Final Report:

Land Cover and Vegetation Mapping In Pinnacles National Monument

2005

Rocky Mountains Cooperative Ecosystem Studies Unit (RM-CESU) - In fulfillment of Cooperative Agreements 1200-99-007, UMT-57 and UMT-75; and Cooperative Agreement H1200040001, UMT-06.

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

Roland Redmond, Director
Jim Schumacher
Will Gustafson

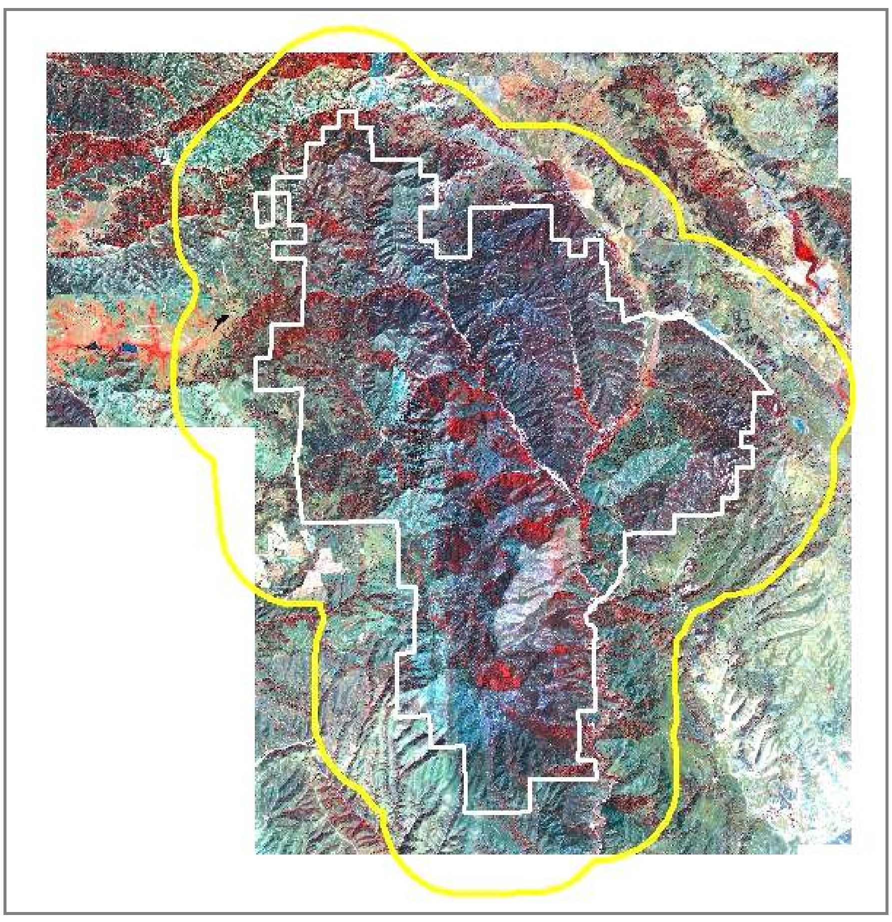
December 8, 2005

INTRODUCTION, IMAGERY, SOFTWARE

For use in resource management and long-term monitoring by scientists and planners at Pinnacles National Monument, California, the Wildlife Spatial Analysis Lab at the University of Montana has produced a digital database of existing vegetation and land cover from high-resolution Ikonos satellite imagery. Pinnacles National Monument (referred to as PINN throughout this report) had significant acreage added in expansion of November 2000, and this new mapping provides additional resource information to cover the full area plus an adjacent buffer zone.

Two Ikonos images were assembled into a seamless layer for classification. Figure 1 below shows the Ikonos imagery with the park boundary as a white line. This park boundary encompasses 10,950 hectares (27,059 acres). An additional buffer zone of 2 kilometers surrounding the park was also classified. The yellow line shows the full area that was classified and included in the digital database, and this area is then approximately double in size at 22,239 ha (54,954 acres). The white background not covered by imagery on three edges, as can be seen in the figure, is not included in that number: we have classified roughly 55,000 acres.

Figure 1: Ikonos imagery displaying r-g-b color values as image bands 4-3-1, with National Monument and classified area boundaries. The top of map points due north.



The Ikonos imagery and resulting classification grid are 4-meter resolution. The imagery from the Ikonos-2 satellite sensor is 4-band multispectral: 3 visible bands plus near-infrared. (The panchromatic black&white 1-meter resolution band was used in visual analysis but not in the computer classification.) The two images are dated 05-03-2000 and 06-05-2000. There are some band differences between the two images due to the acquisition dates being one month apart during the end of the wet season and/or plant phenology. Thus, the mosaic of the two images has some differences across the project area but is deemed okay for classification with good distribution of training samples across the whole study area in both images. (See distribution in Figure 2 below.)

Preparation of the imagery and some visual analyses were done in Erdas Imagine software. Image segmentation was output from eCognition software (using parameters of Scale 17.5, Shape Factor .2, Smoothness .8, Compactness .2). The layers input for segmentation in eCognition included: 3 spectral bands (red, green, near-infrared; blue band was not used); 2 of the 3 Principle Components Analysis (PCA) layers (PCA-3 not used); and layer for NDVI – Normalized Difference Vegetation Index.

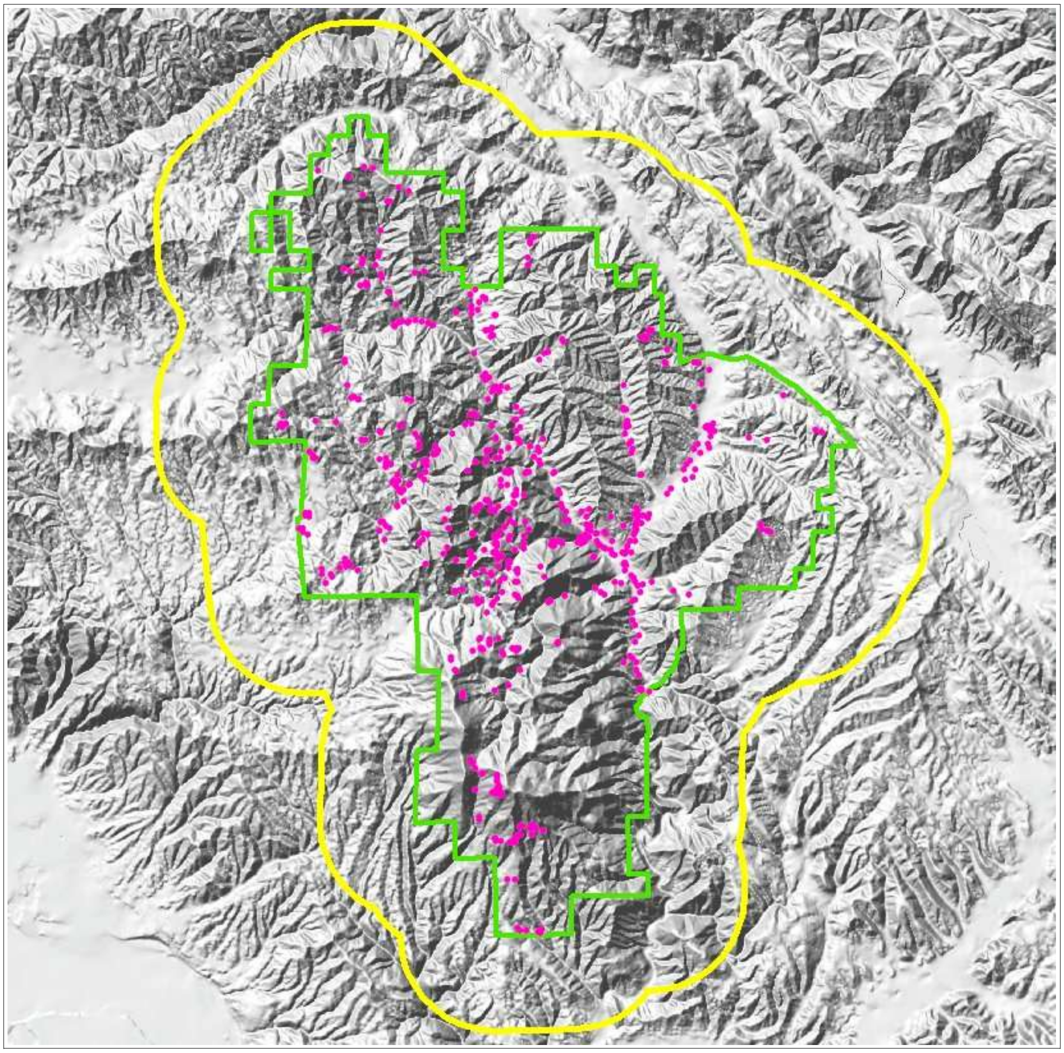
Each segmented region (raster polygon area) was attributed for spectral and topographic statistics. After converting output to ESRI ArcInfo software grid format, an attribute was also added separately in ArcInfo for proximity to water, to aid in classifying vegetation types adjacent to streams. Training data were examined and coded in ESRI ArcGIS.

Classification, as described below, is accomplished using the data-mining software WEKA. It utilizes multiple machine-learning algorithms, rather than a more traditional single "Nearest Neighbor" approach. Classification output was visually checked in ArcGIS. Canopy cover is classified from NDVI in ArcGIS as described below, and attached as an attribute in the database.

TRAINING DATA

Data from field plots ("Releve" data) were taken from the most recent spreadsheet file provided by PINN. These point data were converted into an ArcInfo coverage consisting of 591 points labeled with attributes for vegetation *Group, Alliance, Association,* plot shape and size, along with their original Releve plot identification number and locational x,y coordinates. 2 plots labeled by PINN as "Bad point" and "Garbage" were immediately coded for exclusion: although they did contain attribute for veg *Association*, without knowing why PINN had labeled them as bad, such as an incorrect location, we exclude them. Frequency tables were built to verify numbers of points per vegetation types to assist with comparison of target list of cover types to be mapped.

Figure 2: Distribution of training plots over hillshade background.



Analysis of Training Points with Segmented Regions

After initial trial classifications, each of the 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 1-meter black&white panchromatic image as background display helped confirm the validity of plot location and suitability. Also somewhat useful was older PINN classified polygons from photo-interpretive classification. Photography of all the vegetation types, from the University of California Digital Library, was also helpful for double-checking basic plant form, as were our photos from visit to Pinnacles.

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 either judged as not meant for that particular patch of vegetation, or simply a duplicate; or if very near the edge, move one point to adjacent similar polygon, as described below. As our polygons are small in size, this occurred for a small number of points.
2. Points on the very edge of regions/polygons where it could not be positively determined which vegetation patch was intended.
3. Points simply deemed poor training examples in comparison to the majority, after gaining expertise in identifying vegetation in the imagery.
4. Plot is more than likely not located in intended vegetation, with plot placed on the edge of a patch instead of actually in it. Field crew's attribute for plot shape and size perhaps is meant to cover these cases (or they figured it would be obvious during mapping). Some of these are plots lying in a shadow polygon on the image rather than the vegetation polygon (typically a tree). The field collector may have simply been standing beside the tree rather than under the tree, and plot dimension noted. Where obvious, many of these points were then manually moved into the veg polygon as described below.
5. Percent cover of the intended vegetation is judged too low overall for that particular segmented region/polygon to be a suitable training plot. For example, a plot that is coded for one vegetation type but appears to be in a particular polygon shaped so that for its entirety it is not very densely vegetated, and is likely to be misclassified and confuse the good training data.

It is important to note that the PINN field crew was not placing plots based on our delineated regions, as the field data were collected earlier. So plots are not placed perfectly in the center of visually appropriate polygons. Data also may not have been 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, which are computer generated polygons of a certain size and shape. In the field, the Releve plots are sometimes not good samples of our polygons but that is due more to how the polygons split up the vegetation patches rather than the field crew's collection methods.

Sometimes the problems above reflect drawbacks of the image segmentation process: occasionally a region/polygon might be an odd shape. Overall the polygons appear to fit the landscape viewable on the imagery very well, but when we're dealing with over 100,000 polygons, a small percentage may be an odd fit – not how one would draw the shape if one could manually draw every single polygon.

Figures 3, 4, and 5 show an example of how computer-derived segmentation splits up the imagery. Shown is a zoom-in detail of the project area near the westside Chaparral Ranger Station/Picnic Area/Trailhead. If we compare these labeled regions with earlier polygon mapping from photo-interpretive methods, the computer-derived regions are a much finer detail than the generalized polygons.

Figure 3: Detail of Ikonos Imagery with Segmented Regions (Image bands 4-3-1).

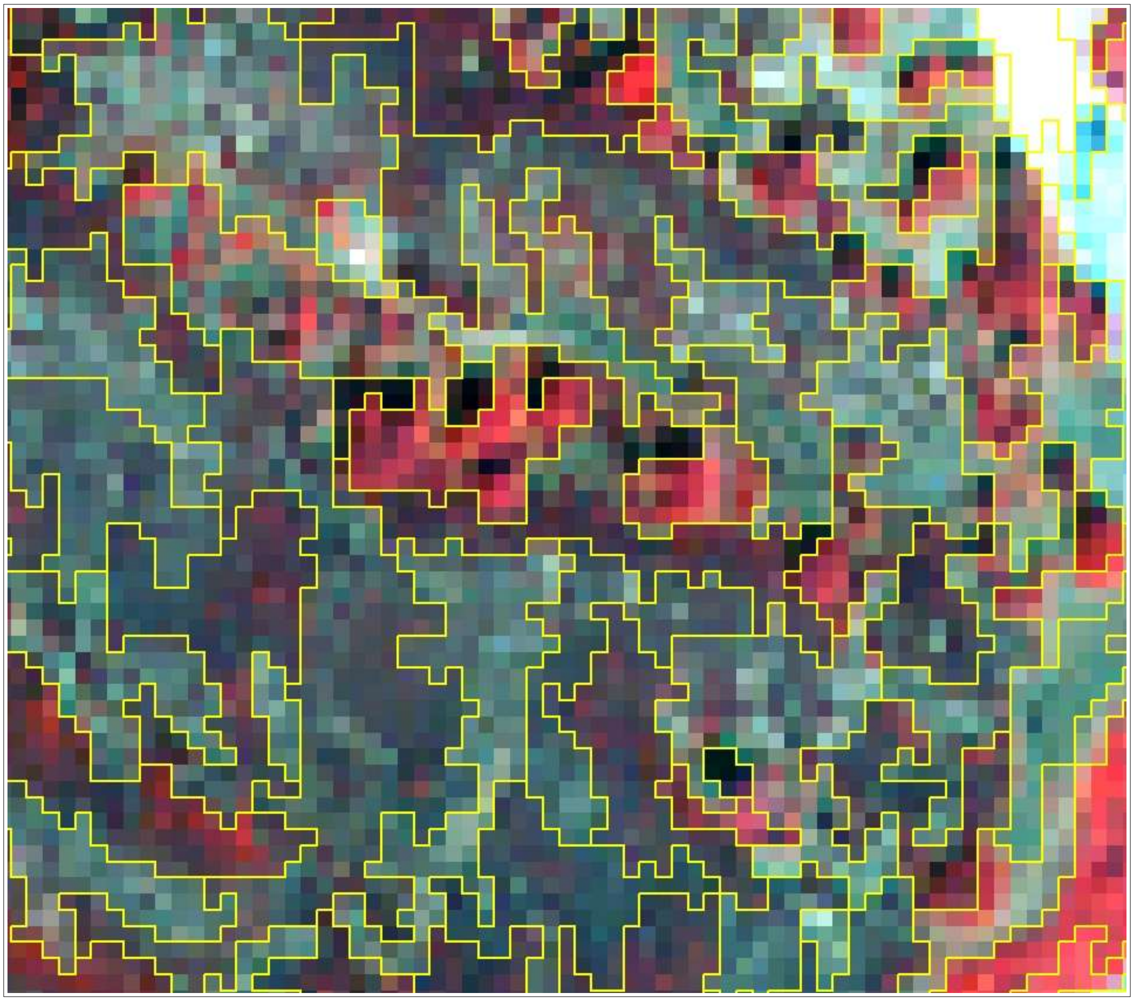
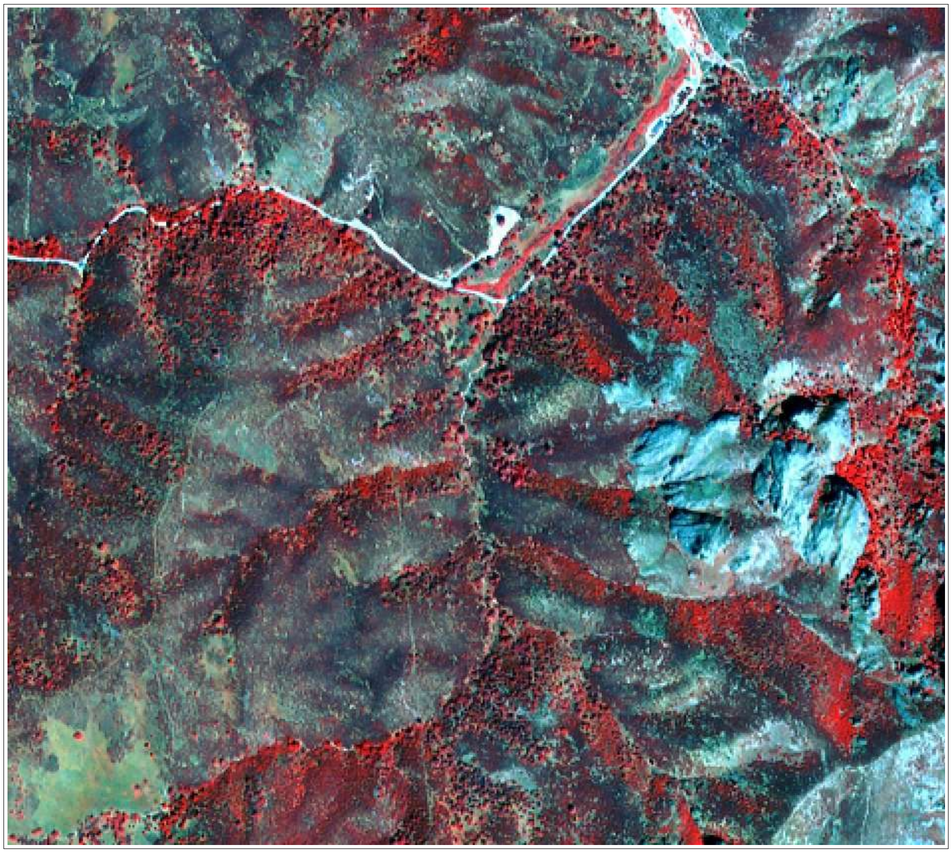
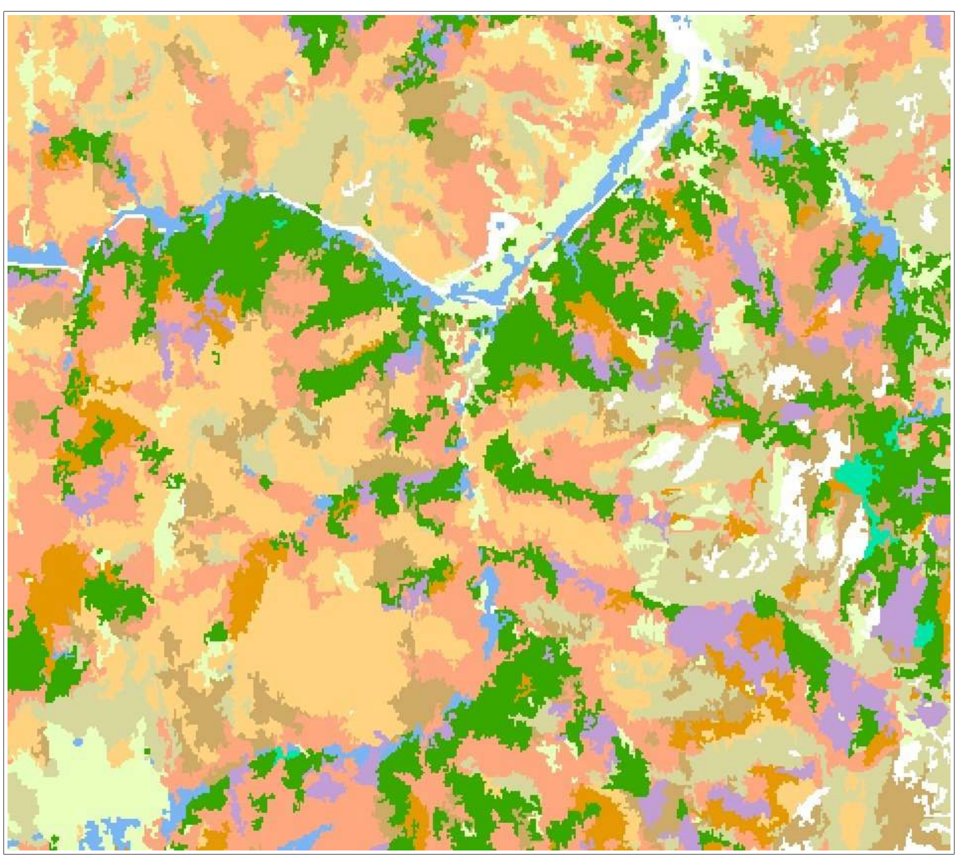


Figure 3 is a closeup detail from the top of the next two figures. By displaying the regions as yellow polygons atop the imagery, we see the results of segmentation with individual 4-meter pixels aggregated into regions representing vegetation patches. Each of these regions is attributed with all the image statistics and topographic variables and run through the classification program.

Figures 4 & 5 on the next page are zoomed out to a slightly bigger area of the west side. Park features that are recognizable: the west entrance road as it approaches Chaparral Ranger Station at top; round overflow parking area as barren white color, upper middle; rock formations of Resurrection Wall at right; Juniper Canyon on the far right with the "red" vegetation; and a grassy field just north of the West Entrance at lower left. Figure 4 again shows the Ikonos imagery in a color band combination that makes lusher green vegetation stand out here as brighter red. You can see individual trees as red, with their shadow. Figure 5 then shows polygon regions color-coded for land cover type. Looking at them next to each other, it is an example of how the segmented regions are classified.

Figures 4 & 5: Ikonos Imagery and Coded Land Cover for a Comparison Area.





Manual Editing & Moving of Training Points

Where possible during inspection, problematic 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 (sometimes exactly on the edge of the 4-meter pixels), or to move a plot out of a shadow. Points were usually moved only a few meters – less than or equal one pixel, a single 4-meter cell – and never more than a couple cells, in order to prevent mistakenly moving it into an incorrect patch, even if the field crew stated that the plot represented a large patch. Plot sizes from the field data are often as large as 1000 square meters, whereas our delineated polygons are based on groupings of 4 meter pixels, so the original point can end up in an incorrect adjacent polygon than intended. For example, a tree point is located in an obviously grassy area <u>between</u> the trees in a large field plot, so we move the point to the tree polygon. Or the field crew was standing in the <u>shadow</u> of the tree, in grass or shrub or barren, and with a large plot size noted; so the point is moved to the tree polygon. On the other hand, training plots in shadows (either woodland or topographic shadow) sometimes could not be moved if they're too far from edge of region and then might be incorrect if moved; so then these are coded as unsuitable for the training set.

Sometimes when more than one plot occurred within a polygon, one of them was moved to the adjacent polygon, but only if it was close enough to an edge of what also appeared to be intended sample vegetation and the Releve plot dimension included it. Otherwise it was coded for exclusion. Field plot size and shape helped guide this.

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. 99 points were moved and 154 points were coded as unsuitable by the final classification. Of the 154, 15 of those were cases where multiple plots occurred in a polygon and were unable to move and thus invalidated. The remainder were judged either poor training examples or an under-represented type that would confuse the other types. As mentioned already, bad training samples could be points lying in shadows but too far away from the edge of polygon to be moved manually with confidence. Specific vegetation types were under-represented and omitted, as described below: if allowed to remain, they do not introduce sufficient meaningful training information and confuse the remaining data and bring down accuracy. It is important to have a minimum number of training samples for each cover type.

By the final classification, some plots had visually been analyzed numerous times over. For difficult classes, grouping of vegetation types in slightly different ways, while aiming for the desired target classes, required tweaking of the training data by re-coding or omitting or putting back in, and re-running multiple times. Notes on specific problems with the classes are elaborated below.

Further evaluation of the training set was provided by the classification software WEKA. It 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.

LAND COVER CLASSES

We attempted to group vegetation types to follow a target list of cover types to be mapped after earlier meeting with PINN. The final classification includes 11 cover types: 6 Chaparral classes, 3 Woodland/Riparian, Herbaceous/Grass, and Sparse/Non-veg. Below is a description of each class with how and why the groups were combined. Chaparral is said to cover approximately 80% of Pinnacles and thus has a larger number of classes.

Field data were coded by PINN with a *Group* label, which are groupings of vegetation *Alliances*. *Alliance* and *Association* are part of the classification hierarchy of the U.S. National Vegetation Classification System (USNVCS), being followed nation-wide by the National Park Service. Field data contain a far greater quantity of *Groups* than are intended to be classified successfully from remote sensing methods, and so they have been re-grouped further towards a lesser number of targeted cover types. These classes could be considered a mid-level classification of vegetation types that share similar ecological habitat.

Many *Groups* were severely under-represented in the training set, with some having only a few plots, probably due to their rarity or knowing that our classification would re-group them. Because of this, initial classification rounds were done first on a very broad scheme and then later rounds refined to aggregate *Groups* into different classes. At the broadest level, we classified basic life-form as just 4 classes: Woodland/Riparian, Chaparral, Herbaceous/Grass, and Sparse/Non-veg. Final classification at this life-form level is at a very high accuracy (95% -- see the Evaluation section below), as it does not have trouble distinguishing between those 4 types. We are including in the database for PINN this 4-class classification, and also delivering a simple 4-class covertype grid along with an 11-class covertype grid. This is in addition to the large database grid containing all of the associated attributes, which can be somewhat unwieldy in use.

As you try to break out more and more classes from the field data, with lesser and lesser numbers of training plots per each class, accuracy drops proportionally and to a very poor result when trying for certain types with insufficient training plots and/or less readily distinguishable by the program. We repeatedly tried to classify certain groups without very good success, and finally settle on the 11 classes to reach the target goal of at least 80% accuracy, with our final classification of those 11 at 85% accuracy. Quite good via remote sensing. (See further notes in Results and Discussion sections below.)

Some of the field data were coded by PINN as mixed species, undoubtedly due to their growing patterns in the landscape. PINN's primarily chaparral vegetation is more heterogenous than some landscapes containing stands of pure single species. Thus we end up with mixed-species groups in the classification. For example, Chamise-Buckbrush or Chamise-Manzanita, as there were no Manzanita single-species plots.

PINN's targeted list of cover types to be mapped was originally somewhat greater than the final 11: approximately 20 types, including a few rare types that might need to be mapped manually rather than with the program. It included some classes as both separate vegetation species and a mixed class, whereas the field data later was primarily mixed species with insufficient quantity of training plots for single species mapping. For example, initial desire for separate classes for both Buckbrush and Chamise-Buckbrush; whereas we end up with just one Chamise-Buckbrush class.

The original list had a few veg species as potentially being mapped manually, rather than thru the computer classification, but this would be quite difficult on this imagery. We ended up with a single mixed Riparian class and Cottonwood and Sycamore points were included in that; Buckeye we did end up mapping on its own as one of the 11 computer-derived classes; Foothill Ash (with no training points provided) ended up as association with Prunus/Mixed Chaparral; and Calif.Sagebrush did well classified with Buckwheat.

For the final classification, each training example was cast into one of the following classes based on its *Group* code, with *Alliance* or *Association* used as secondary information for some points. Table 1 shows the number of points per class in the final training set. These are the basic class names: see more detailed descriptions below.

Table 1: Frequency of training points with basic class name.

# POINTS	CLASS NAME	
29	c1	Chamise
39	c2	Chamise-Ceanothus
24	c3	Chamise-Manzanita / Mtn.Mahogany
36	c4	Prunus / Mixed Chaparral
40	c5	Buckwheat / Calif.Sage
27	c6	Chamise-Black Sage
44	g	Herbaceous / Grass
59	n	Sparse / Non-veg
62	w1	Oak / Pine
12	w2	Buckeye
66	w3	Riparian Mix
-----		
437 total points		

Classes were given a short character name, such as C1 or W1, for ease of use as a field in the database. The database also includes attribute for a numeric code for each class (see the database metadata).

When describing how points were grouped below, grouping success also depended on carefully inspecting each point, sometimes multiple times, for how it looked as training sample in the imagery. Does it look like a good patch as training sample, and if getting incorrectly classified, does that patch look confusingly like something else? In the following descriptions, when description refers to one type of veg point getting "confused" with another, it is referring to the fact that cross-validation evaluation in WEKA outputs a "Confusion Matrix" which shows which class any incorrectly classed points fell in. As an example of looking at the confusion matrix for the Prunus class: 28

points classified correctly; 4 points classified as W1-Woodland (which is not surprising, as Prunus is somewhat of a woody shrub that could appear as tree-like in remote sensing imagery and confusing to the program); 1 point as Buckeye; 1 point each as C2 and C3 (also not surprising, as these are more woody shrubs); 1 point tossed out by WEKA as an "outlier", for a total of 36; and no points confused with C1-Chamise. Thus, overall for this class, 80% accurate – not too bad. See the final confusion matrix in the evaluation section of this report, with further example of reading it. Description of the classes follows:

C1: Chamise
Although Chamise is a predominant species at Pinnacles, much of it is in the mixed species following.

C2: Chamise-Ceanothus
This is the Chamise-Buckbrush mix. Included mostly mixed-species Adenostoma-Ceanothus plots, with some Ceanothus plots. We attempted classification of Ceanothus alone, with poor results, because there were insufficient Ceanothus plots and/or they get too confused by the program with Chamise, even if we toss out the large number of mixed-species Chamise-Ceanothus plots. So rather than omit all those as useful training samples, we did not break out Ceanothus alone..

C3: Chamise-Manzanita/Mtn.Mahogany
This is the Arctostaphylos-Adenostoma spectrum, with PINN *Group* definition that it did or did not include significant amounts of Adenostoma (no single-species Arctostaphylos points provided); along with Cercocarpus (Mountain Mahogany). Mtn.Mahogany was a difficult type to place. We attempted to classify it on its own but with very poor result (about 50% accurate), perhaps because not enough plots were provided. When grouped with a mixed Chamise class, it lowered accuracy significantly. Rather than dropping all these plots, it fit best with this class based on how it looks in the imagery and on the accuracy results.

C4: Prunus/Mixed Chaparral
This is the Holly-leaf Cherry class ("anything dominated by Prunus"), along with PINN's "Mixed Chaparral". It includes both of those plot types along with the mixture Prunus-Mixed Chaparral. The program distinguished Prunus quite well right from the start. We first placed Mixed Chaparral plots with mixed chamise/sage/etc types (and tossed out the mixed-species *Group* as confusing), with poor results: "Mixed Chaparral" does not belong with those -- this type is <u>not</u> a mixed Chamise sort of grouping. PINN's definition states it is separate from Prunus by not having significant amounts of Prunus, but on the imagery it was obviously related in the landscape. These are north-slope species "not fitting in other *Alliances*". In the imagery, Mixed Chaparral appears <u>very</u> similar to Prunus and obviously does not fit with Chamise types. The *Alliance* is often Rhamnus. Plots with Fraxinus (Foothill Ash) in *Association* are included in Prunus or Mixed Chaparral *Groups;* thus this class includes Foothill Ash.
Juniper. This class also includes 2 points of the Juniperus Group: 2 which had a Prunus *Association*. Juniperus (Calif.Juniper – a scrub-type juniper) was difficult to

class, either due to insufficient points and/or its sometimes smaller size or scattered growth pattern. We were provided a small number of Juniper points and attempted to classify it by itself without success – none of the points classified correctly, 0%. In the imagery the points appeared in patches of a scattered nature, or else our polygons simply did not capture them well. It was getting confused with both the Oak woodlands and with other shrubs. We tried lumping it in a Pine/Juniper class, lumping the conifers together, but was also far too low an accuracy. It was getting most confused with either Oaks or Manzanita (not with Chamise), and as it is a scrub-type plant, we tried going ahead and putting it with the Manzanita class (not with a mixed Chamise/Scrub class). But then this class started getting confused with woodlands, which it wasn't at all before; and likewise the Woodland class was now getting more confused with shrubs. As training data, that scheme brought both classes down. In the end, for the best success, we split up the Juniper points: the 2 with Prunus *Association* were lumped here with Prunus class (and this class then actually improved in accuracy percentage); 2 points with a Quercus *Association* were lumped with the Oak woodlands class; and the other 7 points were coded as unsuitable.

C5: Buckwheat/Calif. Sagebrush

This is Eriogonum (Buckwheat types) along with Artemisia californica. In early rounds of classification, Eriogonum had only moderate success (in the 40-60% accuracy range). Cal.Sagebrush was very poor by itself (about 30%), and when included in a mixed Chamise-Sage chaparral type of class, was not too accurate there either. But the majority of Cal.Sage plots are mixed *Alliance* with Eriogonum and PINN stated "could possibly go in the Eriogonum Group". So when these two types were combined, accuracy for both increased, as the Eriogonum was also no longer getting confused with some other class. In fact, final classification accuracy for this class ended up <u>very</u> high at 94%.
Lupinus. This class also includes a single point labeled as Lupinus *Group* and having *Association* with Eriogonum. We tried to classify Lupine as a class of its own, but with insufficient good training points, it was extremely poor with only a single point classifying correctly. Or else because even with high-res imagery, pixels that are 13 feet square are not fine enough to capture scattered Lupine; but those polygons were also not of such low-density to put them with the Sparse *Group* either. For the one Lupine point with Eriogonum *Association*, we lumped it here and coded as unsuitable all the others.

C6: Chamise-Black Sage

Includes Salvia (Black Sage) and mixed Salvia-Adenostoma plots. We attempted to classify Salvia alone (with only about 50% accuracy at first), but rather than toss out those mixed-species plots, include them.

G: Grassland/Herbaceous

This is the Grassland forbs/herbaceous *Group*. It also includes the smallish number of points for Heterotheca (Golden Aster) *Group,* (as sometimes Grassland does have an *Association* with Heterotheca)*,* if those plots did not look extremely sparse.

N: Sparse/Non-veg

This class includes the Sparse Group (<10% cover Alliances, Rock, Scree, Lichen, etc),

along with other plots that were too sparsely vegetated to classify with remote sensing as other than Sparse. These included some of the points for Baccharis salicifolia (Mule-fat), Artemisia dracunculus (Wormwood), Mimulus (Sticky Monkeyflower), Selaginella (Spikemoss), and Herbaceous Streambed. Some of these with sufficient quantity of plots, such as Selaginella and Mule-fat, we first attempted to classify on their own, but without success. When inspected, many of these points are stream channel types appearing very sparse in the imagery. We lumped those points in this Sparse class or else disqualified the rest as unsuitable for the training set, as they are not good training data for other classes either. For Selaginella (Spikemoss), perhaps these little plants just grow too sparsely on rocks or barren areas to distinguish on 4-meter pixels. PINN's *Group* definition stated for Mule-fat “some of these may get confused with the Sparse Vegetation Group”. Likewise, PINN definition states Herbaceous Streambed points “could possibly be clumped into the Sparse Vegetation Group”. We first attempted to place Mule-fat with the Riparian class, where it lies ecologically, but it does not appear visually similar in the imagery and confused the Riparian class. Most Mule-fat plots looked very sparse and unlike other Riparian. Mimulus and Wormwood did not classify well when lumped with a mixed Chamise/Scrub class either, as they are quite different. Early rounds tended to over-classify this Sparse class, at a range from 8 to 13% of total area. Points were then inspected for polygon sparsity. Streambed types were often obviously low-density when looking at the 4-meter pixels (very light or white color in the imagery). If not, they were coded as unsuitable and not utilized at all. A substantial portion of the total 154 points deemed unsuitable as training data were of the Sparse and Selaginella *Groups*. Some of these points might be included in the other Grassland/Herbaceous class but are too dissimilar in appearance, and rather than try to create a second and somewhat vague Herbaceous class, we coded them as unsuitable.
Sedum. Sedum is one other Group not described elsewhere. We first tried lumping it here with the other low-density types. But the small number of points were inspected and found to be a total mish-mash of pixel variation, probably due to its growing location, and all were disqualified for training data.

W1: Oak/Pine

This is mixed Oak Woodlands and Gray Pine, and including Scrub Oak. We attempted over and over to break out Pine as a separate class. But with insufficient number of training points, or perhaps as it seems to often be interspersed with the Oaks (and in fact some of the plots were attributed with Quercus *Association*), ended up lumping them together. And PINN's *Group* definition for Pine did state “this group is highly questionable...”. Likewise, we attempted to classify both Blue Oak and Coast Live Oak on their own, but they were always confused with the other Oaks. Perhaps with additional training data or in a landscape of larger and less heterogenous stands, we could separate out one or more of these species. Remote sensing has trouble distinguishing one Oak from another in this landscape. Live Oak and Valley Oak (only a handful of points provided for Valley Oak) are sometimes said to be Riparian types but plots were mostly in mixed Woodlands and not located beside streams, and thus were not included in the Riparian class. For Scrub Oak, we tried classifying it either with the other Oaks or with shrublands. PINN's plant listings place it as shrubland ecologically, but from a remote sensing standpoint, it had far better accuracy when lumped with the other Oaks. On the imagery, it looks far more like Quercus family than it does shrubland. (In comparison, the also scrub-like Juniper was a tougher call, as described above.) This class did contain 2 points in its training set for Juniper with Oak *Association*, as described in discussion of Juniper above in the Prunus class.

W2: Buckeye

This is the Aesculus *Group*. It appeared to be growing on slopes where we hoped the program could distinguish it, although with moderate success with limited training data. It is probably the weakest of class results but in attempt to separate out as many tree species as possible, we leave it as a separate class. When we view the final classification mapping and compare to previous PINN mapping, it does not look at all unreasonable. Again, each one of these training plots was inspected multiple times over in hopes of improvement.

W3: Riparian Mix

This is a Woodland and mixed Riparian class containing the *Groups* Willow, Cottonwood, Sycamore, Mixed Riparian Woodland, Rosa, Coyotebrush, Juncus, and Seep/Spring. Some of those only had small numbers of training plots and thus are lumped together here. We attempted to classify Willow separately but it was too confused with the other Riparian points and less than 30% accurate. Likewise, we attempted to classify Rosa with Coyotebrush as a class of their own (Coyotebrush often has *Association* with Rosa). But they were getting confused with the Riparian points (or to a lesser extent with other shrubs), with all points located in Riparian corridors. They are Riparian types at Pinnacles, as the PINN Plant Checklist states. The couple plots for Seep/Spring could possibly be located basically beneath trees, but visually they appear like the surrounding trees and are included in this class. PINN definition stated Juncus plots “could possibly be clumped into the Grassland Group”, but when we tried that, Juncus had better accuracy when clumped here with the Riparians. On the imagery, Juncus plots looked like Riparian vegetation.

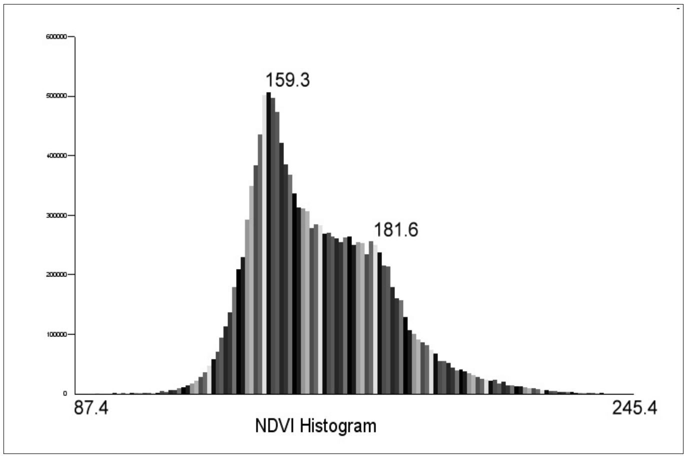
CANOPY COVER CLASSES

One of the project goals besides land cover type was a classification of canopy cover. Without no canopy data coming from the Releve field plots, the other method of accomplishing this attribute was to use NDVI and produce a histogram from that vegetation measure that is used to split canopy cover into low, medium, and high classes. This is produced from Mean-NDVI value for each region. The histogram (see Figure 3 below) did show 2 distinct points in the curve at which to break out 3 classes, so this was preferable to splitting the classes simply on plus and minus one-standard deviation breaks, for example. The range of values for Mean-NDVI was 87.4 to 245.4, with an overall mean of 170.0 and standard deviation 15.2. Our 3 canopy classes were split at these points:

87.4 – 159.3 = LOW CANOPY
159.4 – 181.6 = MEDIUM
181.7 – 245.4 = HIGH

This is a very basic measure of canopy coverage of the landscape. As there is no ground-reference data for canopy cover, there is no post-classification assessment of accuracy of this layer. Upon review, the class breaks could easily be modified: using the Mean-NDVI attribute, calculate the numeric canopycode attribute at different values. For ease of display in ArcGIS or other software, we also output a simpler canopy cover grid from the canopycode attribute.

Figure 3: NDVI Histogram for the 138,360 Regions.



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. (See more about WEKA below.) Imagery statistics are output from eCognition and attributed to each segmented region. Proximity to water features is separately attached, with source input being a streams file obtained from PINN that came from fine-scale digitizing of streams from Ikonos imagery. 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 Pinnacles area imagery and a somewhat different set might be chosen for a different geographic area.

The following list of 38 "best variables" were used in this classification. For brief definitions of each, see the metadata for the product database. But you can see listed by the numbers there are variables (attributed to each of the 138,360 regions) for each of the PCA layers, NDVI values, and each of the spectral bands (as mean value, standard deviation, ratio, and mean difference to neighbor); then "brightness" and "max difference", and then topographic attributes and the proximity to water attribute. Description of the proximity to water variable follows below.

Table 2: Variables used in the classification.

```
MEANPCA1
STDDEVPCA1
RATIOPCA1
MEANDIFFPCA1
MEANPCA2
STDDEVPCA2
RATIOPCA2
MEANDIFFPCA2
MEANPCA3
STDDEVPCA3
RATIOPCA3
MEANDIFFPCA3
MEANNDVI
STDDEVNDVI
RATIONDVI
MEANDIFFNDVI
MEAN1
STDDEV1
RATIO1
MEANDIFF1
MEAN2
STDDEV2
RATIO2
MEANDIFF2
MEAN3
STDDEV3
RATIO3
MEANDIFF3
MEAN4
STDDEV4
RATIO4
MEANDIFF4
BRIGHTNESS
MAXDIFF
DEMMEAN
SLOPEMEAN
ASPECTMEAN
PROXWATER
```

These variables were checked utilizing several attribute fitness tests in WEKA. WEKA has the capability of analyzing datasets with very large numbers 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, 4 other statistics variables from eCognition were ouput for the "shape" variables Lengthwidth, Compactness, Shapeindex, and Density. These were also input for WEKA but then 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) can increase 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 removing those 4 shape variables, and for checking removal of any other weakest variables. For instance, multiple evaluators also agreed that certain variables were weakest as far as providing useful information, such as the variables for Standard-dev-Band2 or Mean-Diff-Neighbor-NDVI or Mean-Diff-Neighbor-Band4. We tried removal of as many as 10 weakest attributes but results showed accuracy dropped: these variables were indeed adding useful information and thus these final 38 variables were kept. The most important measures for the overall classification, according to WEKA, were based on mean pixel values; however we feel the topographic and water-proximity variables are also quite useful for certain types, even if not in an overall index.

DEM-elevation, Slope, and Aspect are derived from a 10-meter resolution elevation dataset provided by PINN. They were resampled to 4-meter cell size to match the imagery and output grid.

The variable based on the water proximity measure was extremely useful in assigning labels to groups that occur near water. It is a focal sum of small radii from stream (8 meters, or about 26 feet from stream, thus ~ 50+ feet diameter stream zone). It is accomplished as a raster model by converting stream lines to raster cells and calculating focalsum for desired radius, then attaching result to database grid using a zonalstats function. It sums pixels within that radius and when displayed atop imagery, the 2-cell radius (8 meters), looks roughly the best. It appears to capture riparian vegetation and although probably misses a little and perhaps could be expanded to 3 cells, on the other hand, it doesn't overdo it so that it would be less meaningful in WEKA as a good variable.

We also assigned somewhat heavier “weighting” in WEKA to the variables for topography and water proximity, and this increased accuracy results. Believing that some veg types are dependent on these and in hopes of capturing north-slope or riparian types better, for example, we weighted Aspect and Water-Proximity heavier, and to a much lesser extent, Elevation (DEM mean) and Slope value.

CLASSIFICATION

Description

Machine learning algorithms have proved to be very successful in classifying landcover from remotely sensed data. For example, 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 38 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 is the "industry workhorse for off-the-shelf machine learning" and typically produces accurate results, so the ensemble includes one of these. (Ross Quinlan, "C4.5: Programs for Machine Learning", 1993, Morgan Kaufmann Publishers, San Mateo, CA.; Witten, I. and Frank, E., "Data Mining – Practical Machine Learning Tools and Techniques", 2000, Morgan Kaufmann Publishers.)

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. (Witten, I. and Frank, E., "Data Mining – Practical Machine Learning Tools and Techniques", 2000, Morgan Kaufmann Publishers.)

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. (Leo Breiman, "Random Forests", Machine Learning 45 (1):5-32, October 2001; Witten, I. and Frank, E., "Data Mining – Practical Machine Learning Tools and Techniques", 2000, Morgan Kaufmann Publishers.)

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. (Witten, I. and Frank, E., "Data Mining – Practical Machine Learning Tools and Techniques", 2000, Morgan Kaufmann Publishers.)

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. (Aha, D., and D. Kibler, "Instance-based learning algorithms", Machine Learning, 1991, vol.6, pp. 37-66; Witten, I. and Frank, E., "Data Mining – Practical Machine Learning Tools and Techniques", 2000, Morgan Kaufmann Publishers.)

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. (Kohavi R., "The Power of Decision Tables", in Proceedings of European Conference on Machine Learning, 1995, Springer-Verlab; Witten, I. and Frank, E., "Data Mining – Practical Machine Learning Tools and Techniques", 2000, Morgan Kaufmann Publishers.)

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. Each run utilizes the full ensemble of learners.

Overall accuracy as shown below was 84.9%, but if we do a doublecheck of our training plots after the classification, we find that actually 89.5% of the points end up in a grid region of that class (391 out of 437 total, with no outliers removed). However, accuracy of the program cannot be stated as 89.5% as that would be evaluating against its own training set, which would be expected to be higher than evaluating against a randomly held out 10%.

We also checked output accuracy of the training samples using a common percentage-split scheme, where a certain portion of points are held out once. 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.

The database attibute "percent confidence" is a measure output by the program itself for each of the regions as to how confident the program is that each region is that particular labeled covertype.

If an independent set of ground-reference data became available, it could be used in post-classification assessment of the thematic accuracy of this land cover product.

RESULTS

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 437 points and after WEKA removed outliers, classified on 392 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 nearly 85%, and Kappa statistic. Kappa statistic is an indication of how much better than random are the overall results – in this case, 83% better than random.

Table 3: WEKA Output Results, Confusion Matrix (continues next page).

```
Scheme:       weka.classifiers.meta.Vote -B
Instances:    392

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	**333**	**84.949 %**
Incorrectly Classified Instances	59	15.051 %
Kappa statistic	**0.8306**	
Mean absolute error	0.0614	
Root mean squared error	0.1541	
Relative absolute error	41.349 %	
Root relative squared error	56.5666 %	
Total Number of Instances	**392**	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	**Precision**	Recall	F-Measure	Class
0.76	0.014	0.792	0.76	0.776	c1
0.765	0.031	0.703	0.765	0.732	c2
0.625	0.005	0.833	0.625	0.714	c3
0.8	0.02	0.8	0.8	0.8	c4
0.943	0.008	0.917	0.943	0.93	c5
0.792	0.014	0.792	0.792	0.792	c6
0.929	0	1	0.929	0.963	g
1	0	1	1	1	n
0.808	0.038	0.764	0.808	0.785	w1
0.444	0.005	0.667	0.444	0.533	w2
0.885	0.033	0.831	0.885	0.857	w3

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k		<-- **classified as**
19	3	0	0	1	1	0	0	0	0	1	25	a = c1
2	**26**	1	1	0	1	0	0	3	0	0	34	b = c2
1	3	**10**	1	0	1	0	0	0	0	0	16	c = c3
0	1	1	**28**	0	0	0	0	4	1	0	35	d = c4
1	0	0	0	**33**	1	0	0	0	0	0	35	e = c5
1	3	0	0	1	**19**	0	0	0	0	0	24	f = c6
0	0	0	0	1	0	**39**	0	0	0	2	42	g = g
0	0	0	0	0	0	0	**59**	0	0	0	59	h = n
0	0	0	2	0	0	0	0	**42**	1	7	52	i = w1
0	0	0	3	0	0	0	0	1	**4**	1	9	j = w2
0	1	0	0	0	1	0	0	5	0	**54**	61	k = w3
24	37	12	35	36	24	39	59	55	6	65	**392**	**TOTAL**

The first column for “TP Rate” above lists “Producer Accuracy” for each of the classes: the number of correctly classified points (moving the decimal place gives you a percentage). The column for “Precision” lists “User Accuracy”.

For an example on reading the confusion matrix, reading the last horizontal line for Class W3 (Riparian), 54 points out of 61 total Riparian points were classified correctly (88.5% Producer Accuracy). Out of <u>all</u> the instances, 65 points were classified as Riparian (totals across the bottom of matrix), of which 54 are correct (83.1% User Accuracy). Also note that from Table 1, the training set included 66 points for Riparian, but 5 were removed during pre-processing by WEKA as outliers, leaving that total of 61 points to train the classification.

As another example, note that one class (N – Nonveg/Sparse) had 100% accuracy. All 59 training plots classified correctly, and no plots from other classes were mistakenly classified as Nonveg/Sparse.

Results for a simple 4-Class Lifeform

We also present the results for a classification of what is basically lifeform, but here labeled as the same four basic types – C,G,N,W – from the 11-class scheme (same lettering scheme). Not quite true lifeform, as in Tree-Shrub-Grass-Nonveg. W is mostly woodlands but includes some riparian shrubs; N is mostly non-vegetated but includes some "Sparse" types; and C contains the large area covered by Buckwheat, which might be more of an herbaceous life form, but PINN plant lists place it with the Chaparral. These four classes then are Chaparral, Grassland/Herbaceous, Non-veg/Sparse, and Woodlands/Riparian.

This output grid is not simply a recombining of the 11-Class grid, although there would be nothing wrong with that. It is a full separate classification for just 4 classes. It was done for comparison to see how well the program could distinguish between the simplest set of classes. Input is the same set of 437 training plots, with points relabeled based on the C,G,N,W scheme above. In this case, you see that out of the 437 points, WEKA removed as outliers only 13 and classified on 424 instances.

Table 4: 4-class covertype results, displaying similar WEKA output tables and confusion matrix as above (but without the highlighting):

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         403               95.0472 %
Incorrectly Classified Instances        21                4.9528 %
Kappa statistic                          0.9247
Mean absolute error                      0.0546
Root mean squared error                  0.1315
Relative absolute error                 20.5033 %
Root relative squared error             36.0905 %
Total Number of Instances              424
```

```
=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0.969     0.057      0.935     0.969     0.952    c
  0.811     0          1         0.811     0.896    g
  1         0          1         1         1        n
  0.94      0.028      0.94      0.94      0.94     w


=== Confusion Matrix ===

   a   b   c   d    <-- classified as
 188   0   0   6      a = c
   5  30   0   2      b = g
   0   0  59   0      c = n
   8   0   0 126      d = w
```

The overall accuracy is over 95%. Grassland is the lowest at 81%, but it's not very surprising that 5 of the points might get confused with Chaparral (see the Confusion Matrix), as perhaps some of the plots were representing polygons that are a mixture.

In reading the Confusion Matrix, again note that N-Nonveg/Sparse classified at 100%. Chaparral and Woodland/Riparian classified at 97 and 94%, with primarily just a few points confused with each other, not surprisingly.

Area Statistics for Both Classifications

Table 5: Classification by Percent Area.

		WITHIN PINNACLES	FULL BUFFER AREA
C1	Chamise	15.4	10.6
C2	Chamise-Ceanothus	20.9	15.0
C3	Chamise-Manzanita/Mtn.Mahogany	5.6	3.7
C4	Prunus/Mixed Chaparral	4.7	2.9
C5	Buckwheat/Calif.Sage	14.6	17.9
C6	Chamise-Black Sage	18.5	14.5
W1	Oak/Pine	8.6	9.9
W2	Buckeye	0.2	0.2
W3	Riparian Mix	2.0	2.4
G	Grassland/Herbaceous	5.8	17.4
N	Sparse/Non-veg	3.7	5.4
		-----	----
		100.0	99.9

Shown is area by percent of each class within Pinnacles N.M., and area for the full buffered area for grid/database being delivered, which is approximately twice as large an area, as noted in Introduction. Area was calculated for Pinnacles by creating a second grid with a “mask” within the full area.

Note that there is a far larger proportion of Grassland <u>outside</u> of Pinnacles: pastures that can be seen in the imagery. Also perhaps notable, the other Chaparral types outside of Pinnacles then drop in proportion, <u>except</u> for Buckwheat/Cal.Sage, perhaps as it gets mixed in with pasture lands.

As a subtotal, adding the 4 major types <u>within</u> Pinnacles, C-W-G-N, from Table 5 above, for the 11-Class grid:

```
Chamise                  = 79.8%
Woodland/Riparian        = 10.7
Grassland/Herbaceous     =  5.8
Non-veg/Sparse           =  3.7
                           -----
                           100.0
```

These amounts jibe very well with earlier estimates from PINN of its land cover. Differences could also be explained by the fact that some polygons with a training plot in them are not the best polygon examples and might be mixes of veg type and density.

Percent amounts from the 4-Class grid (which was estimated at 95% accurate):

Table 6: Area by Percent for 4-type Classification.

		WITHIN PINNACLES	FULL BUFFER AREA
C	Chaparral	84.6	75.6
W	Woodland/Riparian	10.0	12.2
G	Grassland/Herbaceous	1.9	5.8
N	Sparse/Non-veg	3.5	6.4
		-----	-----
		100.0	100.0

These amounts differ somewhat from the subtotals above. As described earlier, these were two separate classifications. This one is stated 95% accurate, and it should be easier to distinguish just these 4 classes from each other with accuracy. Looking at the classified grids atop the imagery, we like the 11-class grid, and perhaps the actual percentage amounts on the ground are in between these 2 sets.

Discussion

It was evident when looking at the imagery that many patches are mixed species, not always nice pure stands of shrubs or trees. Of the 11 classes, Buckeye shows the weakest result but was originally not thought to be computer classified at all and instead be manually mapped; and as it was classified from a small number of training plots, it could be more successful with additional training. For its small number of points, Buckeye shows confusion (see the Confusion Matrix) with the tree types and especially with Prunus class, which is not surprising for its habitat. Likewise, Prunus shows confusion with the Woodland types.

Woodland and Riparian classes show some confusion with each other, which is not surprising as they are both primarily tree cover. The Chaparral classes other than Prunus show confusion almost solely only with each other, which is also not surprising, as the training data and habitat are often a Chamise mixture and look similar on the imagery.

Overall accuracy from initial classifications improved 20-25% and for 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. But in particular, it could allow for breaking out additional classes, or different grouping schemes of the *Groups* or *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 can turn out stronger if not too many outliers are removed and some degree of variation in training plots are kept.

Overall accuracy for this number of classes is quite good despite the fact that the ground truth data was not collected while viewing the segmented polygons. 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.