

***Point Reyes
Classification of Land Cover
from Satellite Imagery:
Final Report***

2005



**Wildlife Spatial Analysis Lab
The University of Montana
Missoula, MT 59812**

**Roland Redmond, Director
Jim Schumacher
Shane Mason
Will Gustafson
Chris Winne**

January 31, 2005

INTRODUCTION, IMAGERY, AND SOFTWARE

As part of the ongoing vegetation mapping effort at Point Reyes National Seashore (referred to as PRNS throughout this report), Ikonos imagery was assembled here at University of Montana into a seamless layer for classification. It encompasses lands containing Golden Gate National Recreation Area units from the north end of Tomales Bay at the top end of PRNS, and extending south to the Golden Gate Bridge, including part of Marin County. Parklands include all of Point Reyes N.S., Samuel P. Taylor State Park, Tomales Bay State Park, Muir Woods National Monument, Mount Tamalpais State Park, Audubon Canyon Ranch, Fort Baker, Marin Headlands, Tennessee Valley, Stinson Beach, and other lands administered by Golden Gate NRA. The total area classified is 78,078 hectares, 192,935 acres, or 302 square miles. Those figures include a small buffer zone of 400 meters of non-vegetation coastal waters in the classification.

The graphic below displays the classification study area with the Ikonos imagery in “natural color”. The top of the map points due north, with Golden Gate at lower right and San Pablo Bay at right. Note the presence of clouds which obscured the classification: the white patch of clouds along the edge of one of the Ikonos images on Point Reyes, and a small patch hugging the coastline southeast of Stinson Beach/Bolinas Bay. (Also note, blue water areas and gray non-classified lands on this map are patched in from a different source: water has a vastly different appearance on the imagery itself.)



The Ikonos imagery and resulting classification grid are 4-meter resolution. The imagery from the Ikonos-2 sensor is 4-band multispectral: 3 visible bands plus near-infrared. (The panchromatic 1-meter resolution band was not utilized.) The imagery is dated early July, 2003. Preparation of the imagery and some visual analyses were done in Erdas Imagine software. Image segmentation was output from eCognition software based on the 4 spectral bands (using parameters of Scale 50, Shape Factor .2, Smoothness .8). Each region (raster polygon area) was attributed for spectral and topographic statistics. After converting imagery to ArcInfo grid format, attributes were also added separately in ESRI ArcInfo software for proximity to water, to aid in classifying vegetation types adjacent to water. Training data were examined and coded in ArcGIS. Classification, as described below, is accomplished using the data-mining software WEKA and then output visually checked in ArcGIS.

TRAINING DATA

Data from field plots were taken from the Microsoft Access file provided by PRNS. These point data were converted into an ArcInfo coverage consisting of 1,915 points labeled by their vegetation *PI-Code* and other attributes such as vegetation *Alliance*, along with their original plot identification number.

Imagery to be classified for this project covers Golden Gate NRA lands north of Golden Gate Bridge, thus field plots south of there were omitted, leaving 1,525 points.

Plots missing attributes for both *PI-Code* vegetation class and *Alliance*, or *PI-Code* = 99999 and *Alliance* “unable to key”, could not be used to classify at the desired level, and thus another 76 points were omitted. Also omitted were another 36 plots, where the field crew was unable to key the *Alliance* (*Alliance* = NA, not available). Many of those points are too general, with *PI-Codes* such as 24000 or 55000 or 67000, to serve as good training examples.

Omitting the points that PRNS was unable to key drops the total from 1,525 to 1,413.

Analysis of Training Points

Each of the 1,413 points was then manually inspected on-screen, one by one, for validity in relation to the segmented regions and the Ikonos imagery. Using finer-scale DAIS one-meter imagery as a background check helped confirm the validity of plot location and suitability.

Points were examined and coded as suitable/unsuitable. Suitability and locational problems consisted of:

1. Multiple field plots within segmented region/polygon (we then omit ones that are judged not meant for that particular patch of vegetation).
2. Points on the very edge of regions/polygons where it could not be positively determined which vegetation patch was intended.

3. Points covered by cloud.
4. Points simply deemed poor training examples in comparison to the majority, after gaining expertise in identifying imagery.
5. Plot is more than likely not located in intended vegetation. Occasionally plots were placed in a donut hole surrounded by the vegetation, such as GPS location of a tree plot as standing on bare ground while surrounded by the intended trees (probably figuring it would be obvious during mapping); or else plot placed on the edge of a patch instead of actually in it.
6. Percent cover of the intended vegetation is far too low for that particular segmented region/polygon to be a suitable training plot. For example, a tree plot where imagery displays about 90% bare ground with just a few scattered trees, or a shrub plot that is in just a small patch but our segmented region is primarily forest.
7. For just a few points, obvious mistakes: such as a grass point in a thick stand of trees.

It is important to note that the PRNS field crew (or other collecting agency) was not placing plots based on our delineated regions, as the field data were collected earlier for a different project. Data were also not originally collected for use with fine-resolution multi-spectral imagery. Thus, the plots are not always a good fit in terms of sampling our regions.

Sometimes the problems above reflect drawbacks of the image segmentation process: occasionally a region/polygon might be a poor shape. Where possible during inspection, points were manually moved to vegetation polygons for which they were obviously intended (and coded as having been moved, in case they were utilized later on). This was often done for plots lying on the edges of polygons.

After the first rounds of classification produced less than desired results, omitting poor examples and manually moving points generated a big jump in classification accuracy. This improvement was accomplished after beginning to utilize the DAIS imagery and would not have been possible with Ikonos alone. Based on how the points looked on both Ikonos and DAIS imagery (utilizing multiple combinations of how the spectral color bands were displayed), 189 points were moved and 161 points were coded as unsuitable.

Ten other training plots from the original set were dropped early on. There was one *Calif. Buckeye 14020* plot received from PRNS, but a single training point is not enough to map and as Buckeye differs in some characteristics from other trees, it was dropped rather than allowed to weaken the training set by lumping it with another class. Likewise, a single *Coast Buckwheat 62050* plot was dropped as it did not appear to belong with the other *62000-duneveg* types. There were 4 *Gorse 24999* original plots, all in a very small geographic area on the northeast shore of Tomales Bay near Preston Point, southwest of Tomales town. The Gorse vegetation patches appeared dissimilar to the other scrub types on the imagery and were dropped since additional training data would be required to classify it separately. The final classification did end up mapping those Gorse vegetation patches with the scrub *Coyotebrush* class, as was hoped. There were 4 *Hazel 30010* original plots, but they occurred in deeply shaded areas of forest and could not be distinguished on the imagery and thus, they were dropped.

Manual Addition of Training Data

New points were added on-screen for a few under-represented *PI-Codes* (and given new unique *PI-Code* numbers to distinguish from the PRNS data). Classification of non-vegetated land cover was never intended for this project and only a small number of plots were provided from PRNS: 3 points for beach, 15 points for water, zero for urban/developed. In order to fill in those land areas with a general non-vegetated class and so avoid confusion within the classification, points were added for developed areas and water. Also added to improve the classification were points for grass and dunegrass, as described below. When adding points, locations were spread around the full geographic area for broader variance. Points added consist of:

PI-CODE	# POINTS
47011	17
67000	40
67001	165
90300	52
90301	7
98001	25
90999	18
3121	6

330 total new points	

Table 1: Manually added training points.

47011 and 67000:

The first rounds of classification showed too much confusion between grass, scrub, and dune types. One of the primary dune types which was fairly distinguishable on the imagery (and thus possible to augment), was *47010 European Dunegrass*; so another 17 points were added to the original 10. Grass was being severely underclassified, perhaps due to its under-representation of points: only 39 original plots, compared to over 300 for coast-scrub types and over 300 for marsh types. 40 more grass points were added in obvious bare grassy areas and coded as 67000.

67001:

Later classification rounds still were missing areas of grass, such as the east side of Olema Valley. Another 165 points were added to give a better representation of the full spectrum of grass types, using both the Ikonos and DAIS imagery as background display.

90300 and 90301:

These were 2 variations of developed land, with 90301 being buildings of a different spectral color than the more common 90300.

98001:

Water points in obvious water bodies: coastal bays or reservoirs up in the east hills.

90999:

Cloud points, in order to classify cloud areas separately from other non-vegetation.

3121:

Additional *Monterey Cypress/Monterey Pine* points. The original twelve *3120* plots were typically located at isolated patches of a single large tree or clump of trees (this is a planted non-native forest type). It was thought that with more points, perhaps an additional 1 to 2 dozen, it might be possible to classify this type separately rather than grouping it with *Bishop Pine 3030*. Not enough of these isolated patches were positively identified to add many more points, but the 6 points that were added were kept in the training set.

After all new points were added, they were attributed with the full set of imagery statistics and other variables being used for classification just as were all the other training data. (See attribute details below.)

After omitting and adding training examples, this leaves 1,511 points judged to be suitable for classification based on visual assessment. Further evaluation was provided by the classification software WEKA, which checks consistency of points with all their attributes using its same ensemble of learner algorithms as described in methods below, and removes more or less points, depending on user input, as outliers in its pre-processing classification filter. For our final classification, WEKA used 1,429 training points in its classifier model.

LAND COVER CLASSES

We attempted to group vegetation types as best as possible to follow the existing classification hierarchy of the U.S. National Vegetation Classification System (USNVCS) or the PRNS modified hierarchy, with a few exceptions as described below. Many *Alliances* were severely under-represented in all training sets and some possible *Alliances* did not occur in the training set at all. Because of this, initial classification rounds were done on the broader *Group* scheme, and then later rounds refined to aggregate *Alliances* into different classes and to be similar to PRNS *Mesocluster* level. The *Mesocluster* is a mid-level classification of vegetation types that share similar ecological habitat. The PRNS *PI-Code* numbering scheme also helped guide the grouping, such as the 2000-level forest types being one class and the 3000-level types the next class. Notes on how and why specific types were combined follow below.

For the final classification, each training example was cast into one of the following classes based on its *Alliance/PI-Code*. Table 2 shows the number of points per class in the training set. Note that the table shows 13 classes (as does the confusion matrix shown later). However, the two nonveg classes should be combined for final land cover types to leave 12 classes total: 10 vegetated types plus nonveg and cloud.

# POINTS	CLASS NAME
81	bishop-montr
36	chaparral
18	cloud
343	coyotebrush
108	dougfir-redw
38	duneveg
183	grass-marsh
288	grass-upland
93	nonveg
7	nonveg2
50	saltmarsh
132	sclero-tree
134	willow-alder

1511	total points

Table 2: Training Points by Class

Following is a list of the PI-Codes by class, including a description of each group. Class names were limited to 12 characters for ease of use as field names in the GIS database.

sclero-tree

1010 *Sclerophyllous Forest: Calif.Bay, Eucalyptus, Tanoak, Chinquapin, and*
1011 *Coast Live Oak. Eucalyptus* could possibly be classified separately as it looks
1012 somewhat distinguishable on imagery, but given only 5 training points, it is
1014 grouped with the other sclerophyllous tree types. *Tanoak* could not be
1030 classified separately in a *Redwood-Tanoak Mesocluster* because there were
1070 only 2 training plots for *Tanoak* and insufficient *Redwood* as well, so *Tanoak*
1091 is left grouped with the other sclerophyllous trees.
12020
12021
12022

dougfir-redw

2010 *Douglas-fir, Coast Redwood.* An attempt was made to classify these two
2011 separately but with very poor results for Redwood at less than 20% correctly
2012 classified; thus, they are grouped together.
2013
2014
2015
2020
2050
2051
2052

bishop-montr

3030 *Bishop Pine, Monterey Cypress/Monterey Pine.* An attempt also was made to
3031 classify these two separately but with very poor results for *Monterey Cypress*
3120 at 33% accuracy; thus, grouped together instead. *Monterey Cypress* is a non-
3121 native species that occupies only a very small percentage of the area.

willow-alder

7000 *Arroyo, Red, Black, and Yellow Willow types; Red Alder; Wax Myrtle; and*
7030 *Salmonberry.* This is identical to PRNS *Riparian Forest Mesocluster*. Initial
7061 classifications had *Wax Myrtle* and *Salmonberry* as part of a “Deciduous
7062 shrub” class but with poor results; thus, they follow the PRNS scheme instead.
7070 The “Deciduous shrub” class had accuracy results ranging from under 10%
7071 to the mid-20's and thus was scrapped. Either input data are lacking, or those
7072 deciduous shrubs form a heterogeneous mix with other types, or they simply
20010 are difficult to distinguish with satellite imagery. Those types were instead
30050 split between this riparian class or the coastal-scrub *coyotebrush* class.
32080 *Wax Myrtle* and *Salmonberry* also had very low accuracies if attempting to
classify them separately on their own, and are appropriate to the riparian class.

coyotebrush

19010 Primarily *Coyote Brush*, along with related scrub types: *Broom*, *Sagebrush*,
20020 *Coffeeberry*, *Yellow Bush Lupine*, *Blue Blossom*, *Poison Oak*, *Iceplant*.
21460 *Iceplant* might be better suited ecologically with the *duneveg* class; however,
24040 it had better results as part of this scrub class, which also occupies much of the
24050 same ecological niche beside the coastal dunes. We surmise *Iceplant* may be
24051 classified more correctly as scrub due to its growth pattern of forming dense
24052 green mats of vegetation, as contrasted with some of the other *duneveg*
24053 plants' habit of being mixed with sandy areas. Attempting to classify *Iceplant*
24054 separately on its own resulted in zero out of 18 points classified correctly.
24055 *Coffeeberry* 21460 has a *PI-Code* in the 21000-*Chaparral* range, but it is cast
24056 with these other shrub types and thus is similar to the PRNS grouping. It also
24057 visually does not look like the *chaparral* vegetation on the imagery.
24058
24059
24060
24061
24062
24063
24064
24065
24066
24068
24080
24099
30040
30041
62040

chaparral

21000 *Manzanita*, *Chamise*. This is identical to the PRNS *Chaparral Supercluster*
21110 and includes all of the manzanitas and chamise. These are primarily xeric
21140 types growing farther inland.
21142
21210
21260
21440
21470
21480

grass-marsh

46020 Marsh grasses and rushes, etc., along with lush green non-native grass such as
46021 lawn grass: *Pacific Reedgrass*, all the *Rushes*, *Bulrush*, *Slough Sedge*, *Cattail*,
46022 and *Spikerush*. These are the fresh-water marsh types. We were not provided
52030 training data for developed lawns, but initial classification rounds continually
52031 grouped these areas along with the “marsh” as most similar. These areas
52032 include the obvious golf courses in southeast Marin portion of the study area.

52060 Thus this class is named *grass-marsh* as a combined marsh and grass class,
52061 to distinguish it from upland grasses. Perhaps with additional training data,
52070 lawn grass could be classified separately. While *European Dunegrass* is
55020 grouped with the *duneveg* class, *Reedgrass* is mixed in with the marsh plants.
55030 In fact, the majority of the training plots for *Reedgrass* are *Reedgrass-Sedge-*
55040 *Rush* mixed-species plots.
55050

grass-upland

47030 Upland grass types. These are the drier native and non-native grasses. *Tufted*
52040 *Hairgrass* 52040 is included as it is listed by PRNS as more of an upland type,
67000 rather than with the other 52000-Marsh types.
67001
67010
67020
67030
67040

duneveg

47010 *European Dunegrass, Dune Sagewort, Dune Lupine-Goldenbush.* This class
47011 is akin to the PRNS *Dune Vegetation Mesocluster*, although as noted above,
62060 *Iceplant* is not included.
62061
62064

saltmarsh

51010 *Saltgrass, Cordgrass, Pickleweed.* This class is identical to the PRNS *Salt*
51011 *Marsh Mesocluster.* Initially these types were combined with the
56010 other marsh plants, but as classification success was improved for
64030 distinguishing the marsh plants from other shrubs or dune plants, this class
64032 was also pulled out separately.

nonveg

90300 Urbanized, developed, beaches, water. As noted earlier, classification of
90301 these was not one of the goals and we were provided little training data.
90402 These were then included in order to fill in those land areas and to prevent the
98000 software from confusing them with other types. 90300 and 90301 are two
98001 spectrally different types of urban areas or buildings, producing two classes
after classification, *nonveg* and *nonveg2*, with the intention of combining
these two for the final land covertype product. *Nonveg2* occupies just a tiny
area in comparison to the rest of the *nonveg* types. 98001 represents points
added for water in obvious spectrally “dark” areas of water in bays or inland
reservoirs. However, input imagery displays water in a wide spectral range for
different parts of the study area – from dark black to light purples and white.
Thus it was all grouped together and agreed that PRNS could replace water
areas from a separate source. Because the *nonveg* types are all grouped
together, display of the classification somewhat oddly shows them all

merging together with no boundary between water and beach, or water and city, such as at Sausalito.

cloud

90999 Areas of a few clouds on the imagery, added as a class so that these areas would not get lumped in with the *nonveg* class. Training points were added in the thicker parts of clouds, but there is undoubtedly some confusion in the classification at the thinner edges of clouds.

REGION ATTRIBUTES SELECTED FOR CLASSIFICATION

A variety of attributes were tested for usefulness in developing accurate learning algorithms for the WEKA data-mining classification software. Imagery statistics are output from eCognition and attributed to each segmented region; proximity to water features is separately attached. An ArcInfo program attaches all variables to each training point by location. Databases for the training set and for the output classification grid are input to WEKA, and the best variables are determined in WEKA. This selected set of variables is specific to the Point Reyes area imagery and a somewhat different set might be chosen for a different geographic area.

The following list of 23 “best variables” were used in this classification:

Ikonos 1 Mean: Mean pixel value for the region for Ikonos spectral band 1
Ikonos 2 Mean: Mean pixel value for the region for Ikonos band 2
Ikonos 3 Mean: Mean pixel value for the region for Ikonos band 3
Ikonos 4 Mean: Mean pixel value for the region for Ikonos band 4
Ikonos 4 Std-dev: Standard deviation from the mean pixel value Ikonos band 4 of the region
Ikonos 1 Ratio: Mean band 1 divided by sum of all spectral layers' mean values for the region
Ikonos 2 Ratio: Mean band 2 divided by sum of all spectral layers' mean values for the region
Ikonos 3 Ratio: Mean band 3 divided by sum of all spectral layers' mean values for the region
Ikonos 4 Ratio: Mean band 4 divided by sum of all spectral layers' mean values for the region
Ikonos 4 Meandiffneighb: Mean difference to neighboring region for band 4
Brightness: Sum of mean values of all Ikonos spectral layers for region, divided by quantity
Max Difference: The maximum difference of region pixel values for all Ikonos spectral layers
Mean NDVI: Mean value Normalized Difference Vegetation Index (NDVI) for each region
Ratio NDVI: Mean NDVI divided by sum of all spectral layers' mean values for the region
Meandiffneighb NDVI: Mean difference to neighboring region for NDVI
Std-dev NDVI: Standard deviation from the mean NDVI
Mean DEM: The mean elevation for each region
Std-dev DEM: The standard deviation from mean elevation
Meandiffneighb DEM: Mean difference to neighboring region for elevation
Mean Slope: The mean slope for each region
Mean Aspect: The mean topographic aspect for each region
Proximity Stream: Focal sum of small radii (12 m. or about 40 ft.) from stream
Proximity Body: Focal sum of small radii (16 m. or about 52 ft.) from body of water

These variables were selected based on several attribute fitness tests in WEKA. WEKA has the capability of analyzing datasets with a large number of attributes, and this allows us to include far more variables than classification programs which input just the basic spectral band values, for example. Initially, 35 statistics variables were input for WEKA and then 12 were removed because they did not add meaningful information for

classification. Inclusion of too many attributes can hinder the classification results by adding extra confusion: the number of correctly classified points during cross-validation accuracy assessment (see below) actually increased with removal of weak attributes.

WEKA analyzes variables with attribute evaluators such as *CfsSubset*, *InfoGain*, and *ChiSquared*, and ranks subsets of attributes for merit. We ran 6 different WEKA attribute evaluators in 10-fold cross-validation evaluation mode, and took the results in ensemble for removal of 12 attributes. For instance, multiple evaluators agreed on removing standard deviation statistic for spectral bands 1, 2, and 3; but all agreed on keeping std-dev for band 4, the near-infrared band. All of the evaluators agreed on removing the “shape” indices such as *lengthwidth* and *compactness*. The most important measures for the overall classification were based on mean pixel values. Those variables based on the water proximity measures were extremely useful in assigning labels to groups that occur near water.

CLASSIFICATION

Description

Machine learning algorithms have proved to be very successful in classifying landcover from remotely sensed data. MODIS data is classified with a decision tree, Feature Analyst uses neural networks, and eCognition uses a single nearest neighbor approach. We employed the machine learning software WEKA for classification of the data. (WEKA is an acronym for Waikato Environment for Knowledge Analysis, University of Waikato, New Zealand; Weka is also the name of a bird - a flightless New Zealand rail.) WEKA has been utilized most frequently in the business world for data-mining techniques and classification of large amounts of data through pattern recognition, but it can employ a great variety of learner algorithms useful to classification of our spectral and topographic statistics, such as nearest neighbor and particularly decision trees. Decision trees predict class membership by recursively partitioning (“pruning”) a data set into more homogeneous subsets. Trees can be better suited than other classification techniques in those situations where a cover type is represented by more than one set of remote sensing characteristics – in our case, the 23 variables listed above – and it is immediately apparent to the software which variables contribute to the discrimination between classes.

The problem with most machine learning algorithms is they cannot fully refine the concept space of a complicated class. This means that each algorithm has a certain amount of error associated with its prediction. There are many methods for mitigating this error, the best of which revolve around combining various types of learning algorithms into a committee or ensemble and allowing them to vote on their final prediction. In this manner, 7 learners each had a vote in labeling each region. Another statistic output from WEKA is a “percent confidence” that the region/polygon is indeed that type, and this attribute is also attached to the final product grid.

Thus, an ensemble of machine learning algorithms was used to classify the segmented

regions. Varying types of base learners were used in the ensemble to induce variation in their error spaces, which allows for more stable population votes. For each region, covertype label was assigned based on a majority vote of the ensemble's component learners. Our ensemble of 7 learners (described below) consisted of 4 decision-tree algorithms (including one that is a “forest” of trees), 2 nearest-neighbor, and 1 rule-based. Thus, it is weighted in those proportions (4-2-1) as each of those 7 has one vote in labeling the regions. The results of the ensemble are greatly improved over the results of any single learner.

Ensemble Components

WEKA J48: Implementation of *Quinlan's C4.5* decision tree. This learner typically produces accurate results so the ensemble includes one of these.

WEKA REPTree: Learner that uses gain and variance to build a tree and then simplifies the tree with reduced error pruning. This method is quite different from the *C4.5* so these learners tend to error in differing areas of the input space. One *REPTree* was present in the ensemble.

WEKA RandomForest: Learner that constructs a forest of 10 random trees rather than a single tree: very robust and increases our classification accuracy. The learner grows 10 decision trees, each tree gives a classification, and the forest chooses the classification having the most votes. A fairly accurate classification could be produced using just one *RandomForest* learner along with one *KNN* learner.

WEKA DecisionStump: Learner which uses regression or classification (based on entropy). Decision stump is usually used in conjunction with a boosting algorithm, but in our case also improved correctly-classified accuracy with our other learners, whereas adding another boosting learner did not.

WEKA KNN: Implementation of the k-nearest neighbor learning algorithm. This learner assigns classification based on Euclidean distance, which produces fairly accurate classifications. As it relies on a distance-based metric, this method arrives at its solution space in a completely different manner than the others, which helps to foster variation in the error space. Two *KNN* learners were present: as a good traditional technique for land cover classification, we wanted *KNN* to cast two votes in the ensemble, not just one. We include one with $k = 5$ and one at $k = 10$.

WEKA Decision Table: This rule-based learner was selected specifically for the variation created by the manner in which it arrives at its solution space. Decision trees work from the top down, seeking to split on the attribute that provides the most information gain at each split. On the other hand, rule-based learners look at each class in turn and seek to create a set of rules that cover every instance in that class.

Ensemble Evaluation, Classification Accuracy

The learning method discussed above was tested for fitness with WEKA's own bootstrapping method of a stratified ten-fold cross-validation, which is the standard for evaluating machine learning schemes. Cross-validation methods hold out a percentage of the training examples with the same distribution as the overall data and then test the trained learner's fitness against the held out examples. At each fold a new ensemble of learners are trained on nine-tenths of the data and evaluated on the remaining one-tenth. In this method, the learning scheme is created, trained, and evaluated ten times so that the learners are never evaluated on any portion of its own training set. At the end of ten runs, the results are averaged to produce a statistical evaluation of the scheme's accuracy performance.

We also checked output accuracy of the training samples using a common percentage-split scheme, where a certain portion of points are held out. The difference in accuracy from the 10-fold cross-validation method was not more than a few percentage points, whether it was set to train on 2/3 of points and hold out 1/3 to test accuracy on those remaining 1/3, or to train on half and test accuracy on half.

As an additional evaluation tool, the output attribute for “percent confidence” could also be used to create a classification accuracy map. A contoured isopleth map can be created using interpolation techniques such as kriging.

The results of the accuracy evaluation on our learning scheme are presented below. As stated in the description of the training data, the training set coverage consisted of 1,511 points and after WEKA removed outliers, classified on 1,429 instances.

Displayed are the summary statistics, accuracy by class, and classification error matrix as a confusion matrix table. In the summary statistics we see the overall percentage accuracy of over 78%, and Kappa statistic. Kappa statistic is an indication of how much better than random are the overall results – in this case, 75% better than random. Note that at the top of this table, WEKA lists 24 attributes, not 23, but the last one is the vegetation label itself (in this classification run, it happened to be called “Group6” as it was part of the sixth major re-shuffling of *Alliances* in the vegetation groups).

Scheme: weka.classifiers.meta.Vote -B
Instances: 1429
Attributes: 24
MEAN1
MEAN2
MEAN3
MEAN4
STDDEV4
MEANASP
MEANDEM
BRIGHTNESS
MAXDIFF
MEANNDVI
STDDEVNDVI
MEANSLOPE
PROXSTREAM
PROXBODY
RATIO1
RATIO2
RATIO3
RATIO4
MEANDIFFNEIGHB4
STDDEVDEM
MEANDIFFNEIGHBDE
RATIONDVI
MEANDIFFNEIGHBND
GROUP6

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Vote combines the probability distributions of these base learners:
weka.classifiers.trees.DecisionStump
weka.classifiers.rules.DecisionTable -X 1 -S 5
weka.classifiers.trees.RandomForest -I 10 -K 0 -S 1
weka.classifiers.trees.J48 -C 0.25 -M 2
weka.classifiers.lazy.IBk -K 5 -W 0
weka.classifiers.lazy.IBk -K 10 -W 0
weka.classifiers.trees.REPTree -M 2 -V 0.0010 -N 3 -S 1 -L -1

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1118	78.2365 %
Incorrectly Classified Instances	311	21.7635 %
Kappa statistic	0.7487	
Mean absolute error	0.0553	
Root mean squared error	0.1508	
Relative absolute error	44.5788 %	
Root relative squared error	60.5871 %	
Total Number of Instances	1429	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.553	0.043	0.548	0.553	0.551	sclero-tree
0.714	0.029	0.664	0.714	0.688	dougfir-redw
0.597	0.014	0.708	0.597	0.648	bishop-montr
0.871	0.001	0.964	0.871	0.915	chaparral
0.718	0.029	0.701	0.718	0.709	willow-alder
0.813	0.065	0.786	0.813	0.799	coyotebrush
0.766	0.039	0.728	0.766	0.746	grass-marsh
0.78	0.005	0.848	0.78	0.813	saltmarsh
0.901	0.016	0.932	0.901	0.917	grass-upland
0.75	0.007	0.73	0.75	0.74	duneveg
0.955	0.001	0.977	0.955	0.966	nonveg
0.833	0	1	0.833	0.909	nonveg2
0.824	0.001	0.933	0.824	0.875	cloud

=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	j	k	l	m		<-- Classified as
68	19	7	0	16	11	2	0	0	0	0	0	0	0	123	a = sclero-tree
25	75	2	0	3	0	0	0	0	0	0	0	0	0	105	b = dougfir-redw
8	16	46	1	1	4	0	0	0	0	0	1	0	0	77	c = bishop-montr
1	0	1	27	0	2	0	0	0	0	0	0	0	0	31	d = chaparral
8	2	2	0	89	7	14	2	0	0	0	0	0	0	124	e = willow-alder
12	1	5	0	5	265	22	1	12	3	0	0	0	0	326	f = coyotebrush
2	0	0	0	13	16	131	2	1	6	0	0	0	0	171	g = grass-marsh
0	0	0	0	0	1	9	39	0	1	0	0	0	0	50	h = saltmarsh
0	0	0	0	0	25	2	0	247	0	0	0	0	0	274	i = grass-upland
0	0	0	0	0	6	0	1	2	27	0	0	0	0	36	j = duneveg
0	0	2	0	0	0	0	1	0	0	85	0	1	0	89	k = nonveg
0	0	0	0	0	0	0	0	1	0	0	5	0	0	6	l = nonveg2
0	0	0	0	0	0	0	0	2	0	1	0	14	0	17	m = cloud
124	113	65	28	127	337	180	46	265	37	87	5	15	1429		TOTAL

The column for “TP Rate” above lists Producer Accuracy for each of the classes. The column for “Precision” lists User Accuracy. For an example on reading the confusion matrix, for the Upland Grass class, 247 points out of 274 total Upland Grass points were classified correctly (90.1% Producer Accuracy). 265 points were classified as Upland Grass (totals across the bottom of matrix), of which 247 are correct (93.2% User Accuracy). Also note that from Table 2, the training set included 288 points for *grass-upland*, but 14 were removed during pre-processing by WEKA as outliers, leaving 274 points to train the classification.

Discussion

Not surprisingly, the tree classes have the lowest accuracy as they tend to be confused with each other. It was evident when looking at the imagery that many patches are mixed species, not always pure stands of trees.

Early classification problems of shrub, dune, and marsh types being over-classified and grass being under-classified were vastly improved primarily due to 4 factors: choosing

the best scheme for how to combine the *Alliances* with given training data, finding which WEKA algorithms worked the best, thorough analysis of all the training points, and addition of a large number of grass training points. Overall accuracy from initial classifications improved 20-25% and individual classes as much as 40-50% for some of the weaker ones. Additional ground-truth data for some species could potentially increase accuracy further, or could allow for different or multiple grouping schemes of the *Alliance* species.

Accuracy of the classification might be improved if each of the specific points which ended up classifying incorrectly, as seen in the confusion matrix, were analyzed again. But if looking at these again simply on a visual basis, there is no guarantee that they can positively be judged for whether to delete or move these points in comparison to the other training data. Some vegetation types are simply difficult to tell apart via satellite view. In addition, sometimes a final product that maps the whole land area can turn out stronger if not too many outliers are removed and some degree of variation in training plots are kept.

Overall accuracy and Kappa Statistic are slightly better than previous PRNS mapping from photo-interpretive methods, despite the fact that the ground truth data was not collected with the imagery in mind. However, the big advantage to this method over photo interpretation is the efficiency and repeatability of the computer model. One could supplement the existing ground truth data, or substitute updated imagery, and run the model again relatively quickly for the same area or a slightly different area.