Comparison of Landscape Patterns Delineated by Automated Versus Manual Techniques

Final Report
Cooperative Agreement Number: CA 1200-99-007
Task Agreement: UMT-60

Submitted to
U.S. National Park Service
Point Reyes National Seashore

Willard A. Gustafson
J. Chris Winne
Roland L. Redmond
Melissa M. Hart

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

22 June 2005

Introduction

Land managers have long been devising ways to visualize and quantify patterns inherent in the landscapes that they manage. By investigating these patterns, land managers can glean information about vegetation distributions and abundance, thereby enabling them to study plant, wildlife, and human interactions, and in turn to make better decisions related to competing interests or uses of natural resources.

Methods for visualizing and quantifying landscape patterns have changed dramatically during the last few decades, starting with manual interpretation of aerial photos, ranging through the beginnings of automated analysis of satellite imagery, and now to object-oriented analysis of high resolution satellite imagery (c.f. Mitri and Gitas 2004, Wang et al. 2004). Because the human mind is well adapted to visual pattern recognition, we often see patterns that computers overlook, and as a result, we have often been dissatisfied with the results of standard image analysis procedures. But with advances in remote sensing and computer science technologies, automated methods are improving in their ability to delineate patterns from satellite imagery. Their outputs represent ground conditions in ways that are not only more realistic and pleasing to the human eye, but are statistically sound as well.

For land cover classification, the goal of automated pattern delineation, or image segmentation, should be to produce spatial units (regions) that conform not only to physical patches on the ground but also to the scale of the available training data. For example, one might not want to classify a segmentation with an average region size of 100 ha with training data that represented 0.04 ha plots because the vegetation measured in the plot may not be representative of the larger region in which it falls. Thus the statistics collected from these training data would not be representative of the stand and would likely lead to misclassification errors. By matching the region sizes with the scale of the training data, the statistics generated are more representative of the location where the data were collected, and thus should produce higher classification accuracies.

The objective of this study was to determine how well a photo-interpreted segmentation could be mimicked by automated methods. We specifically sought to evaluate and compare segmentations derived from two types of satellite imagery, Landsat TM-7 and Ikonos-2, and two software packages, M86 (Barsness 1996, Ford et al. 1997) and eCognition (2001). These automated methods were designed to emulate a set of map units that were manually delineated through air-photo interpretation and according to standardized techniques used by the National Park Service (see Moritsch et al. unpubl). Thus, rather than trying to produce an optimal segmentation of map units to be labeled by image classification, our goal was to mimic or improve on the photo-interpretive process. This required special attention to match the scale of automated patch delineation with that of the photo-interpreted polygons, while avoiding over-segmentation of large, homogeneous landscape units, such as grassland or forest patches.

Study Area

The study area follows the general administrative boundary of the Point Reyes National Seashore (PRNS), north of San Francisco, California (Figure 1). The area covers approximately 40,000 ha (98,775 ac) west of the Pacific Coast Highway and between the towns of Dillon Beach and Stinson Beach. Lying along the eastern edge of the Pacific plate, and extending from sea level to above 1000 m, the area is complex in terms of its geology, flora, and fauna. Any effort to delineate landscape pattern must take into account both the scale and diversity of plant communities found within the area.

Input Data

Polygon coverages. PRNS personnel provided two polygon coverages. One was the general administrative boundary for PRNS; the other contained a set of landscape units, intended to represent discrete patches of land cover types, and derived by manual interpretation of 1994 aerial photography. Both coverages can be downloaded from: *http://www.nps.gov/gis/park_gisdata/california/pore.htm*.

Satellite imagery. We relied on multi-spectral imagery from the Landsat Enhanced Thematic Mapper (TM-7) and Ikonos-2 satellites (Table 1). Both sets of imagery were purchased by the National Park Service, and copies were made available to us. The TM-7 image (Path 34/Row 44) was comprised of six bands covering the visible to shortwave infrared wavelengths at 30 m pixel resolution, one thermal-infrared band at 60 m resolution, plus a 15 m panchromatic band sensitive to blue wavelengths. The Ikonos-2 imagery contained four multi-spectral bands (visible to near-infrared) at 4 m resolution, and a 1 m panchromatic band sensitive to blue wavelengths. The Landsat TM-7 image was acquired on June 10, 2001; the Ikonos-2 data were provided pre-mosaicked from multiple images acquired on 11 and 14 July 2003. Aside from pixel resolution (Figure 2), the primary difference between the two types of imagery is that the Ikonos-2 scanner is insensitive to wavelengths beyond the near-infrared (band 4), whereas the TM-7 scanner is sensitive to thermal and shortwave infrared radiation. Further details about the spectral sensitivities of the two scanning systems may be found at *http://www.infoterra-global.com/data_media/spec_bands.gif*.

Table 1. Comparison of satellite imagery inputs.

	Landsat TM-7	Ikonos-2
Spatial resolution	MS[a]: 30 m Pan: 15 m	MS: 4 m Pan: 1 m
Bit depth (levels of grey)	8 bit (255)	11 bit (2048)
Number of bands	7 MS + 1 Pan	4 MS + 1 Pan
Spectral range	Blue - Thermal Infrared	Blue - Near Infrared
Cost	$0.19/km^2	$30/km^2

[a]MS = multi-spectral; Pan = panchromatic/black and white.

Methods

Segmentation of imagery. WSAL-M86 was used to generate a segmentation of Landsat TM 30 m data. The WSAL-M86 segmentation methodology, designed by the Wildlife Spatial Analysis Lab (WSAL) for classifying Landsat TM 30 m imagery, employs a two-stage segmentation process (see Barsness 1996, Ford et al. 1997). During the first stage, the image is divided into spectral classes using the ISODATA clustering algorithm, and during the second stage, those spectral classes are merged into contiguous regions using the M86 merging algorithm.

In addition to the WSAL-M86 segmentation, eCognition (2001) software was used to produce seven segmentations using both the Landsat TM and Ikonos images as follows: Landsat TM 30 m; Ikonos 4 m; Landsat TM pan-sharpened 15 m; Ikonos 4 m principal components (PC) image; Landsat TM 15 m PC image; Hybrid 1; and Hybrid 2. More detailed descriptions of the seven segmentations, and in particular the two hybrid methods follow below. First, however, we review the general approach.

The goal of the eCognition segmentations was to generate a segmentation with as coarse a scale as possible without missing the small and important land cover patches – thus matching the characteristics of the photo-interpreted regions. To this end, each was initially segmented at a very fine scale, creating a very large number of small regions. By iteratively increasing segmentation size with the eCognition software, these small regions were merged together into progressively larger ones until obviously distinct features were lost in the merge process. The image analyst then stepped back to the last acceptable scale and increased the shape parameter within eCognition. Once again, the analyst began to iteratively increase the segmentation scale until the obviously distinct feature was once more lost in the merging process. Thus the final segmentation for each method resulted from the combined adjustments of the shape and scale parameters in eCognition. By emphasizing shape at the final stages of the segmentation, the resulting regions tended to have smoother boundaries and be more compact, and as such, they better mimicked the manually delineated map unit boundaries.

While the above segmentation process was standard to all of the eCognition segmentations, specific techniques varied for the principal component and hybrid segmentations. Principal component analyses (PCA) of the pan-sharpened TM and Ikonos imagery were required for the former. PCA is a method of compressing an image's spectral information into fewer bands by means of a linear transformation to the spectral values such that the maximum variation in spectral values is contained in the first principal component, and subsequent components are oriented orthogonally to the previous principal components along the axis of greatest variance. This results in an image whose bands are uncorrelated, whose first band contains the largest fraction of spectral variation, and whose last band is primarily noise (ERDAS 1997). The proportions of spectral variation contained in each principal component are referred to as Eigenvalues.

Again, seven eCognition segmentations were generated. The first two used the original Landsat TM 30 m and Ikonos 4 m images, while the rest used image products derived from these originals. The third segmentation was based on a Landsat TM pan-sharpened 15 m image. This image's spatial resolution was increased from 30 m to 15 m

using a process called a resolution merging or pan-sharpening (ERDAS 1997). The next two segmentations, Ikonos 4 m PC and Landsat TM 15 m PC, used images generated from PCAs applied to the pan-sharpened Landsat TM 15 m and Ikonos 4 m images. The Eigenvalues for each PC band were used as a guideline for weighting the bands during segmentation (see Table 2).

Table 2. Eigenvalues and segmentation weights for principal components (PC) bands produced from Landsat TM 15 m pan-sharpened imagery and Ikonos (IK) 4 m imagery.

PC Band	Eigenvalue	Segmentation Weight
TM PC1	76%	2.0
TM PC2	16%	1.5
TM PC3	4%	1.0
TM PC4	3%	0.7
TM PC5	1%	0.5
IK PC1	69%	2.0
IK PC2	28%	1.0
IK PC3	3%	0.5

The last two segmentations sought to combine the strengths of each imagery type and thus involved the most complex and creative strategies. Images from Landsat TM and Ikonos satellites are highly dissimilar, both in terms of pixel size and spectral information (Table 1). Landsat TM has much larger pixels than Ikonos imagery (30 m vs. 4 m in the multispectral bands and 15 m vs. 1 m in the panchromatic band). On the other hand, Landsat TM provides a broader spectral range than Ikonos imagery (seven vs. four bands).

For the first hybrid segmentation, Hybrid1, we used the eCognition Landsat TM 30 m segmentation as a starting point and then re-segmented it to a finer resolution using the Ikonos 4 m imagery. For the second hybrid method, Hybrid2, we resampled the Landsat TM 15 m pan-sharpened PCA image to 4 m, then segmented this simultaneously with the Ikonos PCA 4 m imagery in eCognition.

Comparison of Segmentations. The purpose of undergoing image segmentation is to decompose the image into a jigsaw puzzle of interlocking regions based upon their spectral characteristics, with the goal being to represent distinct patches of vegetation on the ground. Thus, the pixels comprising the regions should be homogeneous, i.e., have similar spectral values, and the pixels between neighboring regions should be heterogeneous, i.e., have dissimilar values. We therefore calculated two statistics, one representing the homogeneity within regions, and one representing the heterogeneity between regions, using the first two bands from the Landsat TM and Ikonos PC images. The remaining bands were ignored because they contained relatively little variation in spectral information (Table 2).

The homogeneity index *(1)* was calculated by scaling the standard deviations calculated for each region according to its size (to compensate for the inherent tendency of smaller

regions to be more homogeneous than larger ones), and then averaging the scaled values for each of the two PC bands and across the entire image [see (1) below]. Finally these mean values were weighted according to each principal component's Eigenvalues (the percent contribution to overall diversity that each principal component represents) to produce an adjusted final score. For this statistic, lower scores indicate more homogeneous, and presumably more uniform, polygons.

(1) *Where:*
Std = Standard Deviation;
Count_R = Number of pixels in each region;
Count_I = Total number of pixels in the image;
PC_1 = Principal component 1;
R_1 = Region 1;
R_n = Region n; and
n = Total number of regions in image;

a) For each of the first two principal components, each region's standard deviation was scaled by its size: $SclStdPC_1R_1 = (StdPC_1R_1 * Count\_R_1/Count\_I)$

b) The mean of these scaled standard deviations was calculated for each of the two principal components as:
$MeanSclStdPC_1 = (\Sigma\ (SclStdPC_1R_1 + SclStdPC_1R_2 + \ldots + SclStdPC_1R_n))/n$

c) The means were then scaled by their Eigenvalues, such that:
$HomogeneityPC_1 = MeanSclStdPC_1 * EigenvaluePC_1$

The heterogeneity index *(2)* was calculated by averaging the "mean distance to neighbor" values for each region – a standard calculation within eCognition – for each principal component. This variable represents the spectral Euclidean distance between neighboring regions, weighted by the length of their shared border. These values were then weighted according to each principal component's Eigenvalues (the percent contribution to overall diversity that each principal component represents) to produce adjusted final scores. Larger scores indicate greater spectral differences between adjacent polygons and, presumably, more distinct regions.

(2) *Where:*
MDN = Mean distance to neighbor (eCognition 2001, pp 111-112);
PC_1 = principal component 1;
R_i = Region 1;
R_n = Region n; and
n = Total number of regions in the image;

$HeterogeneityPC_1 = (\Sigma\ (MDNPC_1R_1 + MDNPC_1R_2 + \ldots + MDN\ PC_1R_n))$
$/n * EigenvaluePC_1$

Ideally, the best segmentation would have low values for homogeneity within regions and high ones for heterogeneity between regions. We sought to explore this relationship by calculating the ratio of heterogeneity to homogeneity, with larger values

representing more optimal results. For ease of comparison with the photo-interpreted (PI) regions, we then divided the results for each segmentation by the PI results. These proportional values enabled us not only to easily see which segmentations performed better or worse than the PI, but also to quantify how much better or worse they were.

Results

Visual differences in the patterns common to each segmentation output are illustrated in Figures 3 and 4, while Table 3 summarizes the homogeneity and heterogeneity statistics for Ikonos and Landsat principal components. The segmentation patterns in Figure 3, when broadly examined, suggest that the PI regions are more closely aligned with the TM segmentation than with the Ikonos ones. At finer scales, however, the Ikonos segmentations appear to have the advantage over TM (Figure 4). The hybrid segmentations, which scored very well overall, have an unusual and somewhat unnatural appearance due to processing artifacts. This appearance tends to distract and somewhat offsets the statistical strength of the hybrid segmentations.

Upon initial examination of Table 3, it is clear that the segmentations tested better in terms of both homogeneity and heterogeneity against the imagery used in their production, i.e., Ikonos-based segmentations tested better against the Ikonos imagery and TM-based segmentations tested better against the TM imagery. According to the Ikonos-based statistics, the Hybrid1 segmentation performed best, with the standard Ikonos and Ikonos PCA scoring second and third, respectively. When looking at the TM-based statistics, however, TM PCA scored best, with TM 15 m and Hybrid1 ranking second and third, respectively. This seeming bias, perhaps stronger for Ikonos than TM, makes it difficult to decide which method is truly best. Because the relative ratios quantify how much better or worse the segmentations are compared to the PI regions, we can get a possible indication of the "best" method by computing the mean of these two ratios: These results (Table 3) follow those for the Ikonos-based statistics: Hybrid1 segmentation ranks highest, followed by the standard Ikonos and Ikonos PCA segmentations.

Discussion and Conclusions

The goal of this research was not to produce the best possible segmentation for land cover classification, but instead to determine whether manually photo-interpreted regions could be mimicked through automated techniques. Superficially, it is easy to grasp that the 4 m resolution of the Ikonos images matches more closely with the resolution of the aerial photos used for manual interpretation than with the 30 m Landsat TM; features not identifiable in the TM imagery may be easily distinguished in the Ikonos imagery (Figure 2b) and presumably the aerial photos. Thus, it is not surprising that the Ikonos-based segmentations performed much better overall than segmentations based on TM imagery. Perhaps more surprising, and certainly noteworthy, is that all of the Ikonos segmentations outperformed the photo-interpreted regions they were intended to emulate, whereas the eCognition Landsat TM 30 m and Landsat TM 15 m segmentations were the only ones to score lower than the PI regions.

This is not to suggest, however, that increased resolution will always produce better segmentations; as resolution increases, the overall variation in the image increases as well. As a result, features such as shadows may be identified as distinct regions instead of being incorporated in the larger features that spawned them, e.g., trees and buildings.

The diverse spectral characteristics of the imagery also contributed to differences in the segmentations. PI regions were delineated from 1:24000 true color aerial photographs (Moritsch et al. unpubl.), whereas Landsat TM has seven bands ranging from ultraviolet through mid-infrared, and Ikonos has four bands ranging from ultraviolet through near-infrared. The multiple spectral bands in Ikonos and TM give them a distinct advantage over aerial photos because they provide data in spectral wavelengths beyond visual range. Although increased spectral information did not seem to override TM's difficulty due to its large pixel size, it certainly contributed to Ikonos' advantage over the PI regions.

The hybrid outputs sought to optimize the segmentation process by combining the smaller pixel size of Ikonos with the greater spectral range of TM. Initial results are quite promising, with the Hybrid1 segmentation placing first overall. However, because of its somewhat oddly shaped regions, the segmentation loses much of its appeal, especially when compared to the PI regions. With more experimentation, a hybrid model may eventually yield the best results.

Again, the objective of this comparison was to determine whether photo- interpreted polygons could be reasonably mimicked through automated methods. We found this to be the case. However, with end-user input and feedback, automated methods can be taken one step further to actually improve upon photo-interpreted outputs. A useful model incorporated in more recent work has been to generate multiple segmentations from many inputs, and then to create maps of each for field surveys. Field crews use these maps to perform ocular assessments of the segmentation patterns and then provide feedback as to which segmentations best match ground conditions, as well as suggestions for improving upon the current segmentations.

Acknowledgments

We thank Dave Schirokauer of PRNS for help with all aspects of this project – from providing data, funding, and guidance, to reviewing products and waiting patiently for this report.

References

Barsness, S. 1996. Aggregating raster polygons derived from large, remotely-sensed images. Unpubl. MS thesis, University of Montana, Missoula. 58 pp.

eCognition. 2001. *User Guide*. Definiens Imaging GmbH, Munich, Germany.

ERDAS. 1997. *ERDAS Field Guide, 4th Edition, Revised and Expanded*. ERDAS, Inc., Atlanta, Georgia. *http://support.erdas.com/documentation/files/FieldGuide.pdf*

Ford, R., Z. Ma, S. Barsness, and R. Redmond. 1997. Rule-based aggregation for region identification. Proceedings of the 1997 American Society for Photogrammetry and Remote Sensing Annual Conference, Seattle, Washington.

Mitri, G.H. and I.Z. Gitas. 2004. A performance evaluation of a burned area object-based classification model when applied to topographically and non-topographically corrected TM imagery. Int. J. Remote Sensing 25(14): 2863-2870.

Moritsch, B.J., Allen, L.J., Allen, S., and T. Keeler-Wolf. Plant community mapping and classification at Point Reyes National Seashore and Golden Gate National Recreation Area. Unpublished manuscript (http://biology.usgs.gov/npsveg/pore/methods.pdf).

Wang, L., W.P. Sousa, and P. Gong. 2004. Integration of object-based and pixel-based classification for mapping mangroves with Ikonos imagery. Int. J. Remote Sensing 25(24): 5655-5668.

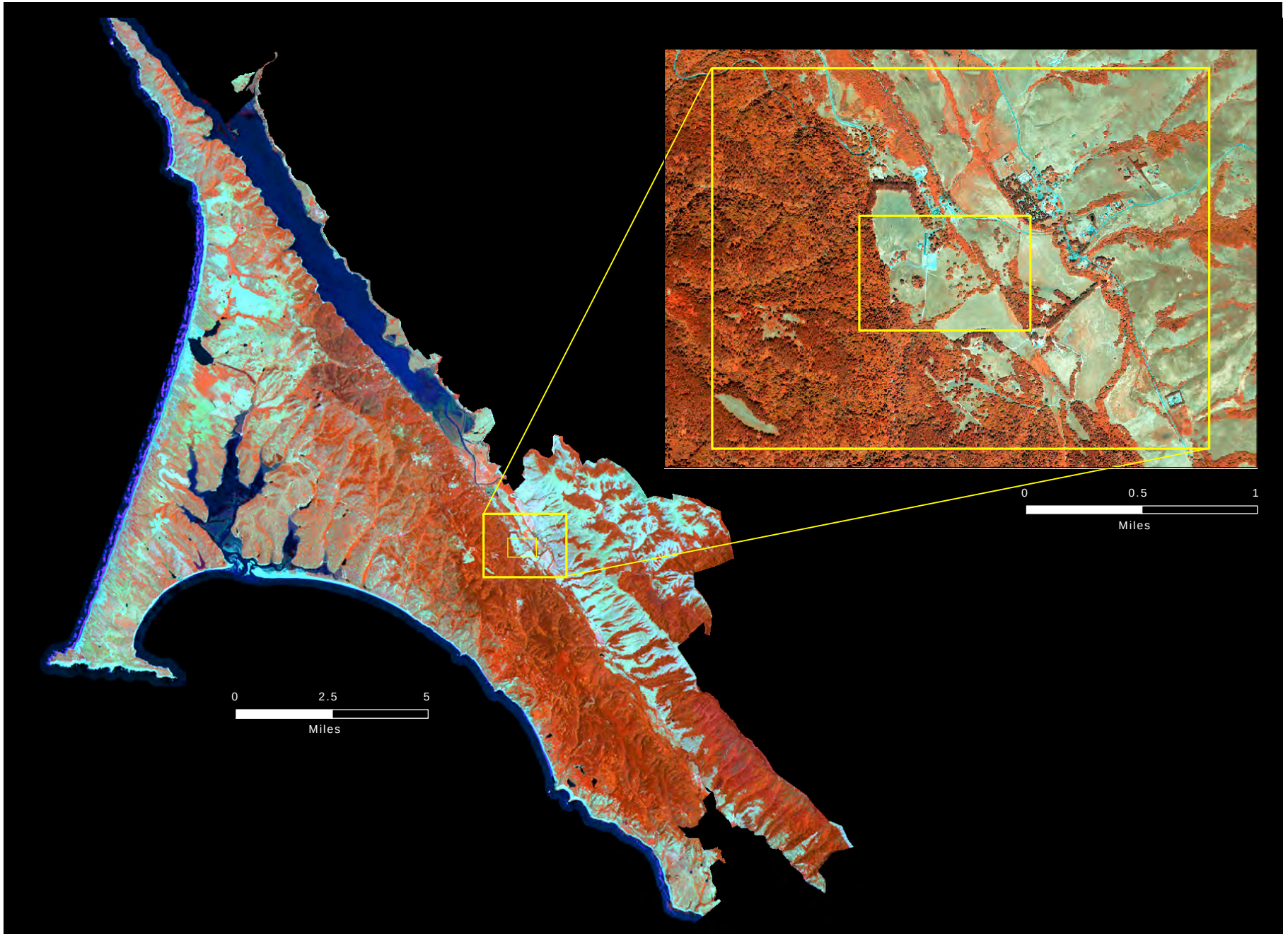


Figure 1. Point Reyes National Seashore as seen from IKONOS imagery. Inset map indicates the two areas shown on figures to follow.

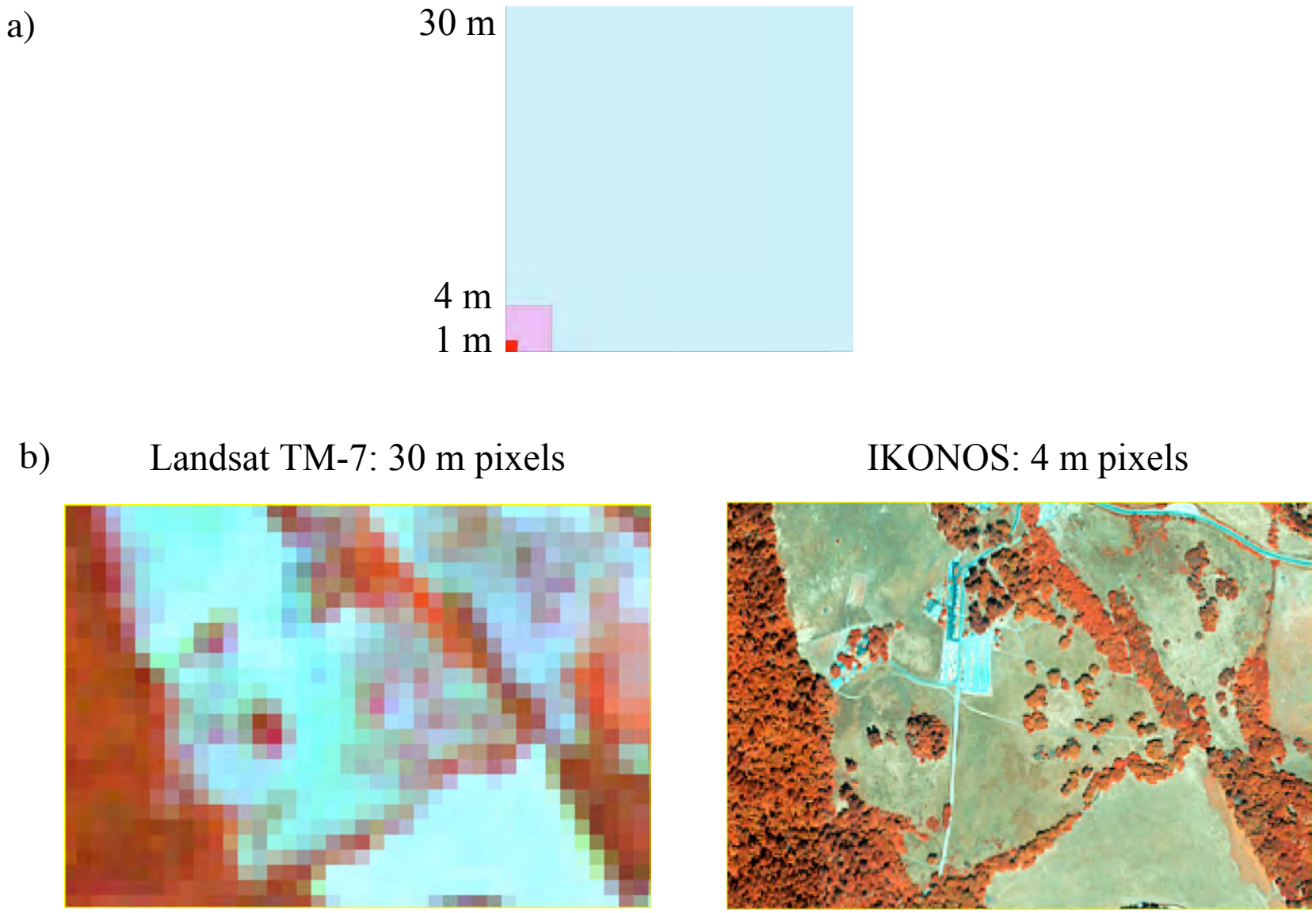


Figure 2. Comparison of image resolution: a) relative pixel sizes; and b) ability to distinguish fine-scale features on the ground (area shown corresponds to smaller inset, Figure 1).

Ikonos Imagery

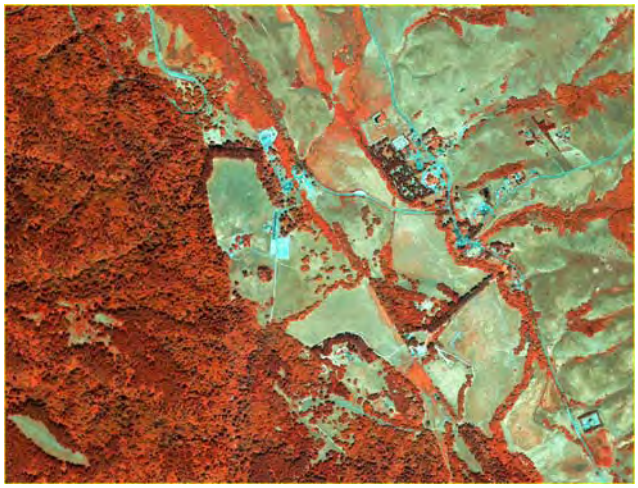
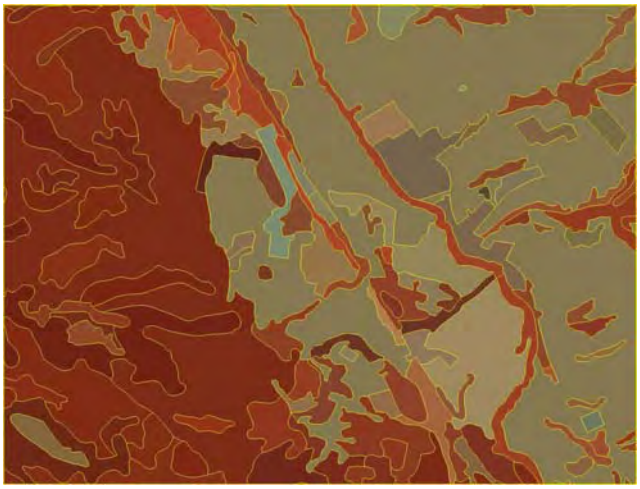
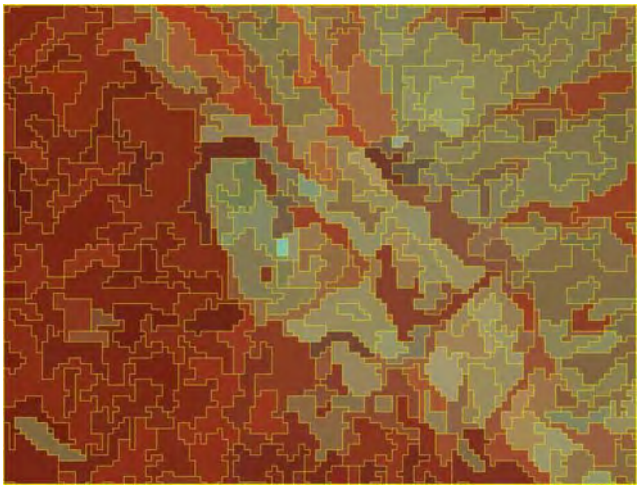


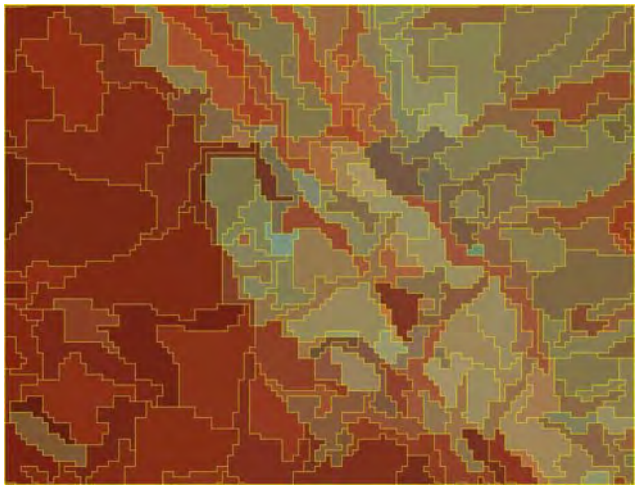
Photo Interpreted



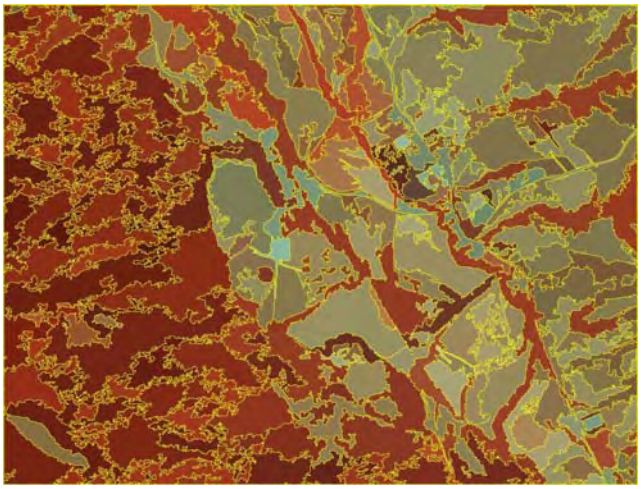
WSAL M86 TM



eCognition TM



eCognition Ikonos



eCognition Hybrid1

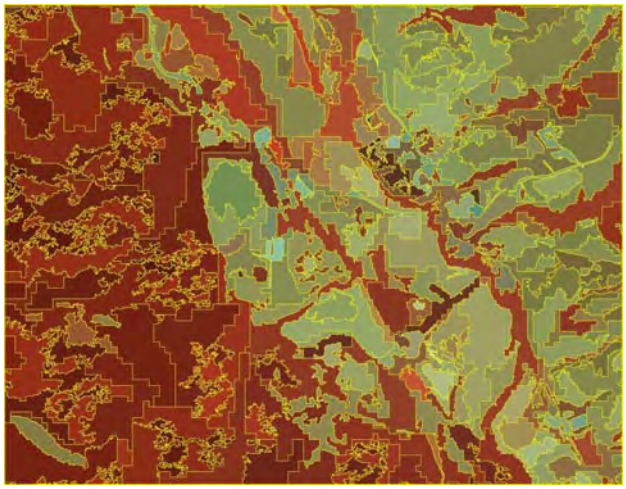
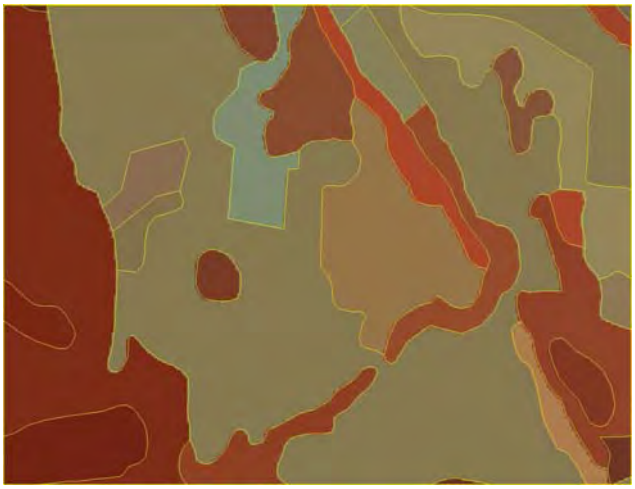


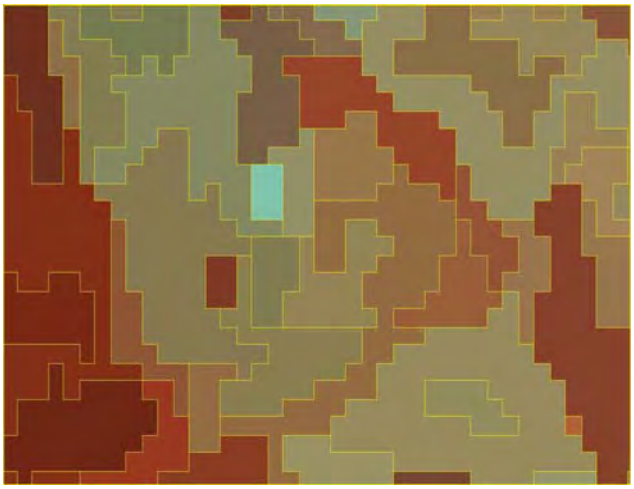
Figure 3. Segmentation outputs, with Ikonos imagery and photo-interpreted polygons shown as reference, for the larger inset area in Figure 1.

Ikonos Imagery

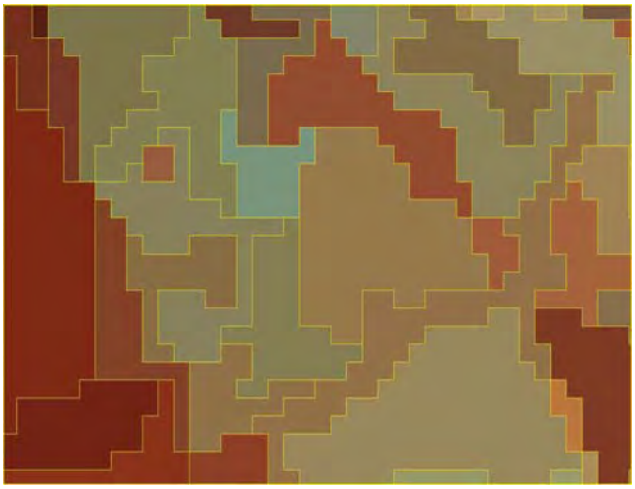
Photo Interpreted



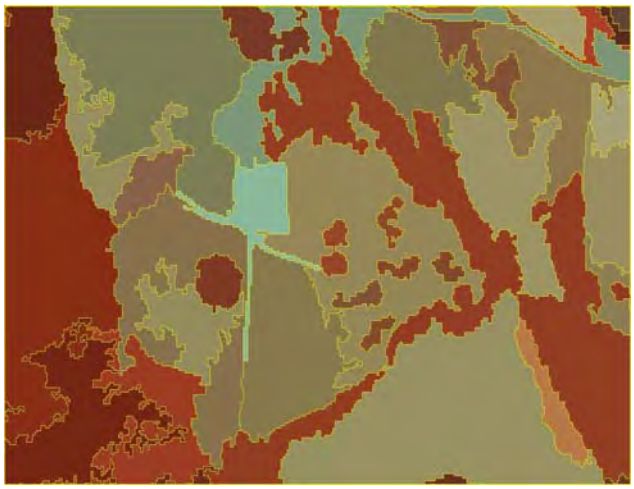
WSAL M86 TM



eCognition TM



eCognition Ikonos



eCognition Hybrid1

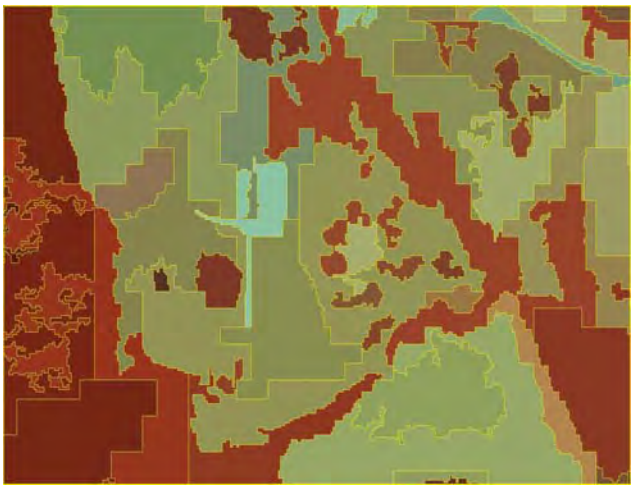


Figure 4. Segmentation outputs, with Ikonos imagery and photo-interpreted polygons shown as reference, for the smaller inset area in Figure 1.

Table 3. Relative heterogeneity/homogeneity ratios generated from both the Ikonos and pan-sharpened TM principal component images and for the eight segmentations plus the photo-interpreted map units (PI). Because the objective was to mimic PI regions through automated methods, ratio values have been scaled relative to PI = 1.0; actual ratio values can be calculated using the actual PI ratio of 27426.7 from the Ikonos PC image and 41709.79 from the pan-sharpened TM PC image. Larger numbers for the adjusted ratios indicate better results.

Adjusted Heterogeneity/Homogeneity Ratios									
PC Image	PI	WSAL-M86	TM30	TM15	TM-PC	IK	IK-PC	HYBRID1	HYBRID2
Ikonos	1.00	1.16	0.75	0.84	0.90	**2.21**	**1.94**	**2.75**	1.60
TM	1.00	1.03	0.73	**1.16**	**1.47**	0.83	0.96	**1.11**	1.03
	1.00	1.10	0.74	1.00	1.19	**1.52**	**1.45**	**1.93**	1.32