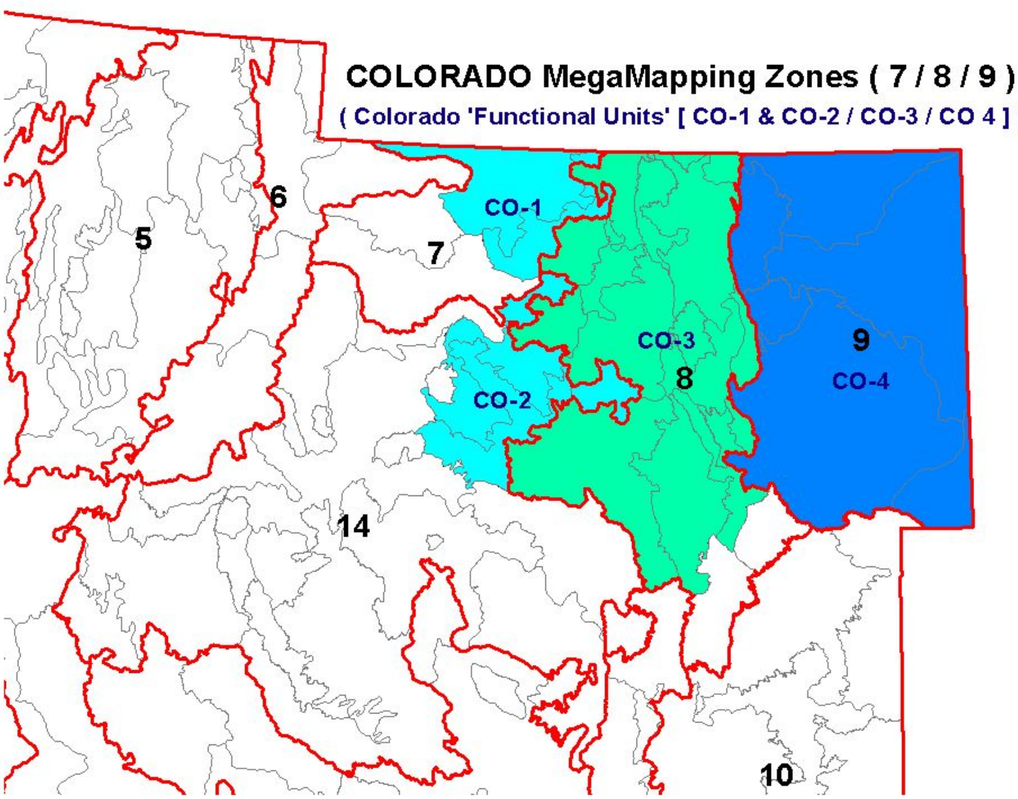
SWReGAP Land Cover Mapping Methods Documentation – "Colorado"

I. Introduction

The “Colorado” component of the land cover mapping for the Southwest Regional Gap Analysis Project (SWReGAP) was conducted at the Colorado Division of Wildlife headquarters in Denver, Colorado, in a collaborative partnership with the United States Bureau of Land Management and the Natural Resource Ecology Laboratory (NREL) at Colorado State University in Fort Collins, Colorado. Colorado was placed in quotation marks above to emphasize the fact that mapping responsibility in the SWReGAP effort did not directly follow state lines, but followed the somewhat ecoregional mapzone boundaries that were developed for this project. As a result, “Colorado” mapping went into Utah and New Mexico in some places, and Utah and New Mexico mapped into Colorado in others. Mega-mapzones (an aggregation of the original smaller mapzones) were partitioned by some of the original mapzone boundaries to create “functional units” for land cover modeling that more closely corresponded to state boundaries. As a result, a piece of northwestern Colorado became Colorado functional unit 1 (CO-1), a portion of southwestern Colorado became Colorado functional unit 2 (CO-2), the core Rocky Mountains of Colorado became Colorado functional unit 3 (CO-3) [our “MMZ8”], and the plains of eastern Colorado became Colorado functional unit 4 (CO-4) [MMZ9]. Colorado functional units 1 and 2 were combined into one “west slope” unit (CO1_CO2) [our “MMZ7”] for convenience in modeling. The figure below illustrates the above:



The methods used for mapping each functional unit are summarized below. These summaries can also be used as a guide to the respective archives. In general, the methodology among the functional units is very similar. While land cover classes differed to some extent among functional units, a bigger difference among functional units relates to the reference data that were obtained. SWReGAP field crews collected substantial field data in functional units CO1_CO2 [MMZ7] and CO-3 [MMZ8]. CO-4 [MMZ9], on the other hand, was minimally covered by SWReGAP field crews, and therefore depended more on additional sources. In addition to some ancillary point data from the Colorado Vegetation Classification Project (often referred to as "Basinwide" data and also used extensively in the other functional units), substantial amounts of polygons from a riparian data set were incorporated for CO-4 [MMZ9]. Additionally, while "augmentation data" (polygons derived from image interpretation) were used fairly heavily in all of the functional units, they were indispensable in CO-4 [MMZ9], and even more extensively used. The accuracy of the land cover classification in CO-4 [MMZ9] does not appear to have suffered as a result of the extensive augmentation, but the dearth of field data does make the "validation" in this functional unit less conclusive.

It should be mentioned that field data collection strategies did vary over time, and therefore among mapzones to some extent. Early field data collection involved unsupervised classification of satellite imagery, with a random sampling of the resulting polygons that were within a specified distance to a road according to a Colorado road coverage. (An ArcView Avenue script provided by Lee O'Brien allowed for this selection process.) Field crews then targeted these sample sites. This strategy was employed for the first couple of field seasons in CO-4 [MMZ9], and halfway into the third year (most of CO-2; in MMZ7). It was then replaced by a strategy whereby image interpreters manually digitized polygons of interest on the imagery along the road network and field crews tracked down the targeted polygons. The field crews could also collect "opportunistic" polygons along the way if there were any other land cover types of interest. This latter approach was used for the last year and a half (primarily CO-1 [in MMZ7] and CO-3 [MMZ8]) and therefore constitutes the bulk of the field polygons (because most of the data were collected in the last two years – 2002 and 2003, and because this approach generally allowed for the collection of much more data).

The functional units were merged using ERDAS Imagine Areas of Interest (AOI's) as "cutlines". These AOI's can be found under CO/PRODUCT/AOI. These cutlines were digitized to ensure the best possible edge-matching between functional units, so that there would be no obvious boundary discontinuities. (Unfortunately, there seems to be a limit on the number of vertices that can be used in an AOI; the "cutlineb" AOI is related to the fact that a second cutline had to be generated after an earlier version lost its finer resolution due to Imagine's limitations.) Colorado used a "matrix" overlay approach to facilitate the edge-matching. This approach involved creating a two-layer image in the areas of overlap between two functional units, and color-coding the areas of agreement according to their class label. Areas of disagreement were left black. A cutline digitized through the areas of agreement ensured that there would be no obvious discontinuity between functional units. This strategy was most effective when there was strong agreement between the two maps, as was the case for the most part.

II. Functional Unit CO1_CO2 (MMZ7) Mapping Methodology

Functional Unit or Mapping Zone:
CO-1 & CO-2 (For convenience, Colorado's relatively small portions of two "Mega-mapzones" were merged into one "Colorado West Slope" mapzone.) Occasionally referred to as our Mega-mapzone 7 (MMZ7).

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:

ETM+ Scene (path/row)	Spring (yr-Julian date)	Summer (yr-Julian date)	Fall (yr-Julian date
34/33	2001-107	1999-182	1999-310
34/34	2001-107	2000-201	1999-310
35/32	2001-162	2000-256	1999-317
35/33	2000-144	2000-256	1999-285
35/34	2000-144	2000-256	1999-285
36/32	2000-103	2000-247	1999-292
36/33	2000-103	2000-167	1999-308
36/34	2000-103	2000-167	1999-292
37/32	2001-128	2000-158	1999-283

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/CO1_CO2_MMZ7/MOSAICS/mz7spri_scene_seq.img
/CO/ARCHIVE/CO1_CO2_MMZ7/MOSAICS/mz7summ_scene_seq.img
/CO/ARCHIVE/CO1_CO2_MMZ7/MOSAICS/mz7fall_scene_seq.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 resulting six band Landsat ETM+ mosaics for each season can be found at:

/CO/ARCHIVE/CO1_CO2_MMZ7/MOSAICS/mz7buff_spr_aea.img
/CO/ARCHIVE/CO1_CO2_MMZ7/MOSAICS/mz7buff_sum_aea.img
/CO/ARCHIVE/CO1_CO2_MMZ7/MOSAICS/mz7buff_fall_aea.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/CO1_CO2_MMZ7/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/CO1_CO2_MMZ7/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/CO1_CO2_MMZ7/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/CO1_CO2_MMZ7/IMG_FILES/MODELS/npv_ratio_c.gmd**

All image derived predictor layers can be found at: **/CO/ARCHIVE/CO1_CO2_MMZ7/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: Expressed in degrees, as calculated by ERDAS Imagine's slope routine.

Aspect: A nine class aspect grid. 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 derived 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/CO1_CO2_MMZ7/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.

Some use was made of the U.S. Forest Service's "R2VEG" vegetation coverage (e.g.http://www.fs.fed.us/r2/sanjuan/projects/gis/index.shtml) to guide the augmentation of polygons. Some consideration was given to "burning in" these forest service polygons directly, but the guided augmentation approach permitted the better modeling of areas lacking R2VEG coverage, and provided a more independent mapping that was consistent with the rest of our effort; a separate assessment could be conducted to determine whether a direct "burn in" would have been more accurate.

b) Summary of samples:

7,180 samples were available for modeling. Two polygon coverages containing all samples are located at: **CO/archive/CO1_CO2/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 hybrid map; this recoding is described in section 5.b and in **CO/ARCHIVE/CO1_CO2/POST_MODEL/recodes.doc**; *the numbers below reflect the aggregation into the final classes.)*

The absence of any samples for S030 and S034 relates to the fact that these mesic versions of S028 and S032, respectively, were difficult to identify in the field (and from other data sources) and were therefore only mapped with post classification tree modeling as described in section 5.b).

CODE	**# SAMPLES**	**DESCRIPTION**
S002	6	ROCKY MOUNTAIN ALPINE BEDROCK AND SCREE
S006	44	ROCKY MOUNTAIN CLIFF AND CANYON COMPLEX
S009	7	INTER-MOUNTAIN BASINS CLIFF AND CANYON COMPLEX
S010	74	COLORADO PLATEAU MIXED BEDROCK CANYON AND TABLELAND
S011	64	INTER-MOUNTAIN BASINS SHALE BADLANDS
S015	4	INTER-MOUNTAIN BASINS PLAYA
S023	292	ROCKY MOUNTAIN ASPEN FOREST AND WOODLAND
S025	6	ROCKY MOUNTAIN SUBALPINE-MONTANE LIMBER-BRISTLECONE PINE WOODLAND
S028	65	ROCKY MOUNTAINS SUBALPINE DRY-MESIC SPRUCE-FIR FOREST AND WOODLAND
S030	0	ROCKY MOUNTAINS SUBALPINE MESIC SPRUCE-FIR FOREST AND WOODLAND
S031	44	ROCKY MOUNTAINS LODGEPOLE PINE FOREST
S032	129	ROCKY MOUNTAINS MONTANE DRY-MESIC MIXED CONIFER FOREST AND WOODLAND
S034	0	ROCKY MOUNTAINS MONTANE MESIC MIXED CONIFER FOREST AND WOODLAND
S036	186	ROCKY MOUNTAINS PONDEROSA PINE WOODLAND
S039	873	COLORADO PLATEAU PINYON-JUNIPER WOODLAND
S042	90	INTER-MOUNTAIN WEST ASPEN-MIXED CONIFER FOREST AND WOODLAND COMPLEX
S045	154	INTER-MOUNTAIN BASINS MAT SALTBUSH SHRUBLAND
S046	467	ROCKY MOUNTAINS GAMBEL OAK - MIXED MONTANE SHRUBLAND
S047	75	ROCKY MOUNTAINS LOWER MONTANE-FOOTHILL SHRUBLAND
S050	9	INTER-MOUNTAIN BASINS MOUNTAIN MAHOGANY WOODLAND AND SHRUBLAND
S052	101	COLORADO PLATEAU PINYON-JUNIPER SHRUBLAND
S054	1059	INTER-MOUNTAIN BASINS BIG SAGEBRUSH SHRUBLAND

S056	16	COLORADO PLATEAU MIXED LOW SAGEBRUSH SHRUBLAND
S059	19	COLORADO PLATEAU BLACKBRUSH-MORMON TEA SHRUBLAND
S065	192	INTER-MOUNTAIN BASINS MIXED SALT DESERT SCRUB
S071	535	INTER-MOUNTAIN BASINS MONTANE SAGEBRUSH STEPPE
S075	95	INTER-MOUNTAIN BASINS JUNIPER SAVANNA
S079	110	INTER-MOUNTAIN BASINS SEMI-DESERT SHRUB STEPPE
S083	69	ROCKY MOUNTAIN SUBALPINE MESIC MEADOW
S085	100	SOUTHERN ROCKY MOUNTAINS MONTANE GRASSLAND
S090	175	INTER-MOUNTAIN BASINS SEMI-DESERT GRASSLAND
S091	22	ROCKY MOUNTAINS SUBALPINE/MONTANE RIPARIAN SHRUBLAND
S092	1	ROCKY MOUNTAINS SUBALPINE/MONTANE RIPARIAN WOODLAND
S093	153	ROCKY MOUNTAINS LOWER MONTANE RIPARIAN WOODLAND AND SHRUBLAND COMPLEX
S096	190	INTER-MOUNTAIN BASINS GREASEWOOD FLAT COMPLEX
S100	23	NORTH AMERICAN ARID WEST EMERGENT MARSH
S102	19	ROCKY MOUNTAINS ALPINE/MONTANE WET MEADOW
S128	25	WYOMING BASINS LOW SAGEBRUSH SHRUBLAND
N11	69	OPEN WATER
N21	47	DEVELOPED, OPEN SPACE-LOW INTENSITY
N22	36	DEVELOPED, MEDIUM - HIGH INTENSITY
N81	598	PASTURE/HAY
N82	296	CULTIVATED CROPS AND IRRIGATED AGRICULTURE
D01	1	NON-SPECIFIC DISTURBED
D02	36	RECENTLY BURNED
D03	68	RECENTLY MINED OR QUARRIED
D04	27	INVASIVE SOUTHWEST RIPARIAN WOODLAND AND SHRUBLAND
D06	62	INVASIVE PERENNIAL GRASSLAND
D08	104	INVASIVE ANNUAL GRASSLAND
D09	47	INVASIVE ANNUAL AND BIENNIAL FORBLAND
D10	18	RECENTLY LOGGED AREAS
D11	40	RECENTLY CHAINED PINYON-JUNIPER AREAS
D15	43	CONSERVATION RESERVE PROGRAM (CRP) LANDS
A788	13	ARCTOSTAPHYLOS PATULA SHRUBLAND ALLIANCE
A913	34	AMELANCHIER ALNIFOLIA SHRUBLAND ALLIANCE
A916	83	AMELANCHIER UTAHENSIS SHRUBLAND ALLIANCE
A1523	65	PURSHIA TRIDENTATA 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 not used, as it was determined that these often had deleterious effects on the modeling.

All single-band predictors can be found at:
/CO/ARCHIVE/CO1_CO2_MMZ7/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/CO1_CO2_MMZ7/TRAIN_DATA/rp4325g.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 then 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/CO1_CO2_MMZ7/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/CO1_CO2_MMZ7/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/CO1_CO2_MMZ7/AOI/.

Some of the recoding was more automated, such as the recoding of certain types within known urban areas. 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). Similarly, any non-forested cover type that fell within a logged AOI (**CO/ARCHIVE/CO1_CO2_MMZ7/AOI/mmz7_gmug_rg_rt_sj.aoi**) was recoded to D10 (Recently Logged). This AOI was derived from the U.S. Forest Service R2VEG coverage. An exact description of the conversions within these AOI’s is given in:**CO/ARCHIVE/CO1_CO2_MMZ7/AOI/MMZ7_AOI_explanations.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/CO1_CO2_MMZ7/POST_MODEL/recodes.doc

Finally, a model was used to separate the Mesic from the Dry-Mesic versions of systems S028 and S032 to create systems S030 and S034, respectively. The .gmd file used for this post-classification model can be found at:
CO/ARCHIVE/CO1_CO2_MMZ7/POST_MODEL/mesic_conifer_classes.gmd
This model is actually a combination of the two models described below:

Discriminating *S028-Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland* and *S030-Rocky Mountain Subalpine Mesic Spruce-Fir Forest and Woodland*: A post-classification model was used to discriminate *S028* from *S030*. The logic and parameters for the model were as follows:

This post-classification model was designed to extract *S030-Rocky Mountain Subalpine Mesic Spruce-Fir Forest and Woodland* from the standard modeled *S028-Rocky*

Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland. Nature Serve describes *S030* to be *"typically found in locations with cold air drainage or ponding, or where snowpacks linger late into the summer, such as north-facing slopes and high elevation ravines. They can extend down in elevation below the subalpine zone in places where cold air ponding occurs; northerly and easterly aspects predominate. These forests are found on gentle to very steep mountain slopes, high elevation ridgetops and upper slopes, plateaulike surfaces, basins, alluvial terraces, well-drained benches, and inactive stream terraces."*

Model methods: A standard model was run withholding *S030*, consequently allowing *S028* to be mapped in its place. A condition statement was created to extract *S030* using the above aspect and landform characteristics in order to identify this subalpine mesic conifer system. The details of the conditional statement follow:

EITHER *S030* 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* = = *S028*)) OR *Standard Modeled Vegetation Output* OTHERWISE

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/CO1_CO2_MMZ7/POST_MODEL/alliance_recodes.doc
Both maps can be found at: **/CO/ARCHIVE/CO1_CO2_MMZ7/FINAL_MAP/.**

<u>6) Validation:</u>

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/CO1_CO2_MMZ7/POST_MODEL/recodes.doc, and from there to final "system" classes as described in section 5.c) and /CO/ARCHIVE/CO1_CO2_MMZ7/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 were no mesic conifer classes 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/CO1_CO2_MMZ7/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. Neither is there a tremendous difference between evaluating with all of the training data or with SWReGAP field-collected polygons only.

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/CO1_CO2/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.

A1523 (Bitterbrush): 62% producer's and 62% user's suggest that this alliance is being mapped reasonably well, except for some omission to big sagebrush (and, to a lesser extent, commission of semi-desert shrub steppe and semi-desert grassland). [**13 sites**]

A788 (Manzanita): 67% producer's and 100% user's suggest that manzanita may occasionally be replaced by Ponderosa, but that it's not overmapping extensively. There were too few samples to get a solid understanding. [**3 sites**]

A913 (Saskatoon Serviceberry): 29% producer's and 100% user's suggest strong omission to montane sagebrush (likely confusion with lush snowberry), but not overmapping. Again, there were too few sites to get a good picture. [**7 sites**]

A916 (Utah Serviceberry): 18% producer's and 60% user's suggest that it is strongly omitted in favor of montane sagebrush, Gambel oak, and foothill shrubland, in that order. There is also a little commission of montane sagebrush, but it does not appear to overmap extensively. [**17 sites**]

D02 (Recent Burn): 100 producer's and 89% user's suggest that burns are mapped well. The one error is likely anomalous. [**8 sites**]

D03 (Mining): 79% producer's and 65% user's indicate that there is some omission to shale badlands, but more of a concern with commission of urban, various naturally barren types, and the salt desert shrublands. The confusion with urban was most likely resolved with post-processing. [**14 sites**]

D04 (Invasive Riparian Woodland): 33% producer's and 100% user's suggest some omission to water, native riparian woodland, and greasewood, but not overmapping. It was not an extensive sampling, however. [**6 sites**]

D06 (Invasive Perennial Grassland): 23% producer's and 100% user's suggest strong omission to big sagebrush (they may be coincident) and some omission to pasture (which may be the same cover type), but not overmapping. [**13 sites**]

D08 (Invasive Annual Grassland): 29% producer's and 75% user's suggest that it is also strongly omitted to big sagebrush (again, they may be coincident) and some omission to pasture (which, again, may be the same cover type), but also not overmapping extensively. [**21 sites**]

D09 (Invasive Annual and Biennial Forbland): 30% producer's and 75% user's demonstrate that it is also often omitted to such types as semi-desert shrub steppe and pasture, but that it is not being overmapped extensively. [**10 sites**]

D10 (Recently Logged): 100% producer's and 100% user's suggest that this type is being mapped very well, but there are too few sites to make any strong conclusions. [**4 sites**]

D11 (Chained Pinyon-Juniper): 63% producer's and 71% user's suggest that there is a fair amount of omission to Pinyon-Juniper woodlands and some commission of montane sagebrush. [**8 sites**]

D15 (CRP): 44% producer's and 44% user's indicates fairly strong omission to big sagebrush, as well as fairly strong commission of pasture and big sagebrush. [**9 sites**]

N11 (Water): 100% producer's and 93% user's indicates that water is mapped well. [**14 sites**]

N21 (Low Intensity Development): 70% producer's and 78% user's indicates some omission to Pinyon-Juniper woodland (which is very surprising) and a little bit of commission of High Intensity Development and Pasture. [**10 sites**]

N22 (High Intensity Development): 50% producer's and 80% user's suggest a fair amount of omission to mining and naturally barren types. These were probably handled with post-processing. The one commission error of salt desert shrubland was likely anomalous and probably handled with post-processing. [**8 sites**]

N81 (Pasture): 69% producer's and 69% user's suggests some omission to big sagebrush, agriculture, montane sagebrush, and riparian woodland (in that order), as well as some commission of montane sagebrush, Gambel oak, agriculture, and big sagebrush (in that order). [**120 sites**]

N82 (Agriculture): 78% producer's and 80% user's indicates that agriculture was only somewhat confused with pasture, and, to a much lesser extent, big sagebrush. [**60 sites**]

S002 (Alpine Bedrock and Scree): 100% producer's and 100% user's are encouraging, but this type was very rare in the area, and there weren't enough samples to adequately evaluate. [**2 sites**]

S006 (Rocky Mountain Cliff and Canyon): 67% producer's and 55% user's indicate various minor problems with this class which generally occurs in areas that are tough to map. The most notable problem is the commission of both mountain and big sagebrush. [**9 sites**]

S009 (Inter-Mountain Basins Cliff and Canyon): 0% producers's (0/0 user's) indicates that these sites were not mapped and were confused with the saltbushes. This class label was rarely used and may not have even been appropriate for Colorado. [**2 sites**]

S010 (Colorado Plateau Canyon and Tableland): 60% producer's and 60% user's indicates that this variable class was confused with a variety of types (most often salt desert scrub). [**15 sites**]

S011 (Inter-Mountain Basins Shale Badlands): 8% producer's and 17% user's indicates that this class was often mapped as mat saltbush, other desert shrublands, and mining, but was also mapped instead of mat saltbush and mining. [**13 sites**]

S015 (Inter-Mountain Basins Playa): 0% producer's and 0% user's (2 errors of commission) indicates that this class was both omitted to mat saltbush and committed to mat saltbush. The actual map was probably much more accurate than this, as the type is probably limited to the few areas where there were training sites. [**1 site**]

S023 (Aspen): 81% producer's and 79% user's suggests that aspen was occasionally omitted to montane sagebrush, lower montane riparian woodland, and Gambel oak, and was occasionally mapped instead of Gambel oak and lower montane riparian woodland. [**59 sites**]

S025 (Subalpine Limber-Bristlecone Pine): 0% producer's (0/0 user's) suggests that this class might not have mapped very well but that at least it probably didn't overmap extensively. More sample sites would be needed to better evaluate the accuracy. [**2 sites**]

S028 (Spruce-Fir): 85% producer's and 73% user's suggests that spruce-fir was only rarely omitted to Lodgepole Pine and Mixed Conifer. More commonly, it was mapped instead of the Aspen-Mixed Conifer (although this could be due, in part, to an overly generous labeling of training sites as Aspen-Mixed Conifer). [**13 sites**]

S031 (Lodgepole Pine): 89% producer's and 67% user's suggests only minimal omission to Aspen-Mixed Conifer, but more common commission of mixed conifer and spruce-fir. [**9 sites**]

S032 (Mixed Conifer): 62% producer's and 67% user's demonstrates that this class was fairly often replaced by Aspen-Mixed Conifer, and occasionally by Lodgepole Pine or Ponderosa Pine. On the other hand, it also committed a fair amount of Ponderosa Pine and even some Gambel oak – one would presume the latter on steep north facing slopes. [**26 sites**]

S036 (Ponderosa Pine): 61% producer's and 63% user's indicates that Ponderosa was often omitted to Pinyon-Juniper and occasionally omitted to mixed conifer and montane sagebrush. On the other hand, it was often mapped instead of Pinyon-Juniper, and occasionally mapped instead of Mixed Conifer and Limber-Bristlecone Pine. [**36 sites**]

S039 (Colorado Plateau Pinyon-Juniper Woodland): 89% producer's and 77% user's indicates that P-J was occasionally replaced by Ponderosa or P-J shrubland, but more commonly it was mapped instead of Ponderosa or Juniper savanna. The latter problem was handled with geographic post-processing. [**175 sites**]

S042 (Aspen-Mixed Conifer): 61% producer's and 65% user's suggest that it is fairly often replaced by spruce-fir and it fairly often replaces Mixed Conifer. This may be more of an issue of training data labeling than mappability. [**18 sites**]

S045 (Mat Saltbush Shrubland): 48% producer's and 50% user's suggests that this type is often confused with Big Sagebrush, Mixed Salt Desert Scrub, and some naturally barren types. [**31 sites**]

S046 (Gambel oak): 77% producer's and 75% user's suggest that it is occasionally replaced by montane sagebrush, and, less commonly, by aspen and (surprisingly) pasture. On the other hand, it occasionally replaces Utah Serviceberry, big sagebrush, and montane sagebrush. [**94 sites**]

S047 (Foothill Shrubland): 33% producer's and 50% user's indicates that it is often replaced by P-J woodland and occasionally replaced by big sagebrush and montane sagebrush. It occasionally replaces Utah Serviceberry and a few other types. [**15 sites**]

S050 (Great Basin Mountain Mahogany Woodland): 0% producer's (0/0 user's) indicates that it might not be getting mapped very well, but that it's probably not overmapping extensively. More sample sites are needed. [**2 sites**]

S052 (Pinyon-Juniper Shrubland): 71% producer's and 56% user's indicates that it is occasionally replaced by P-J woodland and big sagebrush shrubland, but that, more commonly, it replaces P-J woodland. [**21 sites**]

S054 (Big Sagebrush Shrubland): 84% producer's and 59% user's indicate that big sagebrush is only occasionally replaced by montane sagebrush and a smattering of classes, but that, more frequently, it dominates the classification and substantially replaces other classes. Replacement of semi-desert grassland is the most striking (one might wonder the extent to which these were true "grasslands" in the first place), but montane sagebrush, salt desert scrub, pasture, semi-desert shrub steppe, greasewood and a several other classes are often mapped as basin big sagebrush as well. [**212 sites**]

S056 (Colorado Plateau Mixed Low Sagebrush): 0% producer's and 0% user's (1 error of commission) indicates that it's being replaced by big sagebrush and some other types. Alternatively, it doesn't appear to be overmapping extensively. More training data are needed. [**4 sites**]

S059 (Blackbrush-Mormon Tea Shrubland): 100% producer's and 67% user's are encouraging (just a little commission of semi-desert shrub steppe), but there aren't enough sample sites to generate extremely high confidence. [**4 sites**]

S065 (Mixed Salt Desert Scrub): 41% producer's and 46% user's suggest that this type is often replaced by big sagebrush and is occasionally replaced by mat saltbush. On the other hand, it often replaces mat saltbush.[**39 sites**]

S071 (Montane Sagebrush Steppe): 67% producer's and 60% user's indicate that it is fairly often replaced by big sagebrush and pasture, but that more often it replaces other types, including Gambel oak, big sagebrush, the serviceberries, and pasture.[**107 sites**]

S075 (Juniper Savanna): 63% producer's and 80% user's suggest that it was fairly often mapped as P-J woodland and occasionally replaced P-J woodland (both problems were fixed by geographic post-processing). [**19 sites**]

S079 (Semi-Desert Shrub Steppe): 5% producer's and 14% user's confirm that this class is almost always replaced by big sagebrush and a wide range of other classes. Unfortunately, it also maps in some places where it does not belong. [**22 sites**]

S083 (Subalpine Meadow): 64% producer's and 56% user's indicates that meadows are occasionally mapped as something else such as montane sagebrush, and that they are often mapped instead of montane grasslands. [**14 sites**]

S085 (Montane Grasslands): 50% producer's and 71% user's indicates that montane grasslands are often called subalpine meadow and occasionally called montane sagebrush, and that they occasionally replace P-J woodland (surprisingly) and pasture. [**20 sites**]

S090 (Semi-Desert Grassland): 14% producer's and 42% user's indicates that these grasslands are strongly replaced by big sagebrush shrubland, and occasionally replace semi-desert shrub steppe and mixed salt desert scrub. [**35 sites**]

S091 (Subalpine/Montane Riparian Shrubland): 0% producer's (0/0 user's) indicates that this class is being replaced by montane sagebrush, Gambel oak, and Aspen, but that at least it is not overmapping extensively. This class needs more sample sites. [**5 sites**]

S093 (Lower Montane Riparian Woodland): 60% producer's and 64% user's indicates that it is often replaced by Aspen, and occasionally by Gambel oak and pasture. Alternatively, it often replaces marshlands and occasionally replaces aspen and pasture. [**30 sites**]

S096 (Greasewood): 55% producer's and 68% user's indicates that it is often replaced by big sagebrush but also occasionally replaces big sagebrush. [**38 sites**]

S100 (Emergent Marsh): 0% producer's and 0% user's (1 error of commission) indicates that this class is generally replaced by Lower Montane Riparian Woodland, but can also map in place of the same class. This class needs more sample sites. [**4 sites**]

S102 (Alpine/Montane Wet Meadow): 25% producer's and 33% user's indicate that it's generally mapped as pasture or montane sagebrush, but that it can also replace sagebrush and aspen. [**4 sites**]

S128 (Wyoming Basins Low Sagebrush): 40% producer's and 40% user's indicates that it's often replaced by big sagebrush or montane sagebrush, but that it can also replace montane and desert grasslands. [**5 sites**]

D01 (Non-specific disturbed): 0% producer's (0/0 user's) indicates that the one site got mapped as sagebrush. This class would need more sample sites. [**1 site**]

S092 (Subalpine/Montane Riparian Woodland): 0% producer's (0/0 user's) indicates that the one site got mapped as aspen (which may not be entirely incorrect). This class would need more sample sites. [**1 site**]

III. Functional Unit CO-3 (MMZ8) Mapping Methodology

Functional Unit or Mapping Zone:
CO-3 (Colorado's portion of a southern Rockies "Mega-mapzone"). Occasionally referred to as MMZ8 or MMZ8CO (the latter to reflect that it is only the portion of the original mega-mapzone 8 within Colorado)

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:

Good summer mosaics were especially hard to create for the core Rockies. Snow is acceptable and not surprising in the spring and fall mosaics, but it is critical to get summer imagery that is as snow free as possible. This is compounded by a serious problem of cloud cover in the summer months. In several cases (33/33, 33/34, 34/32, and 34/33), image alternates from outside of our standard temporal window (1999-2001) had to be acquired to patch up problem areas of our imagery (clouds for 33/33 and 33/34, snow for 34/32 and 34/33). In two cases we had to resort to Landsat 5 alternates. While the replacement of entire images might have been easier, replacement was restricted to the problem areas, in deference to the temporal window, through the use of AOI's. The AOI's (per scene) can be found at:

/CO/archive/CO3_MMZ8/AOI/PRE/.

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:

ETM+ Scene (path/row)	Spring (yr-Julian date)	Summer (yr-Julian date	Fall (yr-Julian date
33/33	2000-113	2000-226 2003-162 (L5 alt.)	1999-303
33/34	2000-130	1999-191 1996-175 (L5 alt.)	1999-287
33/35	2000-130	2000-258	1999-287
34/32	2001-107	2000-153 2002-174 (alt.)	1999-310
34/33	2001-107	1999-182 2002-174 (alt.)	1999-310
34/34	2001-107	2000-201	1999-310
34/35	2000-105	1999-182	1999-310
35/32	2001-162	2000-256	1999-317
35/33	2000-144	2000-256	1999-285
35/34	2000-144	2000-256	1999-285

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/CO3_MMZ8/MOSAICS/mz8_sceneseq_sprcl.img
/CO/ARCHIVE/CO3_MMZ8/MOSAICS/mz8coseq_sum_newimag.img
/CO/ARCHIVE/CO3_MMZ8/MOSAICS/mz8sseq_fall.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 resulting six band Landsat ETM+ mosaics for each season can be found at:

/CO/ARCHIVE/CO3_MMZ8/MOSAICS/mmz8co_ref_spr_aea.img
/CO/ARCHIVE/CO3_MMZ8/MOSAICS/mmz8co_ref_sum4.img
/CO/ARCHIVE/CO3_MMZ8/MOSAICS/mmz8co_ref_fall_aea1.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/CO3_MMZ8/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/CO3_MMZ8/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/CO3_MMZ8/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/CO3_MMZ8/IMG_FILES/MODELS/npv_ratio_c.gmd

All imagery derived predictor layers can be found at:

/CO/ARCHIVE/CO3_MMZ8/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: Expressed in degrees, as calculated by ERDAS Imagine's slope routine.

Aspect: A nine class aspect grid. 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 derived 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/CO3_MMZ8/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.

Extensive use was made of the U.S. Forest Service's "R2VEG" vegetation coverage (e.g.http://www.fs.fed.us/r2/sanjuan/projects/gis/index.shtml) to guide the augmentation of polygons. Some consideration was given to "burning in" these forest service polygons directly, but the guided augmentation approach permitted the better modeling of areas lacking R2VEG coverage, and provided a more independent mapping that was consistent with the rest of our effort; a separate assessment could be conducted to determine whether a direct "burn in" would have been more accurate.

b) Summary of samples:

11,144 samples were available for modeling. The polygon coverage containing all samples is located at: **CO/ARCHIVE/CO3_MMZ8/TRAIN_DATA/.**
In this particular mapzone, there is only a post-"shrunk" coverage, relating to a step necessary for modeling described in section 5.a). A pre-"shrunk" coverage would have had to have been created from scratch, as successive sets of shrunken augmentation polygons were added to a shrunken version of base layer (SWReGAP and Basinwide) polygons. (This approach saved a tremendous amount of time computationally.)

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/CO3_MMZ8/POST_MODEL/recodes.doc**; *the numbers below reflect the aggregation into the final classes.)*

The absence of any samples for S030 and S034 relates to the fact that these mesic versions of S028 and S032, respectively, were difficult to identify in the field (and from other data sources) and were therefore only mapped with post classification tree modeling as described in section 5.b).

CODE	**# SAMPLES**	**DESCRIPTION**
S001	2	NORTH AMERICAN ALPINE ICE FIELD
S002	239	ROCKY MOUNTAIN ALPINE BEDROCK AND SCREE
S004	100	ROCKY MOUNTAINS ALPINE FELL-FIELD
S006	153	ROCKY MOUNTAIN CLIFF AND CANYON COMPLEX
S008	2	WESTERN GREAT PLAINS CLIFF AND OUTCROP
S012	20	INTER-MOUNTAIN BASINS ACTIVE AND STABILIZED DUNES
S015	9	INTER-MOUNTAIN BASINS PLAYA
S023	782	ROCKY MOUNTAIN ASPEN FOREST AND WOODLAND
S025	87	ROCKY MOUNTAIN SUBALPINE-MONTANE LIMBER-BRISTLECONE PINE WOODLAND
S028	1197	ROCKY MOUNTAINS SUBALPINE DRY-MESIC SPRUCE-FIR FOREST AND WOODLAND
S030	0	ROCKY MOUNTAINS SUBALPINE MESIC SPRUCE-FIR FOREST AND WOODLAND
S031	696	ROCKY MOUNTAINS LODGEPOLE PINE FOREST
S032	767	ROCKY MOUNTAINS MONTANE DRY-MESIC MIXED CONIFER FOREST AND WOODLAND

S034	0	ROCKY MOUNTAINS MONTANE MESIC MIXED CONIFER FOREST AND WOODLAND
S036	644	ROCKY MOUNTAINS PONDEROSA PINE WOODLAND
S038	293	ROCKY MOUNTAIN PINYON-JUNIPER WOODLAND
S039	146	COLORADO PLATEAU PINYON-JUNIPER WOODLAND
S042	232	INTER-MOUNTAIN WEST ASPEN-MIXED CONIFER FOREST AND WOODLAND COMPLEX
S046	420	ROCKY MOUNTAINS GAMBEL OAK - MIXED MONTANE SHRUBLAND
S047	185	ROCKY MOUNTAINS LOWER MONTANE-FOOTHILL SHRUBLAND
S054	331	INTER-MOUNTAIN BASINS BIG SAGEBRUSH SHRUBLAND
S065	10	INTER-MOUNTAIN BASINS MIXED SALT DESERT SCRUB
S071	532	INTER-MOUNTAIN BASINS MONTANE SAGEBRUSH STEPPE
S079	443	INTER-MOUNTAIN BASINS SEMI-DESERT SHRUB STEPPE
S081	213	ROCKY MOUNTAINS DRY TUNDRA
S083	238	ROCKY MOUNTAIN SUBALPINE MESIC MEADOW
S085	739	SOUTHERN ROCKY MOUNTAINS MONTANE GRASSLAND
S086	68	ROCKY MOUNTAINS FOOTHILL GRASSLAND
S088	10	WESTERN GREAT PLAINS SHORTGRASS PRAIRIE
S090	85	INTER-MOUNTAIN BASINS SEMI-DESERT GRASSLAND
S091	478	ROCKY MOUNTAINS SUBALPINE/MONTANE RIPARIAN SHRUBLAND
S092	164	ROCKY MOUNTAINS SUBALPINE/MONTANE RIPARIAN WOODLAND
S093	112	ROCKY MOUNTAINS LOWER MONTANE RIPARIAN WOODLAND AND SHRUBLAND COMPLEX
S095	9	WESTERN GREAT PLAINS RIPARIAN WOODLAND AND SHRUBLAND
S096	147	INTER-MOUNTAIN BASINS GREASEWOOD FLAT COMPLEX
S100	25	NORTH AMERICAN ARID WEST EMERGENT MARSH
S102	344	ROCKY MOUNTAINS ALPINE/MONTANE WET MEADOW
N11	116	OPEN WATER
N21	31	DEVELOPED, OPEN SPACE-LOW INTENSITY
N22	46	DEVELOPED, MEDIUM - HIGH INTENSITY
N31	5	BARREN LANDS
N81	448	PASTURE/HAY
N82	135	CULTIVATED CROPS AND IRRIGATED AGRICULTURE
D02	12	RECENTLY BURNED
D03	34	RECENTLY MINED OR QUARRIED
D06	126	INVASIVE PERENNIAL GRASSLAND

D07	8	INVASIVE PERENNIAL FORBLAND
D08	24	INVASIVE ANNUAL GRASSLAND
D09	58	INVASIVE ANNUAL AND BIENNIAL FORBLAND
D10	154	RECENTLY LOGGED AREAS
D11	3	RECENTLY CHAINED PINYON-JUNIPER AREAS
A788	2	ARCTOSTAPHYLOS PATULA SHRUBLAND ALLIANCE
A1523	5	PURSHIA TRIDENTATA SHRUB HERBACEOUS ALLIANCE

3) Cover types:

a) Classification Tree (CT) modeled cover types:

See the 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 not used, as it was determined that these often had deleterious effects on the modeling.

All single-band predictors can be found at:
/CO/ARCHIVE/CO3_MMZ8/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/CO3_MMZ8/TRAIN_DATA/rp200623_100.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/CO3_MMZ8/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/CO3_MMZ8/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/CO3_MMZ8/AOI/.

Some of the recoding was more automated, such as the recoding of certain types within known urban areas. 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/mmz8_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). Similarly, any non-forested cover type that fell within a logged AOI (**CO/ARCHIVE/CO1_CO2_MMZ7/AOI/mmz7_gmug_rg_rt_sj.aoi**) was recoded to D10 (Recently Logged). This AOI was derived from the U.S. Forest Service R2VEG coverage.

An exact description of the conversions within these AOI's is given in:
CO/ARCHIVE/CO3_MMZ8/AOI/MMZ8_AOI_explanations.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/CO3_MMZ8/POST_MODEL/recodes.doc

Finally, a model was used to separate the Mesic from the Dry-Mesic versions of systems S028 and S032 to create systems S030 and S034, respectively. The .gmd file used for this

post-classification model can be found at:
CO/ARCHIVE/CO3_MMZ8/POST_MODEL/mesic_conifer_classes.gmd
This model is actually a combination of the two models described below:

Discriminating *S028-Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland* and *S030-Rocky Mountain Subalpine Mesic Spruce-Fir Forest and Woodland*: A post-classification model was used to discriminate *S028* from *S030*. The logic and parameters for the model were as follows:

This post-classification model was designed to extract *S030-Rocky Mountain Subalpine Mesic Spruce-Fir Forest and Woodland* from the standard modeled S028-*Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland*. Nature Serve describes *S030* to be *"typically found in locations with cold air drainage or ponding, or where snowpacks linger late into the summer, such as north-facing slopes and high elevation ravines. They can extend down in elevation below the subalpine zone in places where cold air ponding occurs; northerly and easterly aspects predominate. These forests are found on gentle to very steep mountain slopes, high elevation ridgetops and upper slopes, plateaulike surfaces, basins, alluvial terraces, well-drained benches, and inactive stream terraces."*

Model methods: A standard model was run withholding *S030*, consequently allowing *S028* to be mapped in its place. A condition statement was created to extract *S030* using the above aspect and landform characteristics in order to identify this subalpine mesic conifer system. The details of the conditional statement follow:

EITHER *S030* 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* = = *S028*)) OR *Standard Modeled Vegetation Output* OTHERWISE

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/CO3_MMZ8/POST_MODEL/alliance_recodes.doc

Both maps can be found at: **/CO/ARCHIVE/CO3_MMZ8/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/CO3_MMZ8/POST_MODEL/recodes.doc**, and from there to final "system" classes as described in section 5.c) and **/CO/ARCHIVE/CO3_MMZ8/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 were no mesic conifer classes 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/CO3_MMZ8/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. Neither is there a tremendous difference between evaluating with all of the training data or with SWReGAP field-collected polygons only.

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/CO3/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.

A1523 (Bitterbrush): 100% producer's and 50% user's may be reasonable, but not enough sites to adequately evaluate. [**1 site**]

A788 (Manzanita): 100% producer's and 100% user's appears good, but not enough sites to adequately evaluate. [**1 site**]

D02 (Recent Burn): 67% producer's and 100 user's suggest this is mapping well, but more sample sites would be preferable. [**3 sites**]

D03 (Mining): 14% producer's and 20% user's suggest that mining is not being mapped very well beyond the training data; confusion is with a variety of classes. [**7 sites**]

D06 (Invasive Perennial Grassland): 12% producer's and 50% user's suggest that it is generally being omitted, primarily to pasture, montane sagebrush, and montane grassland, but that at least it's not overmapping extensively, except for a fair amount of commission of montane grassland. [**26 sites**]

D07 (Invasive Perennial Forbland): 0% producer's (0/0 user's) suggests that it's being omitted but no commission suggests that at least it's not overmapping. It would need more sites to be better evaluated. [**2 sites**]

D08 (Invasive Annual Grassland): 0% producer's and 0% user's (3 errors of commission) suggest that it's both being omitted and overmapped a little bit. [**5 sites**]

D09 (Invasive Annual and Biennial Forbland): 17% producer's and 50% user's suggest that it's being heavily omitted – mostly to semi-desert shrub steppe, but that it's not overmapping extensively. [**12 sites**]

D10 (Recently Logged): 29% producer's and 90% user's indicates that it is has a strong tendency to be omitted, primarily in favor of spruce-fir and alpine/montane wet meadow, but that it is not overmapping very much at all. [**31 sites**]

D11 (Chained Pinyon-Juniper): 100% producer's and 100% user's are encouraging, but not enough sites to adequately evaluate. [**1 site**]

N11 (Water): 88% producer's and 100% user's indicate that it is mapped very well, except for some slight omission to alpine bedrock, which may be related to water looking like steep shaded bare slopes. [**24 sites**]

N21 (Low Intensity Development): 29% producer's and 50% user's indicate that it is often being omitted, especially in favor of pasture, but not overmapping extensively. [**7 sites**]

N22 (High Intensity Development): 50% producer's and 42% user's indicate that it is fairly regularly omitted, often to Rocky Mountain Cliff and Canyon, but also overmapped, particularly in place of Rocky Mountain Cliff and Canyon. The former problem was handled with urban area post-processing. [**10 sites**]

N30 (Barren): 100% producer's and 100% user's are encouraging, but not enough sites to adequately evaluate. [**1 site**]

N81 (Pasture): 71% producer's and 54% user's indicate that there is some omission of pasture to such classes as big sagebrush and Gambel oak, but, more commonly, it is being mapped in place of alpine/montane wet meadow, agriculture, and, to a lesser extent, Gambel oak and invasive perennial grass. [**90 sites**]

N82 (Agriculture): 37% producer's and 67% user's indicate that it is being fairly heavily omitted in favor of pasture, but that it's not being overmapped extensively. [**27 sites**]

S001 (Alpine Ice Field): 0% producer's (0/0 user's) suggests that it's being omitted but no commission suggests that at least it's not overmapping. It would need more sites to be better evaluated. [**1 site**]

S002 (Alpine Bedrock and Scree): 83% producer's and 80% user's suggest that it is being mapped fairly well – just a little bit of omission to Alpine Fell-Field and Rocky Mountain Cliff and Canyon, as well as some commission of Alpine Fell-Field. [**48 sites**]

S004 (Alpine Fell-Field): 45% producer's and 60% user's indicate that it is fairly frequently omitted to Dry Tundra and Alpine Bedrock, but also commits those two classes to some extent. [**20 sites**]

S006 (Rocky Mountain Cliff and Canyon): 58% producer's and 55% user's suggest that there is some omission to Rocky Mountain P-J Woodland and High Intensity Development, as well as some commission of several classes, primarily Alpine Bedrock and Scree and Montane Grassland. [**31 sites**]

S008 (Great Plains Cliff and Outcrops): 0% producer's (0/0 user's) suggests that it's being omitted but no commission suggests that at least it's not overmapping. It would need more sites to be better evaluated. [**1 site**]

S012 (Active and Stabilized Dunes): 50% producer's and 50% user's indicate that this class is being mapped reasonably well, with no clear pattern of confusion with other classes, but more sites would be needed to thoroughly evaluate its accuracy. [**4 sites**]

S015 (Playa): 50% producer's and 100% user's are somewhat encouraging, but there are too few sites to be conclusive. [**2 sites**]

S023 (Aspen): 86% producer's and 73% user's indicate that it's only rarely being omitted to other classes such as Subalpine/Montane Riparian Shrubland and Mixed Conifer, but that it more frequently is mapped instead of Subalpine/Montane Riparian Shrubland and Aspen-Mixed Conifer, and, to a lesser extent, Alpine/Montane Wet Meadow and Subalpine/Montane Riparian Woodland. [**157 sites**]

S025 (Subalpine Limber-Bristlecone Pine): 6% producer's and 25% user's suggest that it is being heavily omitted in favor of Spruce-Fir, Lodgepole Pine, and Ponderosa Pine, while also sometimes mapping in the place of similar high elevation classes. [**18 sites**]

S028 (Spruce-Fir): 83% producer's and 70% user's suggest that it is only occasionally mapped as Lodgepole Pine or Mixed Conifer, but that, more often, it maps in place of these classes, in addition to the Aspen-Mixed Conifer and Recently Logged areas. [**240 sites**]

S031 (Lodgepole Pine): 60% producer's and 60% user's suggest that it is somewhat frequently mapped as Spruce-Fir, and occasionally as Mixed Conifer, and that it also maps instead of these classes fairly often. [**140 sites**]

S032 (Mixed Conifer): 60% producer's and 54% user's suggest that it is fairly often being replaced by Ponderosa Pine, Lodgepole Pine, and Spruce-Fir, but that it is also mapping in place of these classes, in addition to Aspen-Mixed Conifer and Subalpine / Montane Riparian Woodland. [**154 sites**]

S036 (Ponderosa Pine): 74% producer's and 65% user's indicate that this class is only replaced by Mixed Conifer, but that it also replaces Mixed Conifer, in addition to Rocky Mountain Pinyon-Juniper, Lodgepole Pine, and Subalpine Limber-Bristlecone Pine. [**129 sites**]

S038 (Rocky Mountain Pinyon-Juniper): 70% producer's and 77% user's demonstrate that is fairly often mapped as Ponderosa Pine, and occasionally as Montane Grassland, while it also maps instead of Ponderosa occasionally. [**59 sites**]

S039 (Colorado Plateau Pinyon-Juniper): 73% producer's and 76% user's indicates that this class is being mapped reasonably well, with no strong confusion with any other class. [**30 sites**]

S042 (Aspen-Mixed Conifer): 28% producer's and 65% user's indicate that this class is strongly replaced by spruce-fir, aspen, and mixed conifer, but that it does not overmap extensively, only occasionally replacing aspen and Subalpine/Montane Riparian Woodland. [**47 sites**]

S046 (Gambel Oak): 76% producer's and 63% user's suggest that it is occasionally mapped as montane sagebrush and pasture, but that it is more frequently mapped instead of other classes, such as Lower Montane Riparian Woodland, Subalpine/Montane Riparian Shrubland, and Pasture. [**84 sites**]

S047 (Foothill Shrubland): 38% producer's and 58% user's indicate that it is often replaced by such classes as montane grassland, montane sagebrush, and pasture, and that it occasionally replaces a variety of classes. [**37 sites**]

S054 (Big Sagebrush): 76% producer's and 64% user's indicate that this class is only occasionally replaced by montane sagebrush and semi-desert shrub steppe, but that it more frequently replaces such classes as semi-desert shrub steppe and pasture. [**67 sites**]

S065 (Mixed Salt Desert Scrub): 0% producer's and 0% user's (1 error of commission) indicate that this class is getting replaced by Pasture, but that it also maps in place of Big Sagebrush in one instance. More sites are required to better evaluate this class. [**2 sites**]

S071 (Montane Sagebrush Steppe): 75% producer's and 63% user's indicate that it is occasionally replaced by montane grassland and subalpine meadow, but that it just as often replaces these classes, in addition to many other classes such as Gambel Oak and Big Sagebrush. [**107 sites**]

S075 (Juniper Savanna): 0% producer’s (0/0 user’s) indicates that this class is being heavily replaced by other classes such as montane sagebrush, but that it’s not overmapping. [**3 sites**]

S079 (Semi-Desert Shrub Steppe): 63% producer’s and 62% user’s indicate that there is a fair amount of replacement by Greasewood, with some replacement by Montane Grassland and Big Sagebrush, while this class replaces Desert Grassland, Greasewood, and Big Sagebrush. [**89 sites**]

S081 (Rocky Mountain Dry Tundra): 86% producer’s and 79% user’s indicate that this class is being mapped very well except for occasionally replacing Alpine Fell-Field. [**43 sites**]

S083 (Subalpine Meadow): 50% producer’s and 50% user’s indicate that this class is replaced by a variety of classes, including Montane Grassland, Subalpine/Montane Riparian Shrubland, Montane Sagebrush, and Alpine/Montane Wet Meadow, while it just as frequently replaces these same classes. [**48 sites**]

S085 (Montane Grassland): 72% producer’s and 62% user’s indicate that it is sometimes replaced by montane sagebrush and subalpine meadow, but more frequently it replaces other classes such as Semi-Desert Shrub Steppe, Montane Sagebrush, Subalpine Meadow, and Foothill Shrubland. [**148 sites**]

S086 (Foothill Grassland): 64% producer’s and 50% user’s indicate that it is occasionally mapped as other classes such as Semi-Desert Shrub Steppe, but that more frequently it replaces other classes such as Foothill Shrubland and Invasive Perennial Grassland. [**14 sites**]

S088 (Shortgrass Prairie): 0% producer’s (0/0 user’s) indicates that it may get replaced by other grasslands away from the training sites, but that at least it’s not overmapping extensively. More sites would be needed to evaluate it more thoroughly. [**2 sites**]

S090 (Semi-Desert Grassland): 12% producer’s and 40% user’s indicate that it’s generally getting replaced by Semi-Desert Shrub Steppe and Greasewood, while it also occasionally replaces Semi-Desert Shrub Steppe. [**17 sites**]

S091 (Subalpine/Montane Riparian Shrubland): 53% producer’s and 60% user’s indicate that this class is often replaced by Aspen, sometimes by Gambel Oak, occasionally by Montane Grassland and Montane Sagebrush, and often replaces Alpine/Montane Wet Meadow, Subalpine Meadow, and Aspen. [**96 sites**]

S092 (Subalpine/Montane Riparian Woodland): 0% producer’s (0/0 user’s) is amazingly low for a class with 33 samples – but at least it’s not overmapping extensively(!) – there’s no commission error. This class is being entirely replaced by such classes as Mixed Conifer, Aspen, and Lodgepole Pine. [**33 sites**]

S093 (Lower Montane Riparian Woodland and Shrubland): 30% producer’s and 70% user’s indicate that this class is often being replaced by other classes such as Gambel Oak, but that it less frequently maps instead of other classes such as Arid West Emergent Marsh. [**23 sites**]

S095 (Western Great Plains Woodland and Shrubland): 50% producer’s and 50% user’s are probably inconclusive due to the lack of sample sites. [**2 sites**]

S096 (Greasewood Flat): 73% producer’s and 46% user’s indicate that it is occasionally replaced by Semi-Desert Shrub Steppe, but more frequently, it replaces Semi-Desert Shrub Steppe, and occasionally replaces Semi-Desert Grassland. [**30 sites**]

S100 (Arid West Emergent Marsh): 0% producer’s (0/0 user’s) indicates that this is either called Pasture or Lower Montane Riparian Woodland and Shrubland, but that it’s not overmapping extensively. [**5 sites**]

S102 (Alpine/Montane Wet Meadow): 26% producer’s and 41% user’s suggest that it’s being heavily replaced by such classes as Pasture, Subalpine Meadow, Montane Grassland, Aspen, and Subalpine/Montane Riparian Shrubland, but that it’s also overmapping a lot of these same classes, in addtion to Recently Logged Areas. [**69 sites**]

IV. Functional Unit CO-4 (MMZ9) Mapping Methodology

Functional Unit or Mapping Zone:
CO-4 (Colorado’s Eastern Plains). Sometimes 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

<u>1) Predictor layer preparation:</u>

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	#	DESCRIPTION

	SAMPLES	
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
D04	230	INVASIVE SOUTHWEST RIPARIAN WOODLAND AND 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**]

V. References

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.