060772

Technical Memorandum T3
June 7, 1972

A Cluster-Oriented Analysis of Multispectral Scanner Data

by

Paul E. Anuta

Introduction

The standard procedure at LARS for classification analysis of aircraft scanner data has been to choose training samples (fields) based on the ground truth data and to use these directly or in a form subdivided via clustering as training for the computer. In this process the major emphasis is on defining class statistics which describe actual materials of interest as designated by the user. The approach assumes these materials are spectrally separable and classification is performed to test if this is true. In the approach described here the multispectral data is first clustered to determine what spectrally separable groups exist in the data. After these separable groups are found they are related to their physical meaning or "ground truth". This approach is in a sense the reverse of the ground truth oriented training procedure referred to above.

There are two requirements often cited for multispectral(*) pattern recognition to be useful for a particular material in the scene. First it must be spectrally separable from all the others in the scene and, two, it must be of informational value. In the existing approach the features which are of informational value are first defined. In the cluster oriented approach the spectrally separable classes are found first and time consuming training and test classification analysis to determine the spectral separability of unseparable materials is avoided.

---

* Remote Sensing Analysis: A Basic Preparation, John Lindenlaub, LARS Information Note 110471.

Cluster Analysis Procedures

The clustering analysis was carried out using both the LARS clustering program algorithm (NSCLAS) and group sample clustering program (GRPSAM). The details of these processors are discussed in LARS program abstracts LARSYS 0600 and UNSUP 0031. The goal of this study was detection of corn leaf blight in corn fields; thus, the emphasis was on spectral differences in corn only. The NSCLAS procedure is to cluster all points in the corn fields and then use the results to assign each field to a particular blight class. The criterion for placing a particular field in a spectral class is that the majority of points in the field must fall into the chosen cluster. This approach enables ground truth to be related to cluster in a convenient manner, and furthermore it allows smoothing (averaging) of both ground truth data and spectral data. While this approach is reasonable for relatively homogeneous fields, it fails if, for example, a field is split half and half into two clusters.

Three LARS data sets from the 1971 Corn Blight Watch Experiment were used in the study reported here. Each of the three were processed in some way to remove certain effects in the data. The data set identifier numbers are:

1. Run 71054105: Data with sun angle correction.
2. Run 71054103: Redigitized and line averaged data with sun angle correction.
3. Run 71054107: Principal components transformation on run 71054103 using statistics from all corn fields for the transformation.

Detailed ground truth for blight level was available for the corn fields in this data, and so a large sample analysis could be performed with regard to the levels of corn blight in the fields.

Analysis of Run 71054105

The NSCLAS point clustering program was used to cluster all the corn fields defined in this run. The majority cluster was determined for each of these fields, The cluster assignments are listed in Table 1. The reasons for choosing six clusters will be discussed later. The ground truth data for these fields included an expert observer judgement of the corn blight severity in each field. The blight levels were defined on a scale of 0 to 5, 0 being no blight and 5 being severely blighted corn. Table 1 includes the ground truth blight level for each field as determined by the ground observers.

Table 1. Results of NSCLAS clustering of run 71054105, 12 channels, 6 clusters.

CLUSTER	FIELD	BLIGHT LEVEL	CLUSTER MEAN	VAR
2	TT-7	4	2.5	4.5
3	II-1	1		
	XX-1	0		
	WW-3	2		
	AAA-3	1		
	AAA-5	2		
	OO-4	2		
	OO-2	2		
	SS-6	3		
	A*-18	2		
	UU-4	1		
	VV-3	2		
	J-2	2		
	L-9	0		
	L-8	1		
	L-25	0		
	L-18	0	1.3	1.2
4	W-3	4		
	N-2	4		
	VV-6	3		
	VV-7	3		
	SSS-1	4		
	A*-24	3		
	UU-13	2		
	C*-3	4		
	ZZ-7	2		
	C*-5	2		
	D*-4	4		
	QQQ-1	4	3	1.5
5	JJJ-3	3		
	JJJ-2	1		
	JJJ-5	2		
	EE-11	3		
	EE-10	3		
	KKK-11	3		
	PPP-1	0		
	VVV-1	2		
	YY-6	3		
	RR-2	2		
	C*-8	0		
	WW-1	3		
	UU-15	3		
	UU-7	2		
	RRR-1	2		
	UU-6	2		
	RRR-2	4		
	UU-8	3		
	SSS-2	1		
	R-6	3		
	F-8	1		
	N-1	4		
	L-23	1	2.2	1.3
6	R-9	2		
	UUU-1	2		
	RR-3	2		
	RR-4	2		
	EE-12	3		
	EE-13	2		
	EE-8	3		
	EE-16	2		
	EE-15	3		
	EE-9	2	2.3	.2

The blight levels for all the fields assigned to each cluster were averaged and the variance was computed. These values are listed in Table 2. Also, included in the table are the mean values of multispectral scanner channel 8 for each of the clusters. Channel 8 (.72-.92) is the closest approximation available in scanner data to the infrared sensitive band of color infrared film. This is the band which is related to the high reflectance of green healthy vegetation. It is thought to be related to blight stress in that stressed plants reflect less strongly in the IR and therefore the values in Table 2 would be lower for blight related clusters. This is only a hypothesis at this stage. The blight means were observed to increase as Channel 8 values decreased; however. The color IR film relationship is made here since this form of photography is becoming widely used for remote sensing surveys and tying scanner data results in with photographic evidence is considered to be very desireable.

Table 2. Blight Mean vs. Cluster for 71054105

CLUSTER	Ch 8 MEAN (IR)	BLIGHT MEAN	BLIGHT VARIANCE
2	77	2.5	4.5
3	85	1.3	1.2
4	73	3	1.5
5	78	2.2	1.3
6	78	2.3	.2

Table 3. Results of Group Sample Clustering of Run 71054105 12 channels, 6 groups.

GROUP	FIELDS	NSCLAS CLUSTER	BLIGHT LEVEL	BLIGHT MEAN	BLIGHT VAR
1	JJJ-5	5	2		
	UUU-1	6	2		
	C-16		3		
	C-3		2		
	JJJ-3	5	3		
	JJJ-2	5	1		
	EE-9	6	2		
	EE-16	6	2		
	EE-15	6	3		
	EE-8	6	3		
	KKK-4		1		
	EE-13	6	2		
	EE-12	6	3		
	EE-10	5	3		
	EE-11	5	3		
	PPP-1	5	0		
	VVV-2		2		
	RR-4	6	2		
	RR-3	6	2		
	RR-2	5	2		
	WW-2		0		
	UU-10		3		
	F-8	5	1	2	.86
2	II-1	3	1		
	YY-6	5	3		
	XX-1	3	0		
	WW-3	3	2		
	AAA-3	3	1		
	AAA-5	3	2		
	OO-4	3	2		
	OO-2	3	2		
	SS-6	3	3		
	UU-11		1		
	A*-18	3	2		
	UU-4	3	1		
	SSS-2	5	1		
	VV-3	3	2		
	J-2	3	2		
	F-9		0		
	L-23	5	1		
	L-8	3	1		
	L-9	3	0		
	L-25	3	0		
	L-18	3	0	1.3	.5

Table 3 - continued

GROUP	FIELDS	NSCLAS CLUSTER	BLIGHT LEVEL	BLIGHT MEAN	BLIGHT VAR
3	R-9	6	2		
	NNN-2		1		
	KKK-10		2		
	QQQ-2		3		
	VVV-4	5	2		
	C*-8	5	0		
	ZZ-9		1		
	UU-7	5	2		
	RRR-1	5	2		
	RRR-2	5	4		
	A*-21		2	2.0	1.1
4	UU-6	5	2		
	R-6	5	3		
	L-33		2		
	KKK-11	5	3		
	D*-8		4		
	D*-6		3		
	D*-4	4	4		
	ZZ-7	4	2		
	WW-1	5	3		
	AAA-10		1		
	AAA-12		2		
	UU-13	4	3		
	UU-15	5	3		
	UU-8	5	3		
	UU-2		2		
	VV-7	4	3		
	VV-6	4	3		
	N-1	5	4	2.8	.65
5	B-7	2	1		
	B-6		2		
	TT-7	2	4		
	QQQ-1	4	4		
	C*-5	4	2		
	C*-6		1		
	C*-3	4	4		
	A*-24	4	3		
	A*-27		2		
	SSS-1	4	0		
	F-2		3	2.4	1.9
6	N-2	4	4		
	W-3	4	4	4	0

The results in Table 2 strongly suggest that cluster 4 represents the more severe blight level and that clusters 5 and 6 represent an intermediate level. Cluster 3 has low variance and a lower blight mean indicating that it could be a low blight or normal corn class. Cluster 2 is considered unreliable due to the small sample and high variance.

The group sample clustering program (GRPSAM) was used to cluster the fields as a unit rather than cluster the individual points in the field. Table 3 contains the results of a 12 channel, 6 group clustering of all the corn fields for run 71054105. The relationship between group and blight level mean is:

Table 4. Blight Mean vs. Group for Run 71054105.

GROUP	BLIGHT MEAN	BLIGHT VARIANCE
1	2	.86
2	1.3	.5
3	2	1.1
4	2.8	.65
5	2.4	1.9
6	4	0

The GRPSAM results produce three to four groupings of fields. Groups 1, 2, and 3 have low blight means, groups 4 and 5 have somewhat greater means and group 6 contains two with level 4 blight. The results tend to agree with the point clustering (NSCLAS) results in that group 2 with the lowest blight mean had fields in it predominantly from cluster 3, also having the lowest blight mean, and the same group/cluster relationship exists for the highest blight mean.

The results of the clustering operations were used to select a set of training fields for three corn classes, and the remaining corn fields were separated into three classes for test. Five fields of "other" were selected for training although no test fields were defined for "non-corn"*. The fields chosen are listed in Table 5.

Table 5. Corn Training Fields for 71054105.

STAT CLASS	FIELD DESIGNATIONS	FIELD BLIGHT LEVEL	DISPLAY CLASS
Corn 1	TT-7	4	non corn
Corn 2	SS-6	3	Level 1
Corn 3	JJJ-3,SSS-2	2.5-.5	Level 1
Corn 4	EE-9,RR-4	2-1.75	Level 2
Corn 5	W-3,a-24	4-3.5	Level 3

The LARSYSAA per point classifier was used to classify the data using the best nine channels chosen by the $DIVERG program (channels 1, 2, 3, 5, 8, 9, 10, 11, 12). The results, listed in Table 6, indicate that the clusters found in the corn are not very separable, and even if all the fields in each cluster represented a specific blight level then only about 50% of any one level would be correctly recognized.

Table 6. Per Point Classification Results for 71054105.

CLASS	NO. SAMPS	PCT CORR	NO. CLASSIFIED AS: LEVEL 1	LEVEL 2	LEVEL 3
Corn Lv.1	10615	54	5768	1596	801
Corn Lv.2	3053	69	191	2101	67
Corn Lv.3	3178	50	243	16	1575
Overall		56	(Points not accounted for were classified as "other.")		

* All other cover types, i.e., soybeans, trees, pasture, etc.

The LARSYSAA sample or "per field" classifier was employed to classify each of the fields as a unit rather than as a set of points. The results of the sample classification are presented in Table 7. The higher accuracy is an expected result for sample

Table 7. Sample Classification Results for 71054105.

CLASS	NO. FIELDS	PCT CORR	NO. CLASSIFIED AS		
			LEVEL 1	LEVEL 2	LEVEL 3
Level 1	38	67	24	4	0
Level 2	8	88	0	7	0
Level 3	10	60	0	0	6
Overall		69	(Fields not accounted for were classified as other)		

classification. Even with these improved results the blight separation accuracy appears to be marginal.

A word about channel selection and the number of clusters chosen is in order here. The six corn cluster choice was arrived at in an intuitive manner. It was hoped that at least six corn subclasses could be found. Clustering results for 10 clusters produced a great deal of mixing and several vacant clusters. Six clusters seemed to produce a reasonable number of uniform fields; the results agreed with the color IR photography there was one vacant cluster, and three to four clusters of corn consisteltly turned up. Thus six became somewhat of a standard for the study. The intention in channel selection was to use "all" the information in the data. Thus 12 channels were used for the clustering analysis. Classification time considerations dictated that less than 12 be used for per point classifications. It is known that adjacent visible band channels (channels 1 through 7) tend to be correlated and so it was assumed that four of these seven would contain about as much information as all seven. It is also known that the five IR channels all tend to be valuable, thus a nine-channel philosophy was adopted. The best nine were determined for the various training sets used in the study. It is interesting to note that $DIVERG always chose all five of the IR channels, as expected.

Table 8. NSCLAS Clustering Results for Line Averaged Data
Run 71054103, 12 channels, 6 clusters.

Cluster	Field	Blight Level	Ave	Var
3	QQQ-1	4		
	II-1	1		
	C-5	3		
	SS-6	3		
	UU-1A	-		
	UU-1B	-		
	A-27	2.5		
	PP-67	-		
	L-33	2.25		
	L-25	0		
	L-18	0		
	L-8	1.5	1.9	1.9
4	J-2	2		
	VV-3	2		
	SSS-2	.5		
	UU-4	1.5		
	RRR-1	2		
	UU-10	3.25		
	A-18	2		
	A-16	-		
	AAA-10A	2		
	OO-2	2		
	GG-4.8	-		
	GG-6.11	-		
	AAA-5	1.5		
	WW-3	2.75		
	RR-5	2		
	RR-2	2		
	LLL-1	-		
	EEE-8	-		
	EE-11	3		
	JJJ-5	2		
	OO-11	-		
	JJJ-4	-		
	JJJ-3	2.5	2	.4
5	QQQ-2	2-3		
	QQ-6	-		
	NN-2	-		
	D-8	3.5		
	D-4	4		
5	ZZ-7	2		
	ZZ-6	-		
	C-3	3.5		
	WW-1	3		
	ZZ-9	2		
	AAA-12	2		
	UU-11	1		
	UU-15	3		
	UU-6	2		
	UU-8	3		
	UU-2	2		
	SSS-1	0		
	VV-8	-		
	N-6	-		
	N-2	4		
	N-1	3.5		
	W-3	4	2.6	1.4
6	F-8	1		
	F-9	0		
	R-9	1.5		
	R-6	3		
	UUU-1	2		
	UU-7	2		
	RR-3	1.5		
	YY-7	-		
	YY-6	2.5		
	RR-4	1.75		
	VVV-4	2		
	EEE-5	-		
	EEE-1	-		
	EEE-4	-		
	EE-10	3-3.5		
	PP-2	-		
	EE-12	3		
	EE-13	2.5		
	EE-8	2-3		
	EE-16	1.8		
	EE-15	2.5		
	EE-9	2		
	NNN-6	-		
	C-3	-	2	.6

Analysis of Smoothed Scanner Data: Run 71054103

Run 71054103 embodies several preprocessing steps aimed at reducing the random variability of the data by averaging redundant scan lines. Also the sun sensor signal was processed in such a way as to prevent roll movements of the scanner aircraft from erroneously altering the illumination estimate although this information was not used in this study. These data had slightly different geometry and new field cards had to be obtained. Clustering was carried out on the data using both the point and group sample programs. Table 8 contains the NSCLAS results for all the corn fields. The mean blight level and cluster relate as shown in Table 9.

Table 9. Blight Mean vs. Cluster for Run 71054103.

CLUSTER	Ch 8 MEAN (IR)	MEAN BLIGHT LEVEL	BLIGHT LEVEL VARIANCE
3	106	1.5	1.8
4	108	1.5	.8
5	93	2.0	1.8
6	101	1.35	1.0

Clusters 1 and 2 occurred in very small numbers and no field represented these clusters to any extent. There appears to be a trend here in that cluster 5 had more high blight level fields than the others.

To get a better understanding of the cluster results, the color IR photography from both the 9" x 9" NASA RB57 camera and the 70mm Michigan camera were inspected in order to compare color with cluster. It was observed that cluster and color correlated very well in that increasing cluster number was related to darkening field tone. Table 10 lists the author's color judgement and the associated cluster for several fields. Three different major corn field colors were identified and the different wordings are variations of these. It was very difficult to describe color absolutely due to the subtle variations in tone. Another problem was that there existed a wide variation in color and brightness radially from the center of the photo to the edge; a field appearing in the overlap of two photos with one more toward the center than the other would have very different colors and brightnesses. There also appeared

Table 10. Comparison of Color and Cluster for 71054103, Seg. 212, 43M. Michigan CIR film. 12 channels, 6 cluster NSCLAS Results.

FIELD	BLIGHT LEVEL	COLOR DESCRIPTION	MAJORITY CLUSTER
C-16	2.5,3	Pastel voilet with white spots	3
C-6,7	N-1,T-2	White (mostly) in C-6	2
		Dark violet in lower part C-7	4
C-3		Very dark purple	6
JJJ-3	2.5	Violet stripes with gray-blue (purple background)	4
JJJ-2	2.8	Same as JJJ-3	4
JJJ-4	NA	Same as JJJ-3	4
OO-11		Mixed violet with white spots	4
EE-9	2	Dark purple (very uniform)	6
EE-15	2.5	Same as EE-9	6
EEE-16	1.8	Same as EE-9	6
EE-8	2.5	Same as EE-9	6
JJJ-5	2	Mixed violet and purple (some 6)	4
EE-13	2.5	Dark purple	6
EE-12	3	Same as EE-13	6
PPP-2	NA	Purple	6
QQQ-2	2.5	Bluish (5) Some violet (4) striped	5
KKK-1	3	(6) West purple; (5) Center, blue	5
ZZ-7	2	Dark purple-blue (mixed edges)	5
WW-1	3	Dark purple-blue (mixed)	5
ZZ-9	No-T2	Very mixed (3,4,5,6)	5
AAA-12	2	Blue (5) with violet (4) spots	5
VV-8	3	Blue with 3,4 edges (violet)	5
SSS-1	0	Solid blue, blue-gray	5
L-33	2.25	Pale pastel violet with white (1)	3
N-6	NA	Deep purple-blue	5
N-2	4	Deep blue	5
N-1	3.5	Deep purple-blue	5
W-3	4	Blue	5

to be a sun angle effect which shifted the hot spot from the center to the northwest quadrant of the picture. Even against this confusion there appeared to be three corn field colors: violet, purple, and blue. The blue fields included the ones having blight level 4 and they always fell into cluster 5. This result led to the conclusion that severely blighted corn was identifiable and cluster 5 defined this level. Clusters 3 and 4 were mostly mixed in the fields and appeared to represent normal or slightly blighted corn. The meaning of cluster 6 was confusing. Many level 3 fields fell into cluster 6 which would suggest 6 was the intermediate blight level cluster; however, many low blight levels also fell into 6 pulling the average down.

To get further insight into what the clusters represented, the spectral response means for the six clusters were plotted in Figure 1. The spectral response curves for clusters 4, 5, 6 are all very close together with the darkest cluster (6) having the lowest channel means, as is expected. The red color in the film is represented by channel 8 (.72-.92μ), and the data values here are ordered according to the proposed blight level relationship. The cluster having the lowest response in channel 8 is cluster 5, the blue, blight level 4 cluster (channel 8 mean of 93). The next highest value is for cluster 6 (mean of 101), which is hypothesized as the next less severe blight level. Clusters 3 and 4 have essentially the same value (106 and 108) and are hypothesized as the lowest blight level. The shapes of the curves for clusters 3, 4, 5, 6 are nearly identical but not ordered the same as for channel 8.

In general, increasing cluster number is related to lower scanner data value for each channel, which accounts for the darker tone of the higher number clusters. The difference in color is due to the fact that the red-producing band is increasing in value while the other bands are decreasing or staying the same; (this is true for the cluster order 5, 6, 4). At cluster 3 the blue and green response starts to increase along with the red, and this accounts for the pale violet-pink color. Then in clusters 2 and 1 the blue and green increase greatly, and white is produced even though the red increased very little. Evidently the film response curve for red is steeper than for green and blue, since a much greater increase in green and blue compared to red was required to produce whites. The color and spectral evidence led to the conclusion that clusters 3 and 4, then 6, then 5 represent three increasingly severe blight levels. The reason for low blight readings to fall into cluster 6 is unknown, but since the fields in the cluster are spectrally similar it is reasonable to question the relevance of the ground truth reading to spectral response.

The group sample clustering program was used to group the individual field statistics into six groups using 12 channels. The list of fields which fell in each group is presented in Table 11. The group/cluster relationship presented is:

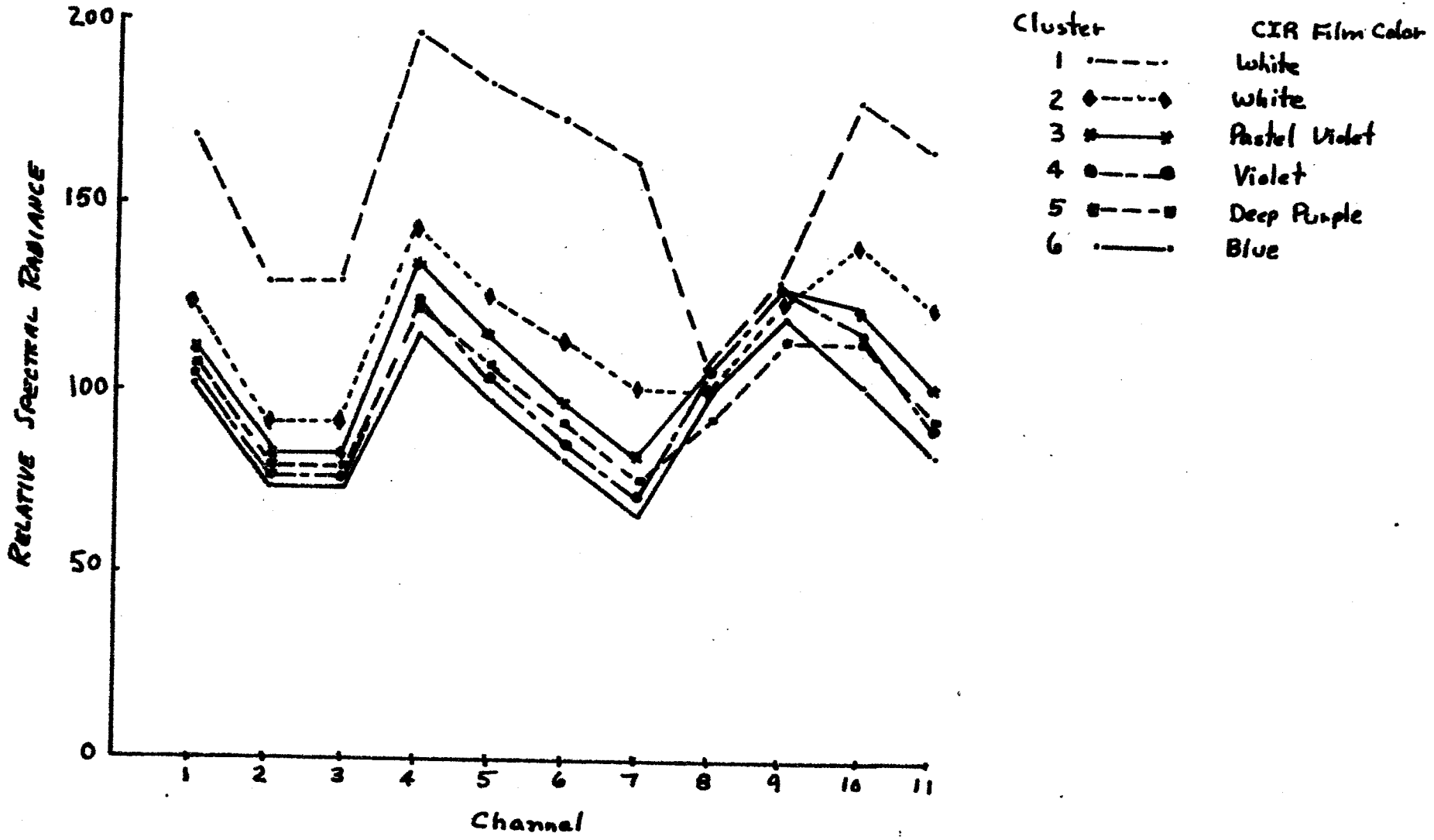

Figure 1. Scanner Data Channel Means for the Corn Field Clusters from Run 71054103.

Table 11. GRPSAM Results for Run 71054103, 6 groups, 12 channels.

GROUP	FIELD	NSCLAS CLUSTER	BLIGHT LEVEL	BLIGHT MEAN	BLIGHT VAR
1	B-6.7				
	KKK-4		1.5		
	KKK-10		3		
	KKK-11		3	2.5	.8
2	QQ-6	5			
	EEE-9				
	II-1	3	1		
	AAA-3		1		
	SS-6	3	3		
	UU-11	5	1		
	A*18		2		
	RRR-1	4	2		
	UU-1A	3			
	UU-1B	3			
	SSS-2	4	.5	1.5	.8
3	EEE-4	6			
	PPP-2	6			
	YY-4				
	NN-1				
	VVV-4	6	2		
	D*-8	5	3.5		
	D*6		3		
	D*4	5	4		
	HHH-2-4				
	C*8		0		
	ZZ-7	5	2		
	ZZ-6	5			
	ZZ-9	5	1		
	AAA-10A	4	1		
	AAA-12	5	2		
	UU-15	5	3		
	UU-7	6	2		
	UU-8	5	3		
	UU-6	5	2		
	UU-2	5	2		
	VV-8	5			
	VV-7		2.5		
	VV-6		3		
	UUU-1	6	2		
	R-6	6	3		
	R-9	6	1.5		
	N-1	5	3.5	2.3	.8

Table 11 - continued

GROUP	FIELD	NSCLAS CLUSTER	BLIGHT LEVEL	BLIGHT MEAN	BLIGHT VAR
4	C-16				
	JJJ-2		1.5		
	JJJ-3	4	2.5		
	JJJ-4	4			
	OO-11	4			
	EE-9	6	2		
	EE-8	6	2.5		
	EE-11	4	3		
	EE-12	6	3		
	EE-13	6	2.5		
	EE-16	6	1.8		
	EE-15	6	2.5		
	QQQ-2	5	2.5		
	PPP-1		0		
	NN-2	5			
	VVV-2		2		
	RR-2	4	2		
	C*5		2		
	C*3	5	3.5		
	WW-2		0		
	GG-4.6	4			
	AAA-10A	4	1		
	UU-10	4	3.25		
	A*16	4			
	RRR-2		3.5		
	A-21		3		
	A*27	3	2.5		
	A*24		3.5	2.3	1.0
5	QQQ-1	3	4		
	C*6		.5		
	WW-1	5	3		
	SSS-1	5	0		
	PP-6.7	3			
	L-33	3	2.25		
	N-6	5			
	N-2	5	4		
	W-3	5	4	2.5	2.9

Table 11 - continued

GROUP	FIELD	NSCLAS CLUSTER	BLIGHT LEVEL	BLIGHT MEAN	BLIGHT VAR
6	TT-4				
	L-3				
	NNN-6	6			
	EE-10	6	3.25		
	JJJ-5	4	2		
	EEE-1	6			
	EEE-5	6			
	EEE-8	4			
	RR-4	6	1.75		
	RR-3	6	1.5		
	YY-6	6	2.5		
	RR-5	4			
	LLL-1	4	2		
	WW-3	4			
	AAA-5	4	1.5		
	GG-6.11	4	1.5		
	OO-4				
	OO-2	4	1.5		
	UU-4	4	2		
	VV-3	4	1.5		
	L-28		2		
	[illegible]-2	4	2		
	F-8	6	1		
	F-9	6	0		
	[illegible]		.75		
	L-8	3	.75		
	L-25	3	0		
	L-18	3	0	1.4	.8

Table 12. Six-Group Clustering Results for Run 71054103

GROUP	BLIGHT MEAN	VARIANCE
1	2.5	.8
2	1.5	.8
3	2.3	.8
4	2.3	1.0
5	2.5	2.9
6	1.4	.8

The blight variance is low for these groups and a very interesting symmetrical blight grouping is seen here. Groups 2 and 6 have almost identical low blight means (1.5 and 1.4), groups 3 and 4 have identical means (2.3) and groups 1 and 5 have identical high means of 2.5. To further explore this three-level grouping, the GRPSAM program was rerun for three groups.

The three-group clustering produced two groups of almost identical mean:

Table 13. Three-Group Clustering Results for 71054103

GROUP	BLIGHT MEAN	VARIANCE
1	1.8	1.)
2	2.6	2.1
3	1.9	1.4

This result tends to reinforce the indication from the NSCLAS results that the low level clusters (3,4) and intermediate level cluster (6) tend to be mixed in blight level. Thus there are apparently two spectrally separate sets of corn fields with similar blight conditions. The third group (group 2) has a significantly higher blight mean, and this indicates that the more severely blighted corn can be separated from all other corn fields.

Samples taken from the corn cluster results were used to train the LARSYSAA pattern classifier for classification analysis. Five classes of corn were defined based on KMEANS clusters 2, 3, 4, 5, and 6, and hay, sprouts and soybeans were defined for the "other" classes. Statistics were obtained and all of segment 212 was classified on a per point basis (355,200 points). All the available corn fields in the segment were grouped according to the cluster results into one of three blight level groups, i.e., Level 1 - cluster 3 & 4; Level 2 - cluster 6; Level 3 - cluster 5. These fields formed the test deck for evaluating the classification results. Cluster 2 was grouped with the other classes since the spectrum for cluster 2 was judged more like bare soil than green vegetation. The training fields selected were the same as for 71054105 since the clusters generally contained the same fields and the same fields could be picked.

The classification results were then evaluated using these groupings. The percent correct recognition for the three corn groups is given in Table 14. The "other" categories were not evaluated.

Table 14. Test Accuracy for Blight Levels (Per Point) Run 71054103

LEVEL	NO. OF POINTS	PERCENT CORRECT	NO. CLASSIFIED INTO: LEVEL 1	LEVEL 2	LEVEL 3
1	17,388	66	11,476	2,476	2,253
2	11,212	50	4,382	5,419	1,253
3	4,500	45	1,903	300	2,019
Overall Percent Correct		57			

Channels 2, 4, 5, 7, 8, 9, 10, 11, 12 were used for the classification. The training statistics were also used to classify the test fields on a per field basis using the LARSYSAA per field classifier. The field classification results are presented in Table 15. These results are consistent with previous experience in which the per field classifier generally gave higher classification accuracy than the per point classifier.

Table 15. Test Accuracy of Blight Levels (Per Field) Run 71054103

LEVEL	NO. OF FIELDS	PERCENT CORRECT	NO. CLASSIFIED INTO: LEVEL 1	LEVEL 2	LEVEL 3
1	55	80	44	6	5
2	28	79	4	22	2
3	16	50	8	0	8
Overall Percent Correct		75			

Since the perfield classifier is many times faster than the per point, it was feasible to do a 12 channel classification for comparison, the results of which are presented in Table 16. The accuracies agree with previous experience: the classification accuracy is not drastically greater when all channels are used.

Table 16. 12 Channel Per Field Classification Results for 71054103

LEVEL	NO. OF FIELDS	PERCENT CORRECT	NO. CLASSIFIED INTO: LEVEL 1	LEVEL 2	LEVEL 3
1	55	80	44	6	5
2	28	79	4	22	2
3	16	50	8	0	8
Overall Correct		75			

It is beyond the scope of this study to search for the optimum channel set; it is assumed that the 9-channel set used is "near" optimum and that differences such as the 2-4% seen above are what separate the present results from the true optimum. A classification based on the GRPSAM results was performed as a final test of the "best" accuracy that could be achieved under the assumption that the true blight effects are described by the clustering results. The two similar blight groups (1 and 3) found by the GRPSAM program were combined as a level 1, or normal, group and group 2 was defined as the blighted group (blight mean of 2.6). A per point classification was run with this field grouping and the results are presented in Table 17. The fact that even with this optimistic grouping the "blighted" category was still only 50% correctly recognized further verifies previous result indications that these corn spectral classes are only about 50% separable from different corn classes and the "other" classes.

Table 17. Classification Results for 2 Corn Groups for Run 7105410[illegible] 9 channels, Per Point Classification.

CLASS	NO. SAMPS	PCT. CORN	NO. CLASSIFIED AS: CORN 1	CORN 2
Corn Level 1	30,[illegible]85	83	25,027	4,153
Corn Level 2	2,852	49	898	1,409
Overall		82		

Table 18. NSCLAS 6 cluster Resu-ts for Run 71054107 (Principal Components on 71054103) First 4 components used.

CLUSTER									
	2		3		4		5		6
FIELD	BLIGHT LEVEL	FIELD	BLIGHT LEVEL	FIELD	BLIGHT LEVEL	FIELD	BLIGHT LEVEL	FIELD	BLIGHT LEVEL
EEE-9	-	C-3		C-16	-	B-6		PPP-2	-
VVV-4	2.0	JJJ-3	2.5	TT-4	-	C-5	N1-T3	KKK-10	3.0
LLL-1	-	JJJ-2	2.8	NNN-6	-	c-6	N0-T1	QQQ-2	2.5
YY-7	-	JJJ-4	-	EE-9	2	ww-1	3.0	EEE-4	-
AAA-3	NOT2	OO-11	-	EE-15	2.5	a-27	2.5	QQ-6	-
AAA-5	1.5	EE-10	3.0	EE-16	1.8	uu-2	2.0	NN-2	-
GG-6	-	EEE-8	-	EE-8	2.5	vv-8	-	II-1	1.0
PP-6	-	NN-1	-	KKK-4	1.5	L-25	0	d-6	3.0
uuu-1	2.0	YY-4	-	JJJ-5	2.0			HH-2	-
		RR-5	2.0	EE-13	2.5			ZZ-7	2.0
		d-4	4.0	EE-12	3.0			ZZ-6	-
		ww-2	0	EE-11	3.0			c-3	3.5
		oo-4	1.5	KKK-11	3.0			ww-3	2.75T
		OO-2	2.0	PPP-1	0.0			ZZ-9	N0 T2
		AAA-10	NOT2	EEE-1	-			GG-4	-
		a-16	-	EEE-5	-			AAA-12	2.0
		uu-11	1.0	vvv-2	2.2			SS-6	3.0
		uu-15	3.0	RR-4				a-18	2.0
		uu-1A	-	YY-6				uu-10	3.25
		uu-6	2.0	d-8	3.5			uu-1B	-
		RRR-2	3.5	RR-3	1.5			RRR-1	2.0
		uu-8	3.0	RR-2	2.0			A-21	3.0
		SSS-1	0	c-8	0			a-24	3.5
				R-6	3.0			uu-4	1.5
				L-33	2.25			sss-4	.5
				w-3	4.0			N-2	4
								N-1	3.5
								L-9	NOT 1.5
MEAN	1.50		2.16		2.22		1.33		2.25
VAR	.75		1.4		.84		1.9		1.3

The results of such a classification as this are only relative since the data and analysis procedure is one sample out of many. The conclusion to be drawn from this analysis is that the apparent blight classes seen in the multispectral data are relatively inseparable. The split of about half of the cluster 5 fields into the low level clusters was due to the fact that about half of the level 3, high blight, fields were mixed with all other clusters appearing in them. The accuracy for the level 3 fields, which were relatively pure in containing cluster 5, was 63% for the 9 channel per point classification.

Analysis of Principal Components Transformed Data

As a final cluster analysis step, the data from run 71054103 was transformed by the principal components transformation, which in effect concentrates the information into a minimum number of channels. The statistics for all the corn fields in the segment were computed and used as a basis for the transformation. The data was transformed and run 71054107 was created. The principal components transformation is discussed in LARS Information Note 072071. The results of NSCLAS clustering using the first four principal components are presented in Table 18. The blight level cluster relationship is presented in Table 19.

Table 19. Blight Mean Cluster Relationship for Principal Component Data

CLUSTER	MEAN BLIGHT LEVEL	VARIANCE
2	1.5	.75
3	2.16	1.4
4	2.22	.84
5	1.33	1.9
6	2.25	1.3

These results indicated a two corn blight cluster situation in which clusters 2 and 5 are the lower blight level and clusters 3, 4, and 6 are the higher level groups. This result was judged to be very similar to the results for the GRPSAM clustering of run 71054103. There the "higher" blight level encompassed four groups averaging a blight level of 2.3 and the principal components produced three clusters having blight means averaging 2.2. The 71054103 GRPSAM

results produced two low level groups averaging 1.45. It was decided not to pursue the principal components further due to this close agreement. It was concluded that the corn field based principal components have results equivalent to GRPSAM results and that the classification results would be similar.

Summary

The analysis described here is cluster oriented rather than blight level oriented, i.e., spectrally separate groups of points were searched for. The blight level ratings of the fields rather than training class definitions, were used to assign meaning to the clusters. The clusters found tend to contain a variety of blight levels thus the spectral class is being determined by either blight effects or "other" effects. These other effects could be soil background, plant density, planting date, seed type, soil moisture, etc. Once the spectral class of a field is determined, a factor analysis should be run to determine the relationship between all these influences and spectral class. Then a reliability figure can be attached to the blight level inferred from the corn classified into any of the spectral classes.

The implications of the results presented here are that even if all the corn fields in cluster 5, for example, actually were at the most severe blight level, only 50% or so of them would be classified correctly. It is not the purpose of this information note to analyse the recognizability of corn blight. Rather a detailed example is presented here which describes the cluster oriented analysis procedure. It is intended as a step toward an optimum training procedure which if acheived will enable more rapid and automatic training of the multispectral pattern classifier.