mapzone, but care had to be taken to avoid any deleterious effects on the modeling. Essentially, the addition of "noisy" predictor variables require the addition of substantial amounts of training data.

All single-band predictors can be found at:
/CO/ARCHIVE/CO4_MMZ9/IMG_FILES/

5) Modeling Methods:

a) See5 Classification Tree modeling:

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.

The pseudo-replication process involved several steps to generate random pixels within each sample polygon. Each sample polygon was first shrunk by 13 meters. This somewhat arbitrary distance was selected so that when random points placed within the shrunken polygon were converted to pixels in a raster grid, a good part of any of the resulting pixels would not lie outside the original sample polygon. Next, 20 random points were placed within each of the shrunken polygons using an **Arcview Avenue script.** Finally, the points were converted to pixels while ensuring that the resulting pixels (the new grid) aligned with the raster predictor layers (failure to ensure this would result in a future resampling that would be more likely to place pixels outside the original sample polygon). The resulting sub-sampled pixels would often be less than 20 per sample polygon, if random points fell within the same pixel. This was far more likely on smaller polygons, such as the buffered Basinwide points, for which there would often only be five or six sub-samples per polygon. This new grid was converted to an Imagine IMG file and is available at:
CO/ARCHIVE/CO4_MMZ9/TRAIN_DATA/rp20sb_g.img

These sub-sampled pixels were then individually drilled through predictor layers using the sampling tool from the CART Module for Imagine (EarthSatellite Corp. 2003). 10% of the ***pixels*** were set aside as "test data" (although these were not considered to be at all independent, and this is not the same as the validation ***polygon*** data set described in Section 6).

See5 Classification Tree:
See5 (Release 1.8) data mining software (www.rulequest.com) was used for generating classification trees. Boosting was employed using 10 trials with no pruning. Boosting involves successive modeling runs on a portion of the data (90% with 10% set aside for evaluation each time, given 10 trials). After the first run, the model tries to focus on the types it erred on in the previous run(s). After 10 runs, voting occurs, whereby the most common classification for a given pixel is assigned to that pixel.

The See5 related files (*.names, *.names.hst, *.data, *.test, *.out, *.tree, and *.set) are located at: **CO/ARCHIVE/CO4_MMZ9/OUTPUT/.**

The *.names file describes the dependent variable and the independent variables. The *.data file contains all of the raw training data for the dataset. The *.test file contains the raw test data. The *.out file describes all the rules – at the very bottom there is information on how the model performed on each of the 10 runs and on the final boost, both against the 90% training data and the 10% "test data". The *.tree is an application file used for applying the final rules.

Spatial application of rules:
The CART Module for Imagine (EarthSatellite Corp. 2003) was used to spatially apply the rules from the .tree file in See5 to create a classified image for the entire mapzone. In the process, an associated "*_error.img" was also created. This latter image is more of a "confidence" map, as higher values indicate greater classification confidence (i.e. less class confusion for similar pixels in the training data). Confidence should not be equated with "accuracy", as the absence of class confusion may simply result from a lack of similar training data, but this map does provide a potentially useful spatial evaluation.

The classified image and the associated "error.img" can be found at **/CO/ARCHIVE/CO4_MMZ9/OUTPUT/.**

b) Post-classification, recoding and other modeling steps:

The output of the C5 classification was evaluated on a class by class basis by isolating individual classes on the map (this can be done in ERDAS IMAGINE by keeping the opacity of the evaluated class at 1 while changing the opacity of the other classes to 0). Anomalous distributions that were not thought to be naturally disjunct outliers were often removed through the generation of "areas of interest" (AOI's) in Imagine, and recoding the problematic type within those areas to a more likely class value. AOI's were named according to the type_to_type conversion and can be found at:
CO/ARCHIVE/CO4_MMZ9/AOI/.
The recoding of certain types within known urban areas was more automated. A Colorado road network coverage draped over Landsat ETM+ imagery was used to guide the digitizing of an urban layer; an AOI of these urban areas can be found at:
CO/ARCHIVE/CO1_CO2_MMZ7/AOI/mmz7_all_urb2ed.aoi. This AOI was used to recode anything that mapped as a naturally barren type within an urban area to N22 (Developed, Medium – High Intensity). An exact description of the conversions within this AOI is given in:
CO/ARCHIVE/CO4_MMZ9/AOI/Urban_recode_explanation.doc
One other recoding step involved converting the classes that were used in the modeling to those that were used in the final maps. (There are different reasons for different classes going into modeling. Some old class labels were grandfathered in to the modeling. In other cases, there was some uncertainty as to which alliance/system a cover type should go into, and class confusion in the final map was a good way of helping that decision.) This recoding is described in:
CO/ARCHIVE/CO4_MMZ9/POST_MODEL/recodes.doc
Finally, a model was used to separate the Mesic from the Dry-Mesic versions of system S032 to create system S034. The .gmd file used for this post-classification model can be found at: **CO/ARCHIVE/CO4_MMZ9/POST_MODEL/mesic_model.gmd**
This model is actually a combination of the two models described below:
Discriminating *S032-Rocky Mountain Montane Dry-Mesic Mixed Conifer Forest and Woodland* and *S034-Rocky Mountain Montane Mesic Mixed Conifer Forest and*

Woodland: A post-classification model was used to discriminate *S032* from *S034*. The logic and parameters for the model were as follows:

This post-classification model was designed to extract *S032-Rocky Mountain Montane Dry-Mesic Mixed Conifer Forest and Woodland* from the standard modeled *S034-Rocky Mountain Montane Mesic Mixed Conifer Forest and Woodland* ecological system cover type. Nature Serve describes this system as *"occurring predominantly in cool ravines and on north-facing slopes. Elevations range from 1200 to 3300 m. Occurrences of this system are found on cooler and more mesic sites than the Rocky Mountain Montane Dry-Mesic Mixed Conifer Forest and Woodland. Such sites include lower and middle slopes of ravines, along stream terraces, moist, concave topographic positions and north and east-facing slopes which burn somewhat infrequently."*

Model methods: A standard model was run withholding *S034*, consequently allowing *S032* to be mapped in its place. A condition statement was created to extract *S034* using the above aspect and landform characteristics in order to identify this montane mesic mixed conifer system. The details of the conditional statement follow:

EITHER *S034* IF ((Landform = = 2 OR Landform = = 5 OR Landform = =6 OR Landform = = 9) AND (Aspect = = 1 OR Aspect = = 2 OR Aspect = = 0) AND (Standard Modeled Vegetation Output = = *S032*)) OR Standard Modeled Vegetation Output OTHERWISE

c) Generalizing to MMU and map completion:

After the spatial application of the CT model to create an *.img file, and post classification steps were taken, the map was generalized using ERDAS Imagine 8.6 GIS Analysis, Clump tool using *4 connected neighbors* (rooks move), and then using Eliminate with a setting of a *minimum of 1 acre*.
Finally, alliances were recoded to systems to create a "Systems only" final map. A description of this simple recoding is found at:
/CO/ARCHIVE/CO4_MMZ9/POST_MODEL/alliance_recodes.doc
Both maps can be found at: **/CO/ARCHIVE/CO4_MMZ9/FINAL_MAP/.**

6) Validation:

a) CT model validation:

Classification assessment was an internal validation in that it used a subset of the reference data collected for modeling purposes. 20% of the reference sample polygons were randomly selected (***using the featureselect.avx***) and set aside for validation. A CT model was then generated with the remaining 80% of sample polygons, using the same procedures described in section 5.a). This model was recoded to final "hybrid" classes as described in section 5.b) and
CO/ARCHIVE/CO4_MMZ9/POST_MODEL/recodes.doc, and from there to final

“system” classes as described in section 5.c) and **/CO/ARCHIVE/CO4_MMZ9/POST_MODEL/alliance_recodes.doc.** Without any post-classification modeling or clump/eliminate, these recoded land cover maps were then assessed with the appropriate recoded version of the 20% reference sample. (The lack of post-processing increased the fidelity between the map and the training data; for example, there was no mesic mixed conifer class in the training data. This was thought to provide a better indication of model performance.)

The CT-modeled classes were assessed using kappa.avx, which intersects the validation sample polygons through the CT modeled land cover map, and considers the site correctly mapped when the most common pixel value within the sample polygon agrees with the sample label. Output from kappa.avx consists of a *.txt, *.dbf, and *.shp file. The *.txt file presents the kappa statistic, the *.dbf file is an error matrix indicating errors of commission and omission and the *.shp file indicates for each reference sample site whether the sample location was considered correctly mapped, or incorrectly mapped and what it was mapped as. The *.dbf files are a bit difficult to read with the default output; they have been modified into cleaner looking matrices and saved as *.xls files. All of these files can be found at: **/CO/ARCHIVE/CO4_MMZ9/VALIDATION/.**

Several (internal) validations were performed. The first Kappa run (and the corresponding .txt file, .xls file, and .shp file) is an evaluation of the hybrid product (generated from the 80% sample) against 20% of all the training data (including augmentation polygons). The second Kappa run is a similar evaluation of the systems product (simply the hybrid with the few alliances recoded to systems). The third run is an evaluation of the systems map against only the portion of the 20% withheld data that was collected in the field (both Basinwide and SWReGAP). The fourth run is an evaluation of the systems map against only the withheld data that was collected by SWReGAP field crews. Finally, the fifth run is an evaluation of the final map (generated using 100% of the training data) against all 100% of the reference data. This last product should not be considered a validation and could be very misleading – but it may demonstrate which classes are simply very difficult for the model to separate.

b) Discussion of mapped cover types:

The following is a somewhat qualitative evaluation of the first Kappa run, reviewing the model’s performance on each of the classes of the hybrid scheme against a 20% sampling of all the training data (including augmented polygons). This evaluation is not much different than it would be for the systems scheme, as there is not much of a difference between the hybrid product and the systems product.

On the other hand, there is a tremendous difference between evaluating with all of the training data and with SWReGAP field-collected polygons only, due to the fact that so few polygons were collected in the field for this mapzone. The lack (and unreliability) of field collected polygons for this mapzone makes evaluation with them similarly unreliable.

(It should be noted that a couple extra classes that are not in the target legend are included in the evalution below: *D18* – stockyards, and *N13* – water for one or two dates of imagery. These classes were later grouped with *N21* and *N11*, respectively, for the hybrid level product, but were included in this Kappa assessment out of curiosity.)

The producer's and user's accuracies described below relate to omission and commission errors, respectively, for the different classes of the hybrid (alliance/systems) level map. [Actual tables are found under **ARCHIVE/CO4/VALIDATION** at **errmtx1.dbf** and **matrix1_adj.xls**.] In some cases a given class was not mapped in any of the training sites; while there were no errors of commission, the class was not mapped correctly anywhere either, and so the resulting 0/0 has no user's accuracy.

A1540 (Yucca): Fairly low producer's accuracy (32%), often omitted in favor of sand sagebrush, shortgrass prairie and foothill grassland. Its user's accuracy was much better (58%), suggesting that at least it's not grossly overmapping. [**34 sites**]

D02 (Recently Burned): 50% producer's (accuracy) and 100% user's (accuracy) suggest that burned areas are not being overmapped, but some may be missed (which is probably fine if the cover type is correct!). [**4 sites**]

D03 (Mining): 50% producer's and 67% user's suggest that mines aren't overmapped badly but they are getting missed. [**4 sites**]

D04 (Invasive Riparian Woodland): 80% producer's and 64% user's suggest that it's not missed very often (a little bit is called plains riparian woodland), but that it is mapped in place of plains riparian woodland fairly often. There is very little other confusion. [**46 sites**]

D06 (Invasive Perennial Grassland): 47% producer's and 56% user's suggest that there is a moderate amount of omission and commission related to confusion with pasture, shortgrass prairie, and foothill grassland. [**58 sites**]

D08 (Invasive Annual Grassland): 20% producer's and 50% user's suggest that it's vastly undermapped (primarily in favor of other grasslands), but at least it's not overmapping. [**15 sites**]

D09 (Invasive Annual and Biennial Forbland): 13% producer's and 43% suggest that it's strongly omitted to shortgrass prairie, but at least not overmapping extensively. Some forbs (e.g. Kochia) might have been more mappable than others (e.g. clover, which might be less distinguishable from shortgrass sites). [**23 sites**]

D11 (Chained PJ): The one site was missed (0% producer's accuracy, 0/0 user's) but there were really too few samples to evaluate fairly. [**1 site**]

D15 (CRP): 80% producer's and 77% user's suggest that it's mapped pretty well, but a little confusion exists with shortgrass prairie, ag, and sand sagebrush (in that order).

[**143 sites**]

D18 (Stockyards): 100% producer's and 60% user's suggest that these are very mappable. Merged with N21 in the final map. [**3 sites**]

N11 (Water): 94% producer's and 93% user's. Only substantial confusion with the following water class. [**89 sites**]

N13 (One or two-date water): 81% producer's and 90% user's demonstrates that this water class is a little tougher, being the marginal class that it is. [**58 sites**]

N21 (Low density development): 90% producer's and 88% user's indicate that this class is mapped quite well. Some confusion with plains riparian woodland is evident on the map, probably due to similar deciduous canopies, but also possibly due to an excess of riparian polygons. [**31 sites**]

N22 (High density development): 100% producer's and 83% user's suggest that it is rarely missed but can sometimes overmap (mines and lake shorelines in particular). [**19 sites**]

N81 (Pasture): 21% producer's and 50% user's indicates that this class is often mapped as various grasslands or agriculture, but does not overmap extensively. This class suffers from a definitional problem in the plains, where almost everything (or almost nothing) could be considered pasture. It's very difficult to be consistent with this label on the plains. [**33 sites**]

N82 (Agriculture): 95% producer's and 92% user's indicate that it's mapped well, but that there is some confusion with herbaceous riparian, shortgrass prairie, CRP land, and sand sagebrush, in that order. [**804 sites**]

S006 (Rocky Mountain Cliff and Canyon): 100% producer's and 100% user's suggest that it's mapped well, but there are too few sites to evaluate realistically. [**2 sites**]

S008 (Great Plains Cliff and Outcrops): 33% producer's and 100% user's suggest that it's often omitted to shortgrass prairie, but that it's not overmapping. [**6 sites**]

S014 (Wash): 40% producer's and 100% user's suggest that it may be omitted in favor of agriculture (fallow lands are similarly bare), but that it's not overmapping. [**5 sites**]

S023 (Aspen): The one site was missed (0% producers, 0/0 user's), but that's just not enough to evaluate fairly. [**1 site**].

S032 (Mixed Conifer): 75% producer's and 86% user's suggest that there's only some confusion with Ponderosa Pine. [**8 sites**]

S036 (Ponderosa): 65% producer's and 77% user's indicate that there's some confusion with Pinyon Pine in particular. (This may be exaggerated by the fact that many polygons were placed in the area of confusion in an effort to resolve the problem. Furthermore, these reference data are by no means "ground truth".) [**40 sites**]

S038 (Pinyon-Juniper): 62% producer's and 70% user's relate to the confusion with both Ponderosa Pine and Juniper savanna, in addition to the potentially misleading factors mentioned for Ponderosa above. [**42 sites**]

S046 (Gambel oak): 89% producer's and 87% user's indicate that Gambel oak is mapped very well, with the only notable confusion being with the Plains Riparian Woodland. [**82 sites**]

S047 (Foothill Shrubland): 79% producer's and 75% user's indicate that these shrublands are a little more likely to be confused with some grasslands, as well as with Pinyon-Juniper woodlands. [**19 sites**]

S048 (Sand Sagebrush): 71% producer's and 74% user's indicate that these shrublands are mapped reasonably well, but that there is still fairly substantial confusion with shortgrass prairie, and with CRP and ag to a lesser extent. [**145 sites**]

S065 (Salt Desert Scrub): 30% producer's and 60% user's relate to omission to shortgrass prairie, but not overmapping extensively. Many of the omitted Atriplex stands may actually contain shortgrass, such that salt desert may not be the most appropriate term anyway. [**20 sites**]

S074 (Juniper Savanna): 81% producer's and 68% user's indicate that some Juniper savanna is being mapped as shortgrass prairie and a fair amount of P-J woodlands are being mapped as Juniper savanna. [**21 sites**]

S079 (Semi-Desert Shrub Steppe): 10% producer's and 50% user's suggest that this class is being omitted to shortgrass prairie and foothill grassland, but that at least the class is not extensively overmapping. [**10 sites**]

S086 (Foothill Grassland): 73% producer's and 63% user's indicate that there is some confusion with shortgrass prairie, and, secondarily, with invasive perennial grassland. [**102 sites**]

S087 (Mixed Grass Prairie): 22% producer's and 67% user's indicate that this class is being omitted to shortgrass prairie and sand sagebrush, but that at least it's not overmapping extensively. [**9 sites**]

S088 (Short Grass Prairie): 86% producer's and 71% user's indicate that the class is mapped reasonably well, but generally overmapped rather than undermapped. It is occassionally omitted in favor of sand sagebrush, foothill grassland, agriculture, CRP, or invasive perennial grass, but much more commonly, it is mapped in the place of these classes, in addition to many other classes in moderate amounts. [**449 sites**]

S089 (Sand Prairie): 50% producer's and 100% user's are encouraging, but too few sample sites to be conclusive. [**2 sites**]

S093 (Foothill Riparian Woodland): This class was primarily obtained through post-processing of S095, so the 0% producer's (0/0 user's) on one site is certainly not very representative. [**1 site**]

S095 (Plains Riparian Woodland): 72% producer's and 79% user's relate to this class sometimes being omitted in favor of invasive riparian woodland and herbaceous riparian, and, to a lesser extent, Gambel oak and agriculture. Commission errors are generally with the same classes but to a lesser degree, except for herbaceous riparian, which it often replaces. [**158 sites**]

S096 (Greasewood): 56% producer's and 60% user's relate to occasional confusion with such classes as shortgrass prairie. [**16 sites**]

S100 (Marsh): 0% producer's and 0% user's (1 error of commission) suggest that this class is not being mapped very well. It's generally called ag or plains riparian woodland. It could probably use more training sites. [**5 sites**]

S120 (Herbaceous Riparian): 54% producer's and 61% user's indicate that is it fairly often omitted to wooded riparian or agriculture. User's errors occur with the same classes to a slightly lesser extent. [**82 sites**]

S125 (Limber Pine): 67% producer's and 100% user's indicate that this class is being mapped fairly well (confirmed by the map), although there are too few sample sites to be very precise. [**3 sites**]

S132 (Tallgrass Prairie): 0% producer's (0/0 user's) on one site not conclusive. [**1 site**]

7) Citations:

Chavez, P. S. Jr., 1988, An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ.*, 14, pp.459-479.
Chavez, P. S. Jr., 1996, Image-based atmospheric corrections—revisited and revised. *Photogrammetric Engineering and Remote Sensing* 62(9): 1025-1036.

Comer, P., D. Faber-Langendoen, R. Evans, S. Gawler, C. Josse, G. Kittel, S. Mernard, M. Pyne, M. Ried, 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.

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

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.