Comparison of Scene Segmentations and Classifications: SMAP, ECHO, and Maximum Likelihood[1]

James Darrell McCauley[2]

Abstract

Sequential maximum *a posteriori* (SMAP) segmentation, the Extraction and Classification of Homogeneous Objects (ECHO) segmentation, and maximum likelihood (ML) estimation were compared in a supervised classification of multispectral data collected from an airborne scanner. The generalization of SMAP provided larger and fewer contiguous areas and a more visually acceptable map than ML or ECHO. Iterative proportional fitting was used to to normalize error matrices and a Tukey multiple comparison test was used to examine differences in classification results. At a risk level of $\alpha = 0.001$, significant differences were found in all mean class classification accuracies: SMAP > ECHO > ML (in order of decreasing accuracy). However, evaluation through pairwise analyses of κ statistics showed no significant difference ($\alpha = 0.1$) between any two classifiers. Iterative proportional fittings and a Tukey multiple comparison test were found more sensitive and preferred over pairwise κ analyses.

1 Classification Results and Comparisons

Results of classifications in remote sensing and other fields are often presented in terms of an *error* or *confusion matrix* (Table 1). This matrix is more generally known as a *contingency table.* Along the diagonal of this table are the number of pixels correctly classified. The percentage of correctly classified pixels is commonly referred to as the *overall classification accuracy.* There is generally no argument that such a global measure is insufficient for comparisons of multiple classifications.

There are two types of errors shown in an error matrix. There are errors of *omission* [15] which occur when the classifier has failed to recognize when a sample belongs to a class of interest. In Table 1, $2 + 3 = 5$ samples from class C were omitted from this class. There are errors of *commission* [15] which occur when the classifier incorrectly assigns a sample to a class of interest. In Table 1, there were 47 samples assigned to class C, but $5 + 1 = 6$ samples should not have been committed to this class.

---

[1] A Report Prepared for Dr. David Landgrebe and EE 577, School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907–1285. Submitted 18 April 1994.

[2] USDA Graduate Fellow, Department of Agricultural Engineering, Purdue University, West Lafayette, Indiana 47907–1146, mccauley@ecn.purdue.edu

Table 1: An example of a confusion matrix (from [15]).

		Ground truth classes			Total
		A	B	C	
Thematic map classes	A	35	2	2	39
	B	10	37	3	50
	C	5	1	41	47
Number of ground truth pixels		50	40	46	136

When examining the results for a particular class, it is important to state the point of view from which the results were interpreted: either the consumer or the producer [5, 20]. Considering again class C in Table 1, the *consumer's accuracy* is $41/47 = 87\%$ while the *producer's accuracy* is $41/46 = 89\%$. Usually the producer's accuracy is not equal to the consumer's accuracy, and neither can be considered *class accuracy.*

In many situations, it may be desirable to examine the interaction between classes in a way that is not confounded by marginals or to make comparisons incorporating effects of off-diagonal elements. Many have developed indices for comparisons between classes and between separate classifications [16], however the kappa (κ) coefficient, developed by Cohen [4], has been the most recommended [5–7, 16]. Details of the computation of the estimate $\hat{\kappa}$ are given by Congalton [5], correcting a typographical error which had propagated into many other works. A formulation for the large sample variance is given by Congalton *et al.* [7]. Given these computational details, $\hat{\kappa}$ can be computed and pairwise comparisons can be made between independent $\hat{\kappa}$'s by:

$$z_{ab} = \frac{\hat{\kappa}_a - \hat{\kappa}_b}{\sqrt{\mathrm{Var}\,(\hat{\kappa}) + \mathrm{Var}\,(\hat{\kappa})}}$$

However, $\hat{\kappa}$ only indirectly incorporates the effects of off-diagonal elements. Perhaps a better approach for comparing several alternative classifications was that taken by Zhuang *et al.* [21]. In their work, error matrices were normalized using *iterative proportional fitting* [8] and class accuracies were compared *simultaneously.*

Iterative proportional fitting is a simple technique which makes row and column totals

Table 2: Sample normalized error matrix.

0.801	0.118	0.081	1.000
0.114	0.831	0.055	1.000
0.086	0.050	0.864	1.000
1.000	1.000	1.000	3.000

simultaneously 100%. After the smoothing of zero counts [9], each iteration of the algorithm requires two steps. First, rows are divided by row totals. Next, column totals are recomputed and columns are then divided by these totals. Row and column totals are recomputed with each iteration. A stopping criteria, such as maximum number of iterations or maximum marginal change, terminates the procedure. Table 2 presents the normalized version of the matrix in Table 1 after 26 iterations of this algorithm.

Given a normalized matrix, such as the one shown in Table 2, class accuracies (along the diagonals) for several classifications can be compared simultaneously using the Tukey multiple comparison method [14]. This method is preferable over pairwise comparisons since the risk level, α, applies to the procedure as a whole and not particular comparisons. Tukey's method is designed to control the probability of making at least one Type I error in the comparison of all pairs of treatment means, *i.e.*, claiming a difference in classifications when a significant difference does not exist. This method of comparison is useful for multiple comparison of results of different treatments when the number of samples is the same (*i.e.*, comparing classifier parameters or classification algorithms). Zhuang *et al.* [21] used this method to compare results of three different algorithms for classifying LANDSAT TM data.

2 Scene Segmentation Algorithms

Unlike traditional point methods such as maximum likelihood (ML) estimation, scene segmentation classifiers take advantage of the *spatial* information of samples in multispectral imagery. Conceptual overviews of two scene segmentation algorithms are given below; refer to the cited literature for details.

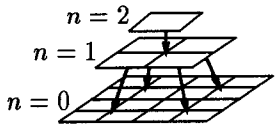


Figure 1: "Pyramid structure of the MSRF. The random field at each scale is causally dependent on the coarser scale field above it. [2]"

2.1 Sequential Maximum a Posteriori Estimation

For many approaches to Bayesian image segmentation, pixel labels are modeled as Markov random fields (MRF) and segmented by approximating the maximum *a posteriori* (MAP) estimate of pixel labels (for an example, see the work of Besag [1]). Based upon this approach to segmentation, Bouman and Shapiro [2,3] replaced the MRF model with a new multiscale random field (MSRF) model. The MSRF is composed of a sequence of random fields having coarse to fine scales (see figure 1). Each finer field is only dependent upon the previous coarser field. Thus, the sequence of these fields form a Markov chain. Also, instead of using the MAP estimate, a new estimate of pixel labels was used. It minimizes the expected size of the largest misclassified region. This new estimator is calculated recursively and was thus termed a sequential MAP (SMAP) estimator [2,3]. The result of this model and estimator form a new method of Bayesian image segmentation which may be used in either a supervised or unsupervised mode. Bouman and Shapiro [3] show that SMAP segmentation provided significantly better results than maximum likelihood estimation for a SPOT image with ground truth along transects. Details of the algorithm and comparisons to ML and a couple of approaches of MAP estimation are given by Bouman and Shapiro [2,3].

2.2 Extraction and Classification of Homogeneous Objects

ECHO (Extraction and Classification of Homogeneous Objects) [11,13] is another algorithm for scene segmentation. The scene segmentation process is a two-stage, conjunctive approach (cell selection and annexation) and may be used in either a supervised or unsupervised mode. First, data are divided into rectangular cells consisting of four or more pixels and subjected to a test of homogeneity. Cells which cannot be called homogeneous (*i.e.*, singular) are

assumed to overlap a boundary in the data. In the second step, annexation, adjacent non-singular cells are tested for statistical similarity and are subject to annexation into a group of cells comprising a homogeneous object. Tunable parameters include the number of pixels comprising cells and threshold values for selection and annexation tests. Landgrebe [13] reviewed results of extensive testing in which ECHO provided greater classification accuracy at greater computational efficiency than did ML classification.

3 Objective

The objective of this study was to compare the SMAP, ECHO, and ML methods for scene segmentation using two different methods: analysis of κ statistics and iterative proportional fitting/Tukey multiple comparison.

4 Methods

The data used in this study, known as "Flightline C1," is a 949 pixel $\times$ 220 pixel 12 band image (wavelengths from 0.40–1.00 μm) acquired from an airborne scanner. The scene, scanned during June 1966, covered a small portion of farmland in Tippecanoe County, Indiana. The scanner had an IFOV of 3 *mrads* and was flown an altitude of 792 m above the terrain. The sensor scanned approximately ± 40 degrees about nadir. A/D conversion used 8 bits. Ground truth was available for most of the image. Figure 2 shows a composite image using band 9 for both blue and green and band 1 for red. The image has a large percentage of areas which appear homogeneous. Rectangular ground truth areas are outlined in white (fig. 2).

Training fields and spectral bands were selected to build spectral signatures that gave adequate results for the ML approach. Then, ML, SMAP, and ECHO classified the known agricultural areas outlined in figure 2 using the same training fields. For the ECHO classification, cell size chosen was 2 pixel $\times$ 2 pixel and the threshold values selected for both selection and annexation tests was 2.0. For all classifications, signature classes were forced to be the same as information classes—no subclass structure was used.

The normalization procedure reviewed by Zhuang *et al.* [21] and briefly described ear-

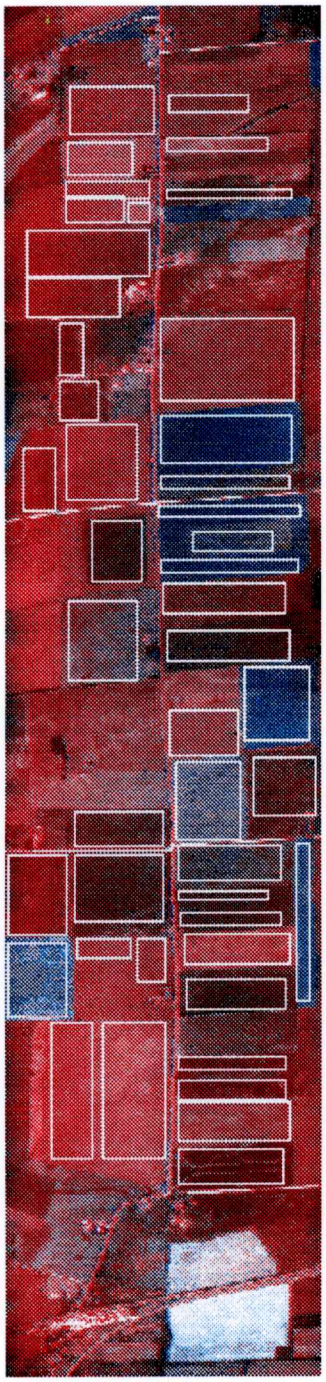


Figure 2: Color composite of flightline C1. Areas with ground truth are outlined in white.

lier was implemented in the GRASS geographic information system [18] (see Appendix A). Error matrices for each method were normalized using iterative proportional fitting (with smoothing of zero counts). By extracting class accuracies along the diagonals of normalized matrices, classifier performance was represented as a two factor statistical experiment.

Following the procedure of Zhuang *et al.* [21] for the two factor experiment, (1) probabilities that individual classes were correctly classified were assumed inherent for a classifier, (2) classifiers were assumed independent, and (3) map classes were assumed approximately independent. The Tukey critical distance (ω) between treatment means was computed using [14]:

$$\omega = q_\alpha(p, v)\frac{s}{\sqrt{n_t}}$$

where

$q_\alpha(p, v)$ = critical value of the Studentized range at a given risk level, α

p = number of sample means of classifiers

v = number of degrees of freedom associated with MSE

s = $\sqrt{MSE}$

n_t = number of observations in each of the p samples

If any two population means differed by a distance greater than the Tukey critical distance, they were judged to be different.

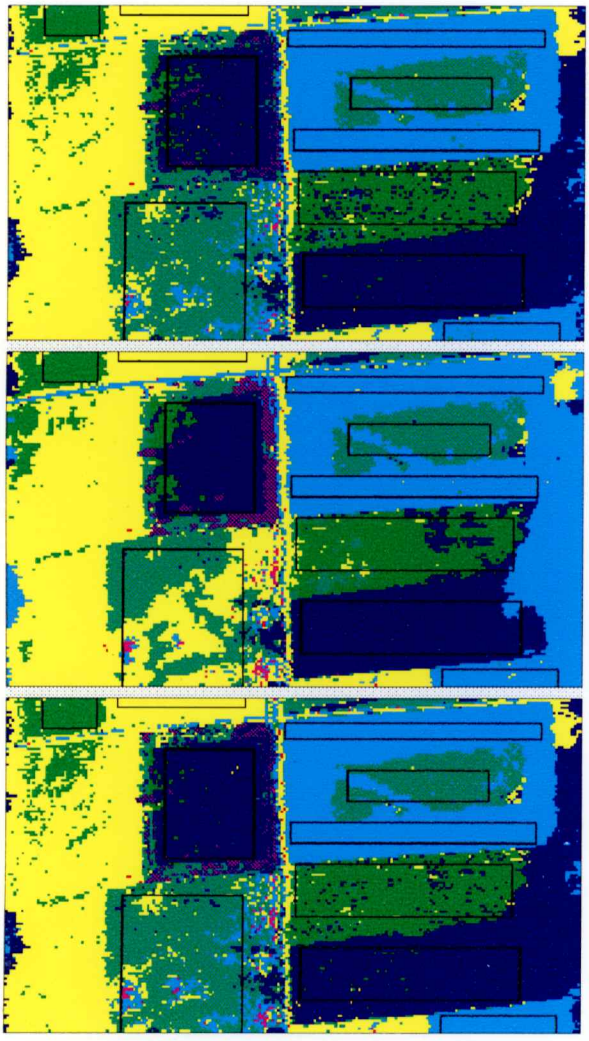
5 Results

Table 3 gives the eight classes defined for the image in figure 2 as well as the distribution of pixels in testing and training fields. For an initial set of training samples, principal component analysis and a preliminary classification with ML indicated that four features were adequate for classification. After examining distance metrics for the defined classes (Table 3) between all possible combinations, four channels were selected. These four bands, 1, 6, 10, and 12, corresponded to wavelength intervals of 0.40–0.44, 0.52–0.55, 0.66–0.72, and 0.80–1.00 μm.

Small sections (128 $\times$ 219) of the classified scene are given for each of ML, SMAP, and ECHO in figure 3. For descriptive purposes only, four of the texture features of Haralick

Table 3: Thematic classes defined for flightline C1.

ID	Class	# Pixels Used	
		Training	Testing
1	Soybeans	175	24423
2	Corn	108	10286
3	Oats	120	5472
4	Wheat	107	9538
5	Red Clover	190	11574
6	Alfalfa	77	3106
7	Rye	21	2340
8	Bare Soil	24	1170
Totals		822	67909



(a) ML estimation.

Contrast:	0.950
Correlation:	0.801
Variance:	3.384
Entropy:	1.007
# Contig. Areas:	1542
Avg Area Size:	18.26

(b) SMAP segmentation.

Contrast:	0.495
Correlation:	0.893
Variance:	3.327
Entropy:	0.892
# Contig. Areas:	535
Avg Area Size:	52.64

(c) ECHO segmentation.

Contrast:	0.944
Correlation:	0.806
Variance:	3.435
Entropy:	1.009
# Contig. Areas:	1442
Avg Area Size:	19.53

Figure 3: Small segments (128×219) of classified images for (a) ML estimation, (b) SMAP segmentation, and (c) ECHO segmentation. Co-occurrence texture features (contrast, correlation, variance, and entropy) were computed for each segment.

et al. [10] (contrast, correlation, variance, and entropy) and the number and average size of contiguous areas were computed for each section. These are also given in figure 3. For the entire image, there were 15349, 7312, and 14531 contiguous areas produced for ML, SMAP, and ECHO, respectively. Average areas were 13.6, 28.6, and 14.3 pixels^2. Texture features were not calculated for the entire image.

Error matrices of the classifiers for all ground truth fields (fig. 2) are given in Table 4. Overall accuracy and κ statistics are given in Table 5. Standard normal deviates for pairwise comparisons were $z_{e,m} = 0.72$, $z_{e,s} = 0.39$, and $z_{s,m} = 1.11$, indicating no significant differences at the $\alpha = 0.1$ level.

Each 8×8 matrix had approximately 20 random zeros. Therefore, a Bayesian estimate of element probabilities (see Appendix A) was used to smooth matrices through the addition of pseudo-counts. Smoothed versions of these are presented in Table 6. Marginal totals (not shown) remained the same as those in Table 4. After smoothing, iterative proportional fitting was used to normalize the matrices (Table 7) so that row and column totals were simultaneously $\approx 100\%$. By extracting class accuracies along the diagonals of the normalized matrices in Table 7, a two factor experiment was completed (Table 8). The results of Tukey's Studentized range test for a risk level of $\alpha = 0.001$ are given in Table 9.

6 Discussion

The SMAP segmentation yielded larger contiguous areas (fig. 3) than did both the ECHO segmentation or the traditional ML estimation. The image classified by SMAP appeared to have the least speckled ("salt and pepper") appearance (fig. 3), which is desirable for many applications of thematic mapping. Quantitatively, the co-occurrence texture features and area calculations in figure 3 agree with this assessment.

From the normalized class accuracies given in Table 8, it is clear that SMAP classification performed better than ML and ECHO classifications for every class. The results of Tukey's multiple comparison test (Table 9) showed that this distinction was statistically significant at the $\alpha = 0.001$ level. All three classifications fell into separate Tukey groupings (indicating statistically significant differences), with ML classification being the least accurate according

Table 4: Error matrices for (a) maximum likelihood classification, (b) SMAP, and (c) ECHO

Map Classes	Ground Truth Classes							
	1	2	3	4	5	6	7	8
1	24005	43	238	41	12	0	44	40
2	1571	8080	274	3	358	0	0	0
3	309	6	4248	340	560	2	7	0
4	5	0	67	9374	0	0	92	0
5	76	276	301	3	10213	705	0	0
6	18	39	272	11	319	2446	1	0
7	10	0	3	736	0	0	1591	0
8	152	0	0	0	0	0	0	1018

(a) maximum likelihood estimation

Map Classes	Ground Truth Classes							
	1	2	3	4	5	6	7	8
1	24241	3	106	20	11	0	28	14
2	1230	8789	137	1	129	0	0	0
3	209	6	4462	295	498	1	1	0
4	1	0	19	9504	0	0	14	0
5	65	32	190	2	10989	296	0	0
6	17	30	229	10	150	2669	1	0
7	8	0	2	642	0	0	1688	0
8	108	0	0	0	0	0	0	1062

(b) SMAP segmentation

Map Classes	Ground Truth Classes							
	1	2	3	4	5	6	7	8
1	24100	18	167	54	10	0	40	34
2	1100	8720	173	4	289	0	0	0
3	302	8	4204	387	548	1	22	0
4	2	0	19	9459	0	0	58	0
5	70	124	195	2	10810	372	1	0
6	20	44	260	12	290	2479	1	0
7	9	0	2	595	0	0	1734	0
8	120	0	0	0	0	0	0	1050

(c) ECHO segmentation

Table 5: Classification statistics.

Overall Accuracy	$\hat{\kappa}$	$\hat{\sigma}_{\kappa}$	METHOD
89.8	0.870	0.000788	ML
92.1	0.899	0.000850	ECHO
93.4	0.915	0.000854	SMAP

Table 6: Smoothed error matrices for (a) maximum likelihood classification, (b) SMAP, and (c) ECHO. Matrices were smoothed through the addition of pseudo-counts.

Map Classes	Ground Truth Classes							
	1	2	3	4	5	6	7	8
1	24005.228	42.953	237.973	40.942	11.936	-0.018	43.991	39.995
2	1570.963	8080.106	273.992	2.975	357.978	-0.007	-0.004	-0.003
3	308.972	5.989	4248.060	339.992	559.994	1.996	6.998	-0.001
4	4.943	-0.019	66.989	9374.123	-0.025	-0.007	91.998	-0.002
5	75.932	275.982	300.990	2.972	10213.129	705.003	-0.005	-0.003
6	17.982	38.995	272.000	10.993	318.997	2446.036	0.999	-0.001
7	9.986	-0.005	2.997	736.006	-0.006	-0.002	1591.024	-0.001
8	151.995	-0.002	-0.001	-0.003	-0.003	-0.001	-0.000	1018.016

(a) maximum likelihood estimation

Map Classes	Ground Truth Classes							
	1	2	3	4	5	6	7	8
1	24241.232	2.950	105.973	19.942	10.934	-0.017	27.991	13.994
2	1229.958	8789.116	136.990	0.975	128.974	-0.007	-0.004	-0.003
3	208.971	5.989	4462.063	294.991	497.993	0.996	0.998	-0.001
4	0.943	-0.019	18.989	9504.125	-0.026	-0.006	13.996	-0.002
5	64.932	31.977	189.989	1.972	10989.140	295.997	-0.005	-0.003
6	16.982	29.994	229.000	9.993	149.994	2669.039	0.999	-0.001
7	7.986	-0.005	1.997	642.004	-0.006	-0.002	1688.025	-0.001
8	107.995	-0.002	-0.001	-0.003	-0.003	-0.001	-0.000	1062.016

(b) SMAP segmentation

Map Classes	Ground Truth Classes							
	1	2	3	4	5	6	7	8
1	24100.230	17.951	166.975	53.942	9.934	-0.016	39.990	33.994
2	1099.957	8720.114	172.991	3.975	288.976	-0.007	-0.004	-0.003
3	301.973	7.989	4204.059	386.993	547.994	0.996	21.998	-0.001
4	1.944	-0.019	18.989	9459.124	-0.026	-0.006	57.997	-0.002
5	69.933	123.978	194.990	1.972	10810.136	371.998	0.995	-0.003
6	19.982	43.994	260.000	11.993	289.996	2479.036	0.999	-0.001
7	8.986	-0.005	1.997	595.004	-0.006	-0.002	1734.026	-0.001
8	119.995	-0.002	-0.001	-0.003	-0.003	-0.001	-0.000	1050.016

(c) ECHO segmentation

Table 7: Normalized error matrices for (a) maximum likelihood classification, (b) SMAP, and (c) ECHO. Matrices were normalized through iterative proportional fitting.

Map Classes	Ground Truth Classes								
	1	2	3	4	5	6	7	8	total
1	0.932	0.010	0.031	0.001	0.002	-0.000	0.009	0.015	1.000
2	0.029	0.932	0.017	0.000	0.022	-0.000	-0.000	-0.000	1.000
3	0.018	0.002	0.851	0.017	0.108	0.001	0.002	-0.000	1.000
4	0.001	-0.000	0.027	0.915	-0.000	-0.000	0.058	-0.000	1.000
5	0.002	0.042	0.025	0.000	0.814	0.117	-0.000	-0.000	1.000
6	0.001	0.013	0.049	0.000	0.055	0.882	0.000	-0.000	1.000
7	0.001	-0.000	0.001	0.067	-0.000	-0.000	0.931	-0.000	0.999
8	0.016	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.985	1.001
total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	8.000

(a) maximum likelihood estimation

Map Classes	Ground Truth Classes								
	1	2	3	4	5	6	7	8	total
1	0.965	0.001	0.017	0.000	0.001	-0.000	0.008	0.007	1.000
2	0.014	0.976	0.006	0.000	0.004	-0.000	-0.000	-0.000	1.000
3	0.010	0.003	0.903	0.008	0.076	0.000	0.000	-0.000	1.000
4	0.000	-0.000	0.015	0.965	-0.000	-0.000	0.020	-0.000	1.000
5	0.002	0.008	0.021	0.000	0.900	0.069	-0.000	-0.000	1.000
6	0.001	0.012	0.038	0.000	0.018	0.931	0.000	-0.000	1.000
7	0.001	-0.000	0.001	0.027	-0.000	-0.000	0.972	-0.000	0.999
8	0.008	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.993	1.001
total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	8.000

(b) SMAP segmentation

Map Classes	Ground Truth Classes								
	1	2	3	4	5	6	7	8	total
1	0.949	0.005	0.027	0.001	0.001	-0.000	0.005	0.012	1.000
2	0.018	0.958	0.011	0.000	0.014	-0.000	-0.000	-0.000	1.000
3	0.016	0.003	0.882	0.010	0.086	0.000	0.004	-0.000	1.000
4	0.000	-0.000	0.016	0.943	-0.000	-0.000	0.041	-0.000	0.999
5	0.002	0.023	0.021	0.000	0.864	0.090	0.000	-0.000	1.000
6	0.001	0.012	0.042	0.000	0.035	0.909	0.000	-0.000	1.000
7	0.001	-0.000	0.001	0.046	-0.000	-0.000	0.950	-0.000	0.999
8	0.013	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.988	1.001
total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	8.000

(c) ECHO segmentation

Table 8: Performance summary for (a) maximum likelihood classification, (b) SMAP, and (c) ECHO.

Map Classes	Segmentation Algorithms		
	ML	SMAP	ECHO
Soybeans	0.932	0.965	0.949
Corn	0.932	0.976	0.958
Oats	0.851	0.903	0.882
Wheat	0.915	0.965	0.943
Red Clover	0.814	0.900	0.864
Alfalfa	0.882	0.931	0.909
Rye	0.931	0.972	0.950
Bare Soil	0.985	0.993	0.988

Table 9: Results of Tukey's Studentized range test for maximum likelihood classification, SMAP segmentation, and ECHO segmentation ($\alpha = 0.001$, $v = 13$, MSE $= 0.00003$, $q_\alpha(p, v) = 10.224$, $\omega = 0.0199$). Means with the same letter are not significantly different.

Tukey Grouping	Mean	n_t	METHOD
A	0.950625	8	SMAP
B	0.930375	8	ECHO
C	0.905250	8	ML

to this test procedure. According to overall accuracies (Table 8), ECHO performed better at classification than ML, though the number of contiguous areas and the average size of these areas in figure 3 were close to the same.

The results of Tukey's multiple comparison test were in contrast with the three pairwise comparisons using z-tests of differences in $\hat{\kappa}$ values. The latter analyses showed no significant differences at a risk level of $\alpha = 0.1$. This indicated that, in this instance, the iterative proportional fitting/Tukey multiple comparison method was more sensitive to different classifications. Visual analyses of classified images led to the conclusion that $\hat{\kappa}$ was too conservative.

Notice in figure 3 that, even though ECHO (and ML) had higher entropy, the segmentation by SMAP was much different at the right edge and in the lower left test field. If the training fields and spectral classes had been optimized for SMAP instead of ML, this may have changed.

Bouman and Shapiro [3] showed in their tests that SMAP outperformed ML. However, their ground truth was located along transects instead of being contained in rectangular regions. The manner in which test samples were selected had an obvious effect on accuracy determination—rectangular test fields will naturally favor methods which yield larger contiguous areas. For classifying remotely-sensed imagery of rectangular and thematically homogeneous (agricultural) fields, this approach to delineating test fields is arguably appropriate. When the objective is to determine the appropriate class and perhaps the corresponding area, a segmentation should be the most accurate for areal units despite spectral impurities.

7 Summary

The SMAP algorithm produced better ($\alpha = 0.001$ for Tukey multiple comparison) results that ML and ECHO, in this instance, for supervised classification of a scene of agricultural fields. Unlike a natural scene, the one used in these classification had the characteristic of being thematically homogeneous. SMAP appeared better at producing maps with larger contiguous areas and with a less speckled appearance. The resulting thematic map for

SMAP was visually more acceptable than those produced by the other methods. ECHO segmentation outperformed ML estimation, which produced the most speckled map.

For these classifications, Tukey multiple comparisons of normalized class accuracies was more sensitive that pairwise κ analyses. For the latter, no significant difference was detected at a risk level of $\alpha = 0.1$.

Acknowledgments

The author acknowledges the assistance of Xin Zhuang, Wyle Laboratories, Arlington, VA and Charles Bouman, School of Electrical Engineering, Purdue University through the early release of their manuscripts [3,21]. Bouman also provided helpful suggestions for describing SMAP training data.

Software Credits

ECHO segmentation was done using *MultiSpec* [12], written by David Landgrebe and Larry Biehl, both of the School of Electrical Engineering, Purdue University. Maximum likelihood classification was done using the GRASS geographic information system (GIS) [19] modules `i.gensig` and `i.maxlik`, written by Michael Shapiro, U. S. Army Construction Engineering Research Laboratory, and Tao Wen, University of Illinois at Urbana-Champaign. SMAP segmentation was also accomplished with GRASS through modules called `i.gensigset` and `i.smap`, written by Shapiro and Bouman. Iterative proportional fitting was accomplished using the GRASS module `m.ipf`, written by the author. Co-occurrence texture features were calculated with the GRASS module `i.texture`, also written by the author. The Tukey multiple comparison test was done using SAS [17].

References

[1] J. Besag. On the statistical analysis of dirty pictures. *J. Royal Statist. Soc.*, 48(3):259–302, 1986.

[2] C. Bouman and M. Shapiro. Multispectral image segmentation using a multiscale image model. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pages III–565–III–568, San Francisco, California, March 23–26 1992. IEEE.

[3] C. Bouman and M. Shapiro. A multiscale random field model for bayesian image segmentation. *IEEE Transactions on Image Processing*, 1994. (in press).

[4] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–40, 1960.

[5] R. G. Congalton. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.*, 37:35–46, 1991.

[6] R. G. Congalton and R. A. Mead. A quantative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, 49(1):69–74, Jan. 1983.

[7] R. G. Congalton, R. G. Oderwald, and R. A. Mead. Assessing landsat classification accuracy using discrete multivariate statistical techniques. *Photogrammetric Engineering and Remote Sensing*, 49(12):1671–1678, Dec. 1983.

[8] S. E. Fienberg. A statistical technique for historians: Standardizing tables of counts. *Annals of Mathematical Statistics*, 1:305–315, 1971.

[9] S. E. Fienberg and P. W. Holland. Methods for eliminating zero counts in contingency tables. In G. P. Patil, editor, *Random Counts in Scientific Work*, volume 1, pages 233–260. Pennsylvania State University Press, 1970.

[10] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man, and Cybertinetics*, SMC-3(6):610–620, Nov. 1973.

[11] R. L. Kettig and D. A. Landgrebe. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics*, GE-14(1):19–26, Jan. 1976.

[12] D. Landgrebe and L. Biehl. *An Introduction to MultiSpec.* School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907–1285 USA, 2.15.94 edition, Feb. 1994. Contact: `landgreb@ecn.purdue.edu`.

[13] D. A. Landgrebe. The development of a spectral-spatial classifier for earth observational data. *Pattern Recognition*, 12:165–175, 1980.

[14] W. Mendenhall and T. Sincich. *A Second Course in Business Statistics: Regression Analysis.* Dellen Pub. Co., San Francisco, 3rd edition, 1989.

[15] J. A. Richards. *Remote Sensing Digital Image Analysis.* Springer-Verlag, Berlin, 2nd edition, 1993.

[16] G. H. Rosenfield and K. Fitzpatrick-Lins. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52(2):223–227, Feb. 1986.

[17] SAS Institute, Cary, NC. *SAS/STAT User's Guide*, 6.03 edition, 1988.

[18] M. Shapiro, J. Westervelt, D. Gerdes, M. Larson, and K. R. Brownfield. *GRASS 4.1 Programmers Manual.* U.S. Army Corps of Engineers, Construction Engineering Research Laboratories, Champaign, Illinois, 1993.

[19] M. Shapiro, J. Westervelt, D. Gerdes, M. Larson, and K. R. Brownfield. *GRASS 4.1 Users Manual.* U.S. Army Corps of Engineers, Construction Engineering Research Laboratories, Champaign, Illinois, 1993.

[20] M. Story and R. G. Congalton. Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3):397–399, Mar. 1986.

[21] X. Zhuang, B. A. Engel, X. Xiong, and C. Johanssen. Analysis of classification results of remotely sensed data and evaluation of classification algorithms. *Photogrammetric Engineering and Remote Sensing*, 1994. (in press).

A Iterative Proportional Fitting with Smoothing of Zero Counts

Before normalization of an error matrix using iterative proportional fitting (as described earlier), *random* zeros (as opposed to *fixed* zeros) in the matrix should be eliminated. Feinberg and Holland [9] developed a method for smoothing with pseudo-counts using a Bayesian estimator of the probability for a matrix element p_{ij}:

$$\hat{p}_{ij} = \frac{N}{N+k}\left(x_{ij} + k\lambda_{ij}\right)$$

where x_{ij} is an entry in the ith row and the jth column of the matrix and $N = \sum_{i,j} x_{ij}$. In Bayesian analysis, the probability p_{ij} is regarded as a random variable. The density function $\pi\left(p_{ij}\right)$ is proportional to:

$$\prod_{i,j} p_{ij}^{k\lambda_{ij}-1}.$$

The expected value is estimated as:

$$\mathrm{E}_{\pi}[p_{ij}] = \lambda_{ij} \approx \frac{X_{i.}X_{.j}}{N^2}$$

and k, the number of pseudo-counts, is calculated as:

$$k = \frac{N^2 - \sum_{i,j} x_{ij}^2}{\sum_{i,j}\left(N\lambda_{ij} - x_{ij}\right)^2}.$$

The following calculation procedure was given by Zhuang *et al.* [21]:

> For smoothing a contingency table with pseudo-counts, first we calculated the "expected value," $N\lambda_{ij}$, instead of λ_{ij} for simplicity of computation. Next, k was computed with the formula given above. Third, the k pseudo-counts were allocated to the individual entries of the *expected value* table, and the entries were multiplied by the ratio k/N. Finally, the contingency table was added to the table obtained in the third step entry by entry, and the result was multiplied by $N/(N+k)$ to preserve the original total of N.

NOT INCLUDED

The manual page and C source code for this procedure implemented in the GRASS GIS [19] follows on the next 9 pages. The program is available via anonymous ftp from

`pasture.ecn.purdue.edu:pub/mccauley/grass/m.ipf.tar.gz`

It is currently under beta testing and will appear in the next release of GRASS (4.2).