Table 52--Classification matrix for September 21, 1972 imagery (MSS bands 4, 5, 6, and 7), using unequa prior probabilities in South Dakota test site.

Class	No. of sample points	Percent Correct	Number of samples classified into										
			Corn	Pasture	Oats	Barley	Rye	Alfalfa	Flax	Sudex	Idle	Fallow	Threshold
Corn.....	1060	0.1	1	753	275	3	0	0	3	0	12	10	3
Pasture..	812	88.4	1	718	86	1	0	0	0	0	2	4	0
Oats.....	243	40.3	0	142	98	0	0	0	0	0	0	3	0
Barley...	97	0.0	0	77	17	0	0	0	1	0	2	0	0
Rye......	16	0.0	0	15	1	0	0	0	0	0	0	0	0
Alfalfa..	303	0.3	0	243	51	0	1	1	0	0	0	6	1
Flax.....	71	4.2	0	45	23	0	0	0	3	0	0	0	0
Sudex....	55	0.0	0	47	7	0	0	0	0	0	0	1	0
Idle.....	18	10.5	0	14	3	0	0	0	0	0	2	0	0
Fallow...	82	4.9	0	59	17	0	0	0	0	0	2	4	0
Totals...	2758		2	2113	578	4	1	1	7	0	20	28	4

Overall performance 30.0%

Table 53--Classification matrix for September 21, 1972 imagery (MSS bands 4, 5, 6, and 7) using quadratic discriminant functions with unequal prior probabilities in South Dakota test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into					
			Corn	Pasture	Oats	Alfalfa	Sudex	Threshold
Corn.....	237	6.8	16	150	54	17	0	0
Pasture..	75	88.0	0	66	7	2	0	0
Oats.....	12	100.0	0	0	12	0	0	0
Alfalfa..	110	25.5	1	56	24	28	0	1
Sudex....	36	0.0	0	30	6	0	0	0
Totals...	470		17	302	103	47	0	1

Overall performance 26.0%

Table 54--Means and covariance matrices for crops in South Dakota on frame 1060-16491, September 21, 1972.

Crop	Means	Number	Covariance Matrix			
Corn	Means	Number 1060	Covariance Matrix			
	22.34		4.84			
	17.69		6.73	13.25		
	31.40		2.67	-0.42	33.40	
	19.38		0.37	-2.95	25.55	18.15
Pasture	Means	Number 812	Covariance Matrix			
	23.94		5.42			
	19.89		7.79	15.13		
	34.34		1.14	-1.48	29.59	
	20.85		-0.69	-3.78	18.72	13.99
Oats	Means	Number 243	Covariance Matrix			
	23.13		9.92			
	19.09		16.72	33.29		
	32.98		10.76	14.40	43.16	
	17.74		4.38	4.48	25.26	16.73
Barley	Means	Number 97	Covariance Matrix			
	24.52		5.47			
	21.46		6.25	11.15		
	30.07		5.93	5.41	25.70	
	17.51		2.65	1.54	16.87	12.53
Rye	Means	Number 16	Covariance Matrix			
	22.31		3.31			
	17.63		2.71	5.43		
	35.06		1.63	3.04	7.40	
	20.94		1.02	1.83	3.78	2.19
Alfalfa	Means	Number 303	Covariance Matrix			
	23.78		6.81			
	19.90		9.62	17.56		
	33.15		3.08	1.94	26.42	
	20.09		0.46	-1.61	16.19	12.25
Flax	Means	Number 71	Covariance Matrix			
	22.30		5.66			
	18.25		5.39	8.64		
	27.63		7.99	6.27	41.73	
	17.55		4.30	2.59	27.63	19.45
Sorghum	Means	Number 55	Covariance Matrix			
	22.51		2.79			
	17.25		3.00	6.60		
	32.15		1.44	-1.97	23.04	
	20.05		0.42	-2.38	15.76	12.74

Table 54 continued

			Covariance Matrix			
Idle	Means	Number 19				
	23.05		9.86			
	19.00		14.74	26.62		
	31.58		7.79	5.45	27.88	
	19.63		0.43	-3.92	14.94	11.90
Winter	Means	Number 82	Covariance Matrix			
Fallow	23.41		5.47			
	19.78		9.58	20.70		
	32.21		-1.27	-5.75	36.24	
	19.27		-2.77	-7.65	20.93	14.59

Idaho:

The test site in Idaho covers nearly four counties. The Crop Reporting District boundaries were bypassed because they did not include some areas of homogeneous types of agriculture that should have been included. Figure 4 shows the test site area.

The results are based on 42 segments in the intensive agriculture stratum in one LANDSAT frame. Two additional segments are not on this frame. The frame that contains these two segments also contains ten segments which are on the first frame. Therefore, it may be possible to use this overlapping data to calibrate from one frame to the next, or to measure the difference due to frames in the means and variance for the overlapped data. A method of using calibration or training data in one frame to adjust parameters or to classify on another frame would be valuable, since, it would increase the value of the segment data. A crop may be different over a large area because of variety, soil type, weather conditions, and state of maturity rather than technical factors associated with acquiring imagery. However, it may be possible in some areas to do signature extension and this problem should be investigated.

The data had serious banding problems. The problems seem to be most apparent in band 5, therefore, that band was left out of the first classification. Table 55 shows this first classification.

Obviously, the classification is not as good as we expected; however, by chance, one would expect only 8% correct classification for 12 crop categories. Another possible problem with the classification is that some field boundaries, sometimes, fall on adjacent points and since the pixels are partially overlapping, these border pixels may be causing some overlap of the crop categories. The grey-scale printout (Figure 14) which follows illustrates this problem.

Figure 14--Gray scale printout of a segment showing how fields are defined.

LANDSAT Column Number

Table 55--Preliminary classification of Idaho study area data using August 1972 data bands 4, 5, and 7 and unequal prior probabilities.

	No. of samples	Percent Correct	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	IDLE	OHAY	PASTURE	SUGBTS	POTATOES	SPWH
Peas and Beans	579	14.5	84	45	1	31	0	0	0	0	327	89	2	0
Harvested Beans	784	71.1	13	562	45	8	0	0	0	0	152	4	0	0
Barley	1019	11.5	33	271	117	27	0	2	6	0	489	64	10	0
Alfalfa	1318	17.3	57	51	2	228	0	0	6	0	527	422	25	0
Corn	542	0.0	10	21	9	119	0	0	0	0	221	161	1	0
Fallow and Other	684	0.4	14	13	3	14	0	3	33	0	575	26	3	0
Idle	206	26.7	4	10	0	1	0	1	55	0	135	0	0	0
Other Hay	11	9.1	0	0	0	0	0	0	0	0	5	3	2	0
Pasture	1484	80.7	38	25	4	78	0	2	49	1	1197	83	8	0
Sugar Beets	527	76.5	12	5	1	43	0	0	6	0	46	403	10	0
Potatoes	533	10.1	29	2	1	80	0	0	0	0	89	278	54	0
Spring Wheat	111	0.0	3	48	3	5	0	0	0	0	49	3	0	0
Total	7798		297	1054	186	634	0	8	155	1	3812	1536	115	0

Overall performance 34.7 percent

It is obvious that many groups are very similar, and therefore, misclassification is high. We will try combining several into groups based on similarity of the estimated parameters, since these initial results indicate a number of crops are not distinct.

The next classification matrix uses equal prior probabilities and is presented in Table 56. The overall classification performance is 21.8%. This points out that prior information in terms of probabilities is also important in this test area.

Since the data had serious banding problems, it was thought that perhaps this caused the extremely poor classification rates. As a result, NASA Goddard was asked to reprocess the image to remove the banding.

The image was reprocessed at considerable expense to Goddard and the classifications were again run. The results are shown in Table 57.

Table 58 is a result of combining classes after classification. It is obvious that going to fewer categories does improve the classification. However, in Idaho, where many crops are grown, the imagery must contain information that will allow users to separate the various crops. Perhaps, temporal information would improve the value of the Idaho imagery.

Results of Classification of Aerial Photography

Since aerial photography is in image form and computer techniques require digital data, it is necessary to convert the photographs to optical densities. A detailed explanation of how this is done may be found in Appendix D. The aerial photography was scanned by a Photometric Data System (now Bolen ahd Chivens) microdensitometer. This instrument records optical densities (or transmissions) of wavelengths of light corresponding to given color filters. Each time a filter is changed, however, the instrument must be recalibrated. The values recorded range from 0.00 to 4.00 in optical density. In brief, the range of values is spread between the chosen calibration point and total darkness.

Initially, the procedure for scanning segments was as follows:

An interval point within the photograph was chosen. This point was the considered lightest spot on the exposed portion of film and it was set at 0.00 on the microdensitometer scale. South Dakota, Kansas, and Idaho photography was scanned and the results brought to light problems in this technique.

Table 56--Preliminary classification of Idaho study area data using August 1972 data bands 4, 5, and 7 with equal prior probabilities.

	No. of samples	Percent Correct	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	IDLE	OHAY	PASTURE	SUGBTS	POTATOES	SPWH
Peas and Beans	597	25.6	148	43	1	29	19	26	109	96	12	25	59	12
Harvested Beans	784	66.1	20	518	40	15	4	18	50	7	8	1	14	89
Barley	1019	9.9	62	214	101	13	19	66	112	59	71	14	78	210
Alfalfa	1318	10.7	119	47	11	141	51	26	80	172	108	115	428	20
Corn	542	1.7	28	18	11	62	9	41	36	56	17	41	198	25
Fallow and Other	684	12.1	23	7	6	5	7	83	416	23	33	5	35	41
Idle	206	70.4	9	4	0	1	1	24	145	3	4	0	0	15
Other Hay	11	72.7	1	0	0	0	2	0	0	8	0	0	0	0
Pasture	1484	8.0	105	15	17	70	14	117	606	54	119	36	148	183
Sugar Beets	527	19.9	3	3	2	18	8	0	8	142	4	105	226	8
Potatoes	533	56.8	10	2	2	25	6	1	4	105	2	72	303	1
Spring Wheat	111	19.8	8	38	0	10	4	6	4	8	5	1	5	22
Total	7798		536	909	191	309	144	408	1570	733	383	415	1494	626

Overall performance 21.8 percent

Table 57--Classification matrix of Idaho Study Area, August 1972 Imagery Using MSS Bands 4, 5, 6, and 7, with unequal Prior Probabilities.

	NO. OF Samples	PERCENT CORRECT	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	PASTURE	SUGBTS	POTATOES	SPWH
Peas and Beans	549	40.6	223	6	9	23	4	61	123	94	5	1
Harvested Beans	813	62.6	19	509	106	11	1	38	121	6	0	2
Barley	957	75.9	68	108	248	65	9	83	331	36	6	3
Alfalfa	1314	29.8	192	30	34	391	30	32	331	250	23	1
Corn	541	8.5	42	13	20	106	46	52	186	69	8	4
Fallow and Other	77 779	37.4	28	1	7	31	3	291	412	3	3	0
Pasture	1433	64.0	107	8	24	115	8	218	917	34	2	0
Sugar Beets	386	56.0	19	1	5	60	8	1	30	216	45	1
Potatoes	395	21.8	15	0	0	115	7	0	92	80	86	0
Spring Wheat	104	3.8	12	27	24	4	1	3	23	4	2	4
Total	7271		725	703	477	921	117	779	2566	787	180	16

Overall performance 40.3 percent

Table 58--Classification matrix of Idaho with unequal prior probability groups - Table 57 collapsed into 7 groups.

Group	No. of samples	Percent Correct	Beans	Small Grains	Corn	Fallow	Pasture	Sugar Beets	Potatoes
Beans...	1362	55.6	757	118	5	99	278	100	5
Small Grains..	1061	26.3	215	279	10	86	423	40	8
Corn....	541	8.5	55	24	46	52	287	69	8
Fallow..	779	37.4	29	7	3	291	443	3	3
Pasture.	2747	73.0	337	59	38	250	1754	284	25
Sugar Beets...	386	56.0	20	6	8	1	90	216	45
Potatoes	395		15	0	7	0	207	80	86
Totals..	7271		1428	493	117	779	3482	792	180

Overall performance 47.2 percent

It was observed that each segment had a different calibration point (lightest spot), hence, there were variations in the scanning results. As a calibration point changed, grey level readings for the same crop in a variety of segments, were different. In fact, when the same segment was scanned twice using two different calibration (light) spots, the crop signatures might not appear similar.

To overcome this defect, a new calibration technique was developed. Emphasis was placed on choosing calibration points which would produce identical results in every segment. The procedure was to focus on the clear, plastic circle which appears on each section of the film as the scanner passes across the image. This circle became 0.00 in every instance. Consequently, reliable crop data was acquired since all calibration factors were now constant in the scanning process. The state of Missouri was scanned using this improved method and the results were found to be more accurate.

Once the data has been scanned, it must be labeled for crop type. Tract and field numbers were provided by the use of a coordinate system and this data was then merged with the ground observation data. This provided crop labels. This labeled data can then be used for both computer training and testing information.

The classification procedure is explained at the beginning of this section. However, since the calibration was done using local calibration points, the classifier training and the computer evaluations were performed in two ways.

For example: 1. All data was pooled and used for both the training and testing.

2. All data in each segment was used for both training and testing one at a time.

The results were then pooled (matrix sum). The prior probabilities in each case were proportional to the training data and since this training data was used to test, it too, was proportional to the data being classified.

In the instance of the pooled training, the prior probabilities were the same for each segment. When interpreting the local training, the prior probabilities were different for each segment and depended on the data in each. For the local training, all conditions were optimal which would mean that the classification accuracy is maximal.

As a preliminary check on the effect that the different calibration points had on the data, a cluster analysis on all data was run. The means for each field were computed by segment and crop. These means were clustered using a program written by C.T. Zahn of Stanford University. 1/ The fact that the means clustered by segments rather than by crops was additional proof of the problems which had arisen because of calibration differences.

Figure 15 provides an overall state by state comparison of classification accuracy. Figure 16-19 summarize the percent correct classification for major crops in each state. These figures compare both methods of training on the same data sets; the difference lies in the results of local versus pooled training data.

Tables 59-65 give the classification matrix for both methods of training. When local training data was used for training, a classification matrix was available for each segment. These segment classification matrices were summed to obtain the final classification matrices in this report.

1/ C.T. "Zahn, Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters," IEEE Transactions on Computers, Vol. C-XX, No. 1, 1971.

Figure 15--Comparison of overall percent classification by states, 1972. (///Slashes indication global classification).

Figure 16--Comparison of classification methods by crop, Kansas August 18, 1972. (///Slashes indicate global classification.

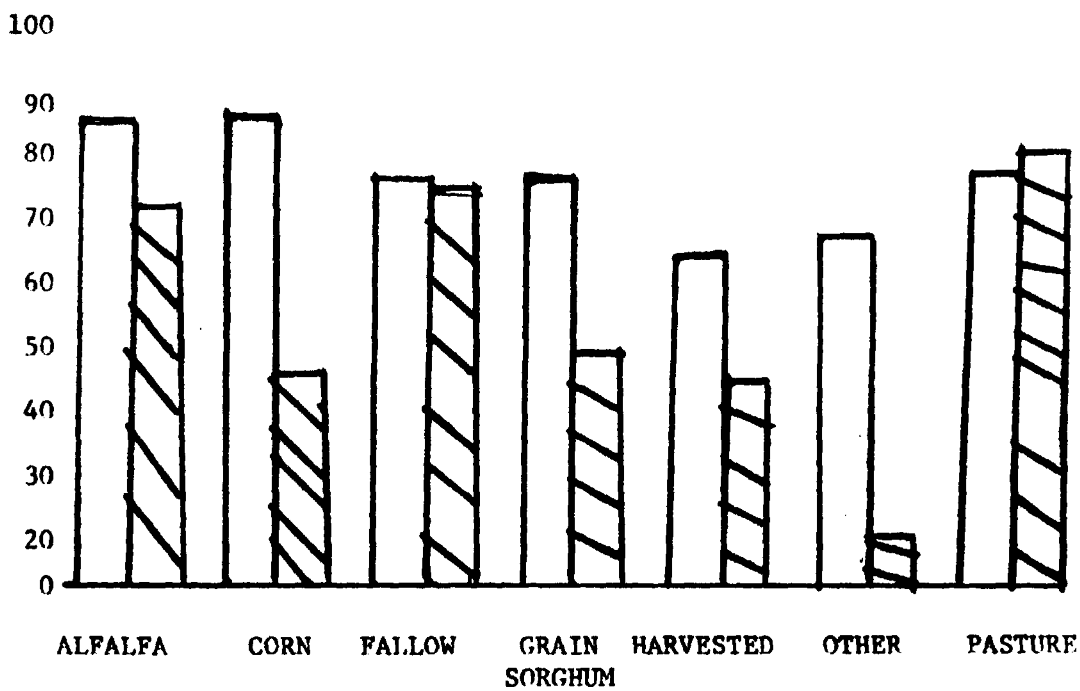


Figure 17--Comparison of classification methods by crop, Missouri, August 29, 1972.
(/// Slashes indicate global classification).

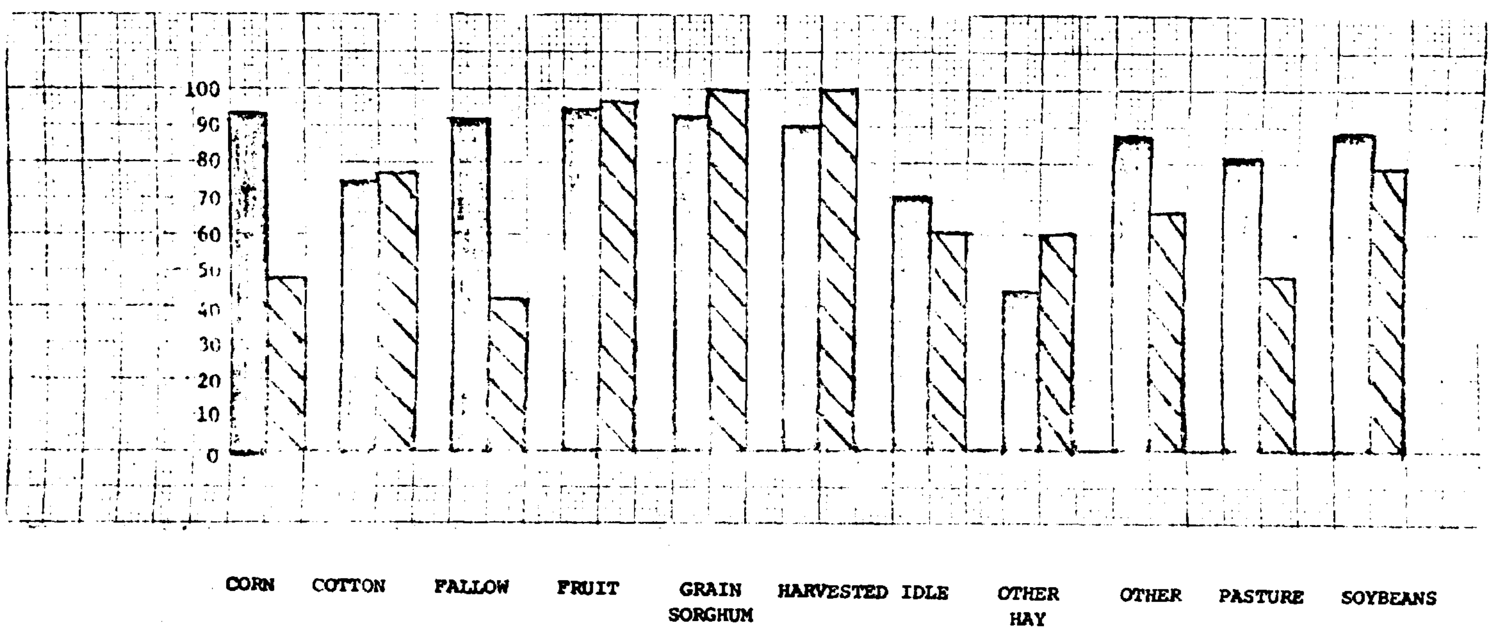


Figure 18--Comparison of classification methods by crop, South Dakota, September 23, 1972. (/// Slashes indicate global classification).

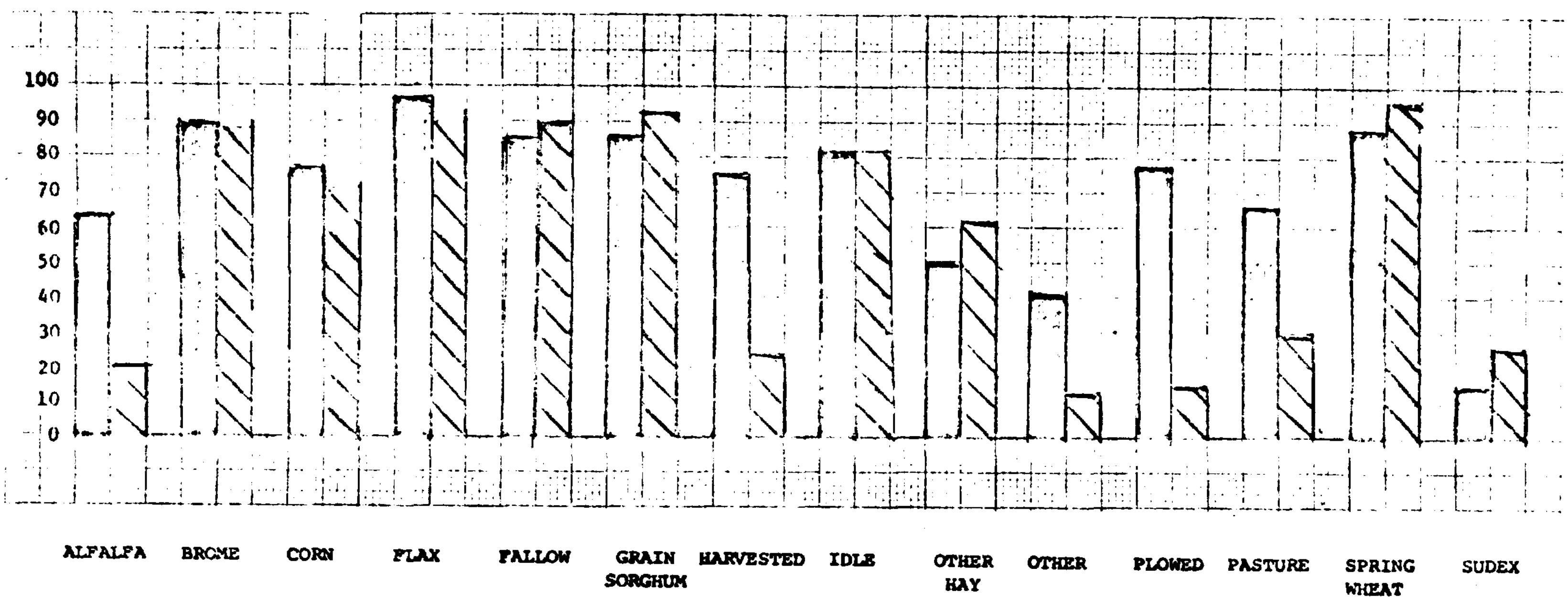


Figure 19--Comparison of classification methods by crop, Idaho, August 12, 1972.
(/// Slashes indicate global classification)

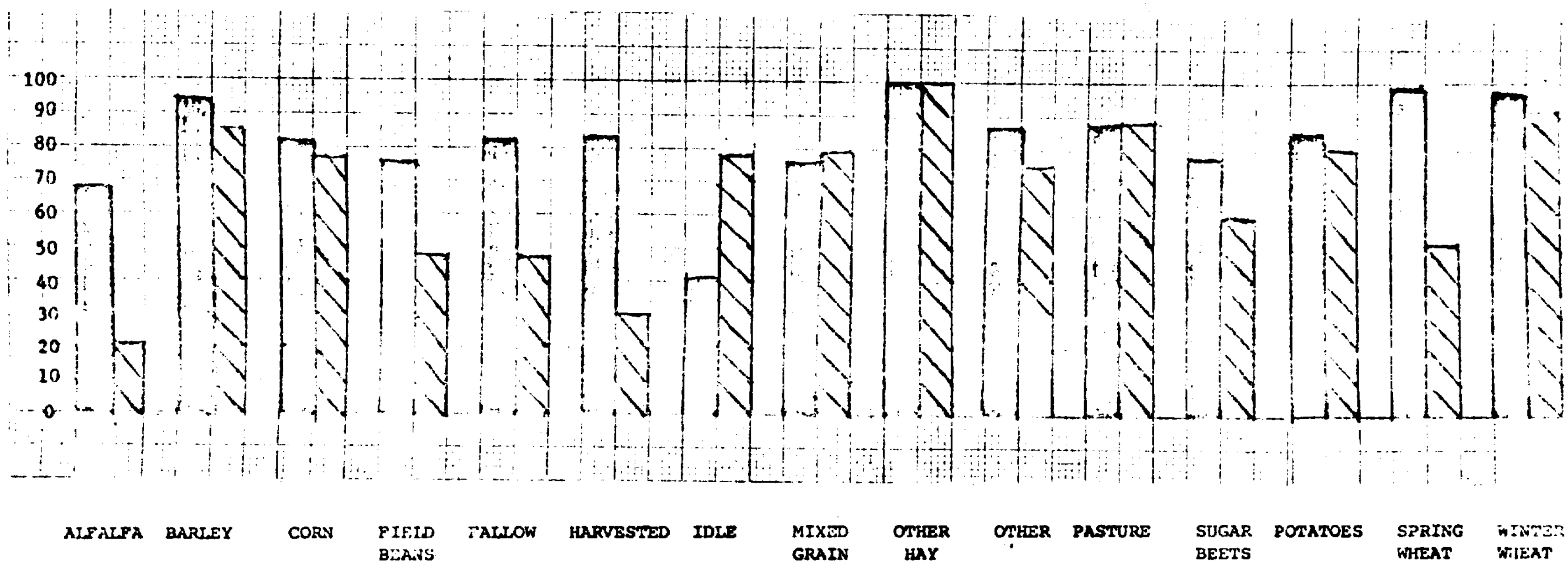


Figure 15-19 indicate that the training by segment (local) gave higher classification percentages. The differences in classification percentages can be attributed to three sources:

1. difference in calibration of data when scanned by a microdensitometer.

2. differences in the number of crop classes and the prior probabilities of each crop class.

3. differences in the variability of a local versus pooled data set.

Interpretation of Figure 15 is quite easy. Kansas (with some calibration effect and only seven crop classes) was not greatly affected by the calibration effects. However, South Dakota, with calibration differences and many crop classes was drastically effected. In Missouri, differences were slight between classification results comparing local training versus pooled training.

Table 59--Classification of flightlines 3 and 10, by segment, using quadratic discriminant functions on all eight spectral variables, Kansas aircraft data, September 1972.

Crop	Percent Correct	ALFA	CORN	FLOW	GSOR	HARV	OTHR	PSTR
ALFA	94.2	1238	0	0	21	8	36	11
CORN	93.9	0	247	0	2	3	11	0
FLOW	80.9	0	0	8383	398	1432	47	100
GSOR	82.2	4	1	51	3525	181	26	498
HARV	66.0	0	0	1031	489	4797	29	922
OTHR	70.1	37	4	13	17	14	312	48
PSTR	83.2	18	0	697	2927	1677	70	26,644
OTHERS		525	129	3186	1606	3290	829	7,166

Overall performance 80.7 percent

Table 60--Classification of flightlines 3 and 10, on all eight spectral variables, Kansas aircraft data, September 1972.

Crop	Percent Correct	ALFA	CORN	FLOW	GSOR	HARV	OTHR	PSTR
ALFA	76.0	999	14	0	107	4	20	170
CORN	41.8	170	147	0	19	0	2	14
FLOW	78.6	8	52	8141	352	708	9	1090
GSOR	46.5	102	107	81	2063	485	62	1540
HARV	41.6	16	42	1816	403	3025	26	1940
OTHR	10.6	43	5	21	126	37	47	166
PSTR	87.6	3	241	222	916	2400	180	28,071
OTHERS		335	596	2023	1402	1462	660	10,728

Overall performance 75.6 percent

Table 61--Classification of flightlines 5 and 6, by segment, using all eight spectral variables, Idaho, September 1972.

Crop	Percent Correct	Number of samples classified into														
		ALFA	BRLY	CORN	FLDB	FLOW	HARV	IDLE	MGRN	OHAY	OTHR	PSTR	SBTS	SPDS	SPWH	WNWH
ALFA	68.3	808	0	13	5	1	7	0	2	0	4	218	84	16	25	0
BRLY	94.6	2	866	0	0	5	5	8	2	0	2	5	5	3	0	12
CORN	82.6	17	4	1171	67	19	10	2	1	1	13	40	67	1	2	0
FLDB	76.3	0	0	7	371	0	0	0	0	0	1	2	40	65	0	0
FLOW	83.1	2	11	98	0	1069	74		1		4	4	16	8	0	0
HARV	83.9	1	9	0	0	26	620	1	0	1	35	27	18	1	0	0
IDLE	41.0	0	15	9	5		5	34		0	0	0	15		0	0
MGRN	76.0	6	7	1	0	6	0		76		1		2		1	
OHAY	1.00	0	0	0	0		0	0		9	0	0	0		0	0
OTHR	87.0	1	2	1	0	6	4		4	1	631	15	51	2	7	0
PSTR	87.9	71	5	35	4	30	59	0	0	0	32	2950	145	26	0	0
SBTS	78.5	0	5	60	136	17	22	7	0	0	6	5	955	3	0	0
SPDS	85.2	1	76	11	2	8	50		0		4	18	1	981	0	
SPWH	99.0	1	0	0	0	0	9	0	0	0	2	0	0	0	410	1
WNWH	97.3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	36
Others		813	1016	703	623	14166	1419	129	302	36	5104	788	788	454	603	17

Overall % = 83.76

Table 62--Classification of flightlines 5 and 6, using eight spectral variables, Idaho, September 1972.

Crop	Percent Correct	Number of samples classified into														
		ALFA	BRLY	CORN	FLDB	FLOW	HARV	IDLE	MGRN	OHAY	OTHR	PSTR	SBTS	SPDS	SPWH	WNWH
ALFA	21.39	253	8	117	1	67	0	4	1	6	29	652	25	11	1	8
BRLY	86.12	1	788	0	0	5	9	35	56	0	7	5	0	0	3	6
CORN	77.74	15	6	1100	19	5	10	21	0	1	7	197	32	2	0	0
FLDB	48.35	0	1	118	235	0	11	0	0	0	0	39	55	27	0	0
FLOW	47.16	2	83	222	21	607	33	4	15	3	199	92	1	5	0	0
HARV	30.72	2	98	59	8	73	227	15	0	39	37	145	1	34	1	0
IDLE	78.31	0	0	12	0	0	0	65	0	0	4	0	2	0	0	0
MGRN	79.00	3	9	0	0	0	0	0	79	0	0	8	0	0	0	1
OHAY	100.00	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0
OTHR	75.45	0	20	3	0	16	4	6	3	2	547	114	5	1	4	0
PSTR	88.35	61	6	66	0	41	37	0	2	1	123	2966	1	53	0	0
SBTS	59.87	50	1	135	66	1	8	55	0	0	3	168	728	1	0	0
SPDS	79.95	2	89	13	71	5	8	0	0	0	2	25	5	921	11	0
SPWH	51.93	0	78	0	0	20	2	0	0	5	5	35	0	0	215	54
WNWH	91.89	0	2	0	0	1	0	0	0	0	0	0	0	0	0	45
Others		312	1513	844	453	599	919	267	640	73	4904	4918	269	316	174	52

Overall % = 66.89

Table 63--Classification of flightlines 2 and 5 by segment, using eight spectral variables, South Dakota, September 1972.

Crop	Percent Correct	Number of samples classified into													
		ALFA	BRME	CORN	FLAX	FLOW	GSOR	HARV	IDLE	OHAY	OTHR	PLOW	PSTR	SPWH	SUDX
ALFA	63.57	719	53	238	0	0	0	24	0	6	14	3	66	0	8
BRME	89.20	15	760	35	0	0	0	28	0	0	0	0	14	0	0
CORN	77.35	313	39	2681	0	1	0	81	10	23	9	43	248	1	18
FLAX	97.57	0	0	0	201	0	0	5	0	0	0	0	0	0	0
FLOW	84.87	1	0	0	0	101	0	3	0	0	0	14	0	0	0
GSOR	85.71	0	0	0	0	0	12	0	0	0	0	2	0	0	0
HARV	75.51	37	117	141	0	14	3	1421	35	6	2	48	43	23	2
IDLE	82.73	3	0	10	0	0	0	5	206	0	0	0	24	0	1
OHAY	50.57	2	0	5	1	0	0	29	0	44	2	4	0	0	0
OTHR	41.35	37	16	81	2	6	0	24	3	19	208	13	94	0	0
PLOW	78.38	8	0	12	0	77	2	85	0	4	11	747	7	0	0
PSTR	66.55	266	141	421	0	0	1	151	31	23	38	20	2246	7	12
SPWH	88.46	0	0	2	0	0	0	1	0	0	0	0	0	23	0
SUDX	15.42	135	0	21	0	0	0	3	0	3	2	15	24	0	37
Others		964	213	2305	117	276	7	1535	195	160	2123	1220	1658	14	77

Overall % correct = 71.73

Table 64 --Classification of flightlines 2 and 5, using eight spectral variables, South Dakota, September 1972.

Crop	Percent Correct	Number of samples classified into													
		ALFA	BRME	CORN	FLAX	FLOW	GSOR	HARV	IDLE	OHAY	OTHR	PLOW	PSTR	SPWH	SUDX
ALFA	16.45	186	196	412	0	10	0	24	13	6	11	3	208	0	62
BRME	88.03	28	750	7	0	0	0	19	1	3	14	3	27	0	0
CORN	71.18	138	371	2467	2	59	0	79	55	44	10	9	197	12	23
FLAX	90.78	0	4	2	187	0	0	3	0	0	2	2	6	0	0
FLOW	89.92	0	6	1	0	107	0	1	0	0	0	4	0	00	0
GSOR	92.86	0	1	0	0	0	13	0	0	0	0	0	0	0	0
HARV	24.05	68	369	381	53	58	0	455	48	35	15	75	258	42	26
IDLE	82.33	0	0	31	0	4	0	0	205	0	0	41	1	0	4
OHAY	62.07	1	2	17	0	1	0	0	0	54	2	1	7	0	1
OTHR	13.52	43	34	63	4	25	0	11	22	39	68	111	77	0	6
PLOW	15.63	15	287	41	4	296	6	32	56	21	26	149	20	0	0
PSTR	30.58	126	456	1320	24	11	1	77	97	74	60	27	1032	25	45
SPWH	96.15	0	1	0	0	0	0	0	0	0	0	0	0	25	0
SUDX	25.83	20	1	26	0	5	0	0	16	0	5	3	2	9	62
Others		666	897	3610	155	563	8	307	508	354	1386	407	1763	34	206

Overall % correct = 44.54

Table 65---Classification of flightlines 2 and 8, by segment, using eight spectral variables, Missouri, September 1972.

Crop	Percent Correct	Number of samples classified into										
		CORN	COTTON	FLOW	FRUIT	GSOR	HARV	IDLE	OHAY	OTHR	PSTR	SOYB
CORN	93.69	104	0	1	0	0	4	0	0	0	0	2
COTTON	75.62	6	273	0	0	1	7	25	0	2	14	33
FLOW	90.91	0	0	30	0	1	0	0	0	0	1	1
FRUIT	95.24	0	0	0	41	0	0	0	0	0	1	1
GSOR	92.86	1	0	0	0	13	0	0	0	0	0	0
HARV	89.06	1	0	0	0	0	57	5	0	0	0	1
IDLE	71.60	0	2	1	0	0	11	232	0	28	42	8
OHAY	44.28	19	0	8	0	0	0	0	182	2	33	167
OTHR	87.47	3	2	1	0	1	0	3	0	342	10	29
PSTR	81.46	10	0	12	6	6	0	2	4	6	268	15
SOYB	88.16	6	51	5	3	5	11	3	5	26	44	1184
Others		154	200	59	68	47	61	228	217	695	485	1469

Overall % correct = 79.61

The microdensitometer can scan a photograph and obtain either density values or transmission values or both. Transmission values are functionally related to density readings by the following equation:

$$\text{Density} = \log \left(\frac{1}{\text{transmission}}\right)$$

Theoretically, all information would be contained in either mode and neither would add anything new to the data. However, in practice, this does not hold true for two reasons:

1. The scanner seems to saturate. The results of this saturation affects density measurements. It becomes difficult to differentiate between brown wheat, brown hay, harvested grains, and bare soil. When the sensor saturates, it gives similar readings even though the colors are quite different. In the use of the transmission values, correct classification is increased but lacks complete reliability.

2. An additional reason for the one mode preference is concerned with the computer operation. The computer algorithm assumes that the data is multivariate normal with equal covariance matrices. Certainly if the data was multivariate normal in the measurement space using density values, it would not be multivariate normal after it had been transformed by a reciprocal of the log transformation. Obviously, they could not both be distributed as multivariate normal data. Thus, it is imperative to investigate the effects of variables on classification groups.

A stepwise discriminant analysis was performed on the training data in South Dakota. The procedure used was program BMD07M of the BMD statistical package. 1/ This program performs a stepwise linear discriminant analysis with proportional group priors on the training data. Variables are entered or deleted from the discriminanting set based upon an F-test of group differences for a particular variable. The variable that has the largest pairwise group F-value is the first variable entered in the discriminant set. This procedure was executed on all eight variables and then upon the subsets of variables corresponding to transmission and density scanning mode respectively. Some of the original nine groups were pooled to eight, six, and, then, four groups and the stepwise classification was performed on the merged groups. The mergers are as follows:

1/ Biomedical Computer Programs. W. H. Dixon, Editor. Berkeley, California; University of California Press, 1973.

a) For the eight groups; "Harvested Grains" and "Harvested Row" were merged to form the classification group "Harvested.

b) For the six groups: Hay, Pasture, and Fallow were merged (a grasses type of cover), in addition to the above merger.

c) For the four groups: Corn, and Soybeans formed a group, Wheat, Pasture, and Harvested Grains formed a group, Plowed, Fallow, and Harvested Row formed a group. This merger in the above groups was a result of a cluster analysis performed on the group means.

The option was specified for the inclusion with no deletion of variables at each step in performing the stepwise discriminant analysis. Thus, supposedly one is adding more information (in the form of more variables) at each step of the stepwise discriminant analysis. The results are astonishing as we can see in Figures 20-27.

Note the following:

(1) Overall percent correct classification increases only slightly, when two variables are in the discriminanting set, irrespective of what variables are used and what the classification groups are. The contention of C. R. RAO 1/ that more variables do not necessarily mean more information and hence more discriminanting power is supported by the data.

(2) The addition of a particular variable influences one classification group greatly. For example, in Figure 20, the variable TGREEN (transmission in GREEN) has a great effect on the classification accuracy of wheat when combined with DRED (density red), and DCLEAR. However, once TRED,has entered,TGREEN's affect is diminished by the confusion variable TRED.

(3) All discriminating information is not contained in one scanning mode (four variables). For example, compare the classification curves for the group FALLOW in Figures 24 and 25 respectively. Fallow was correctly classified about 65 percent of the time when scanned in density mode but had zero recognition in the transmission mode.

(4) The overall classification accuracy increases as the number of groups is decreased, See Figures 30 and 31.

1/ Covariance Adjustment and Related Problems in Multivariate Analysis by C. R. RAO in Multivariate Analysis, editor P. R. Krishnaiah, Academic Press, 1966.

(5) The overall classification accuracy is greater for variables measured in density units, and the use of the variables measured in transmission does not improve the overall classification when only four variables are considered. This can be seen in Figures 20, 23, 26, 29.

(6) This analysis leads one to conclude that if there is interest in only one or perhaps several crop groups that a hierarchial (or layered) classifier might be the best approach to crop identification. At each stage of the hierarchy, a feature selection would be performed to maximize the particular crop or crops of interest.

A single stage classifier with all variables used clearly would not do well on the major crop Wheat in Figure 20, as evidenced by the last stage of the stepwise discriminant analysis.

FIGURE 20

Stepwise discriminant analysis, classification into nine groups, density and transmission mode, South Dakota, 1972.

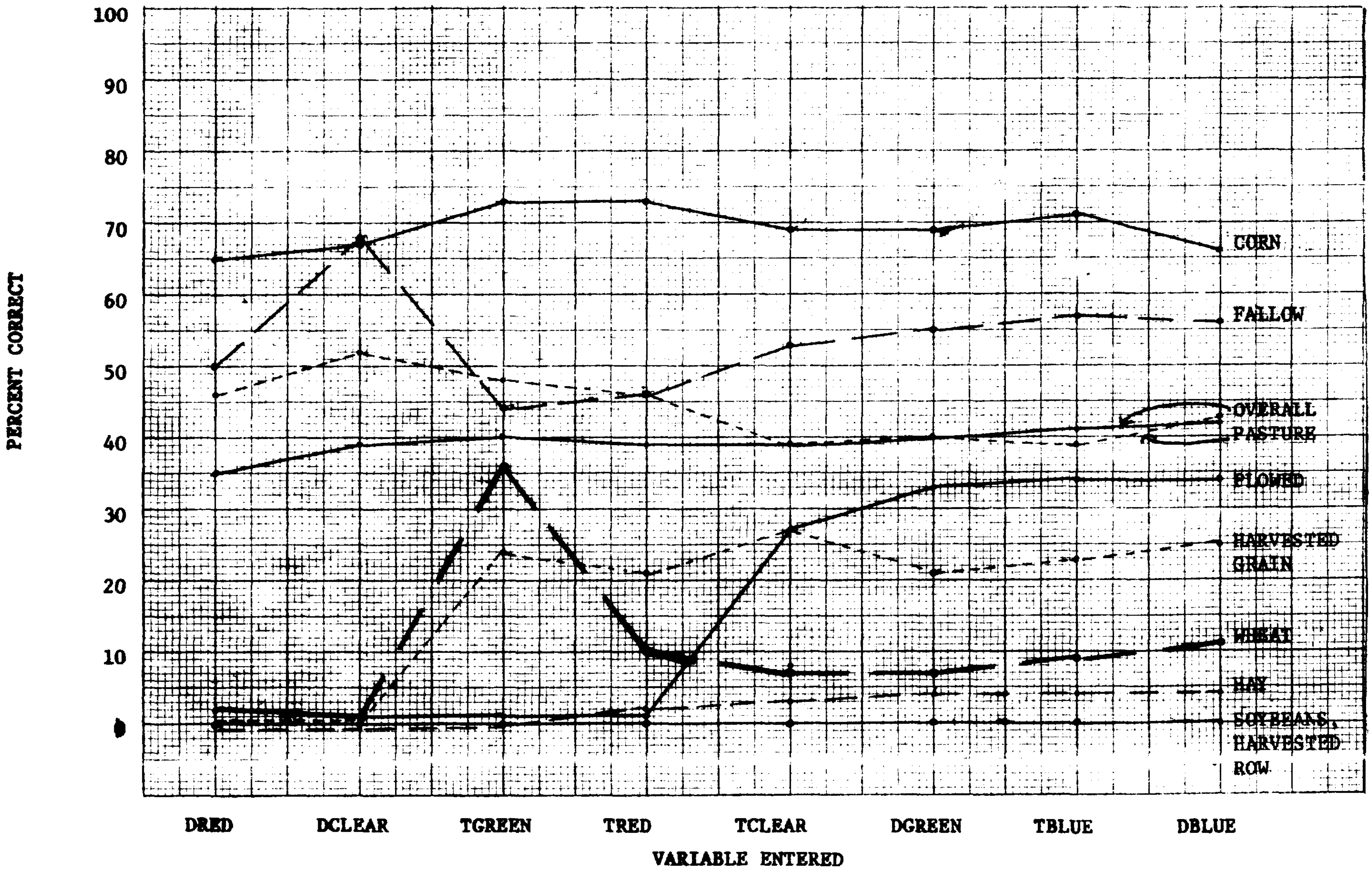


FIGURE 21

Stepwise discriminant analysis, classification into nine groups, transmission scanning mode, South Dakota, 1972.

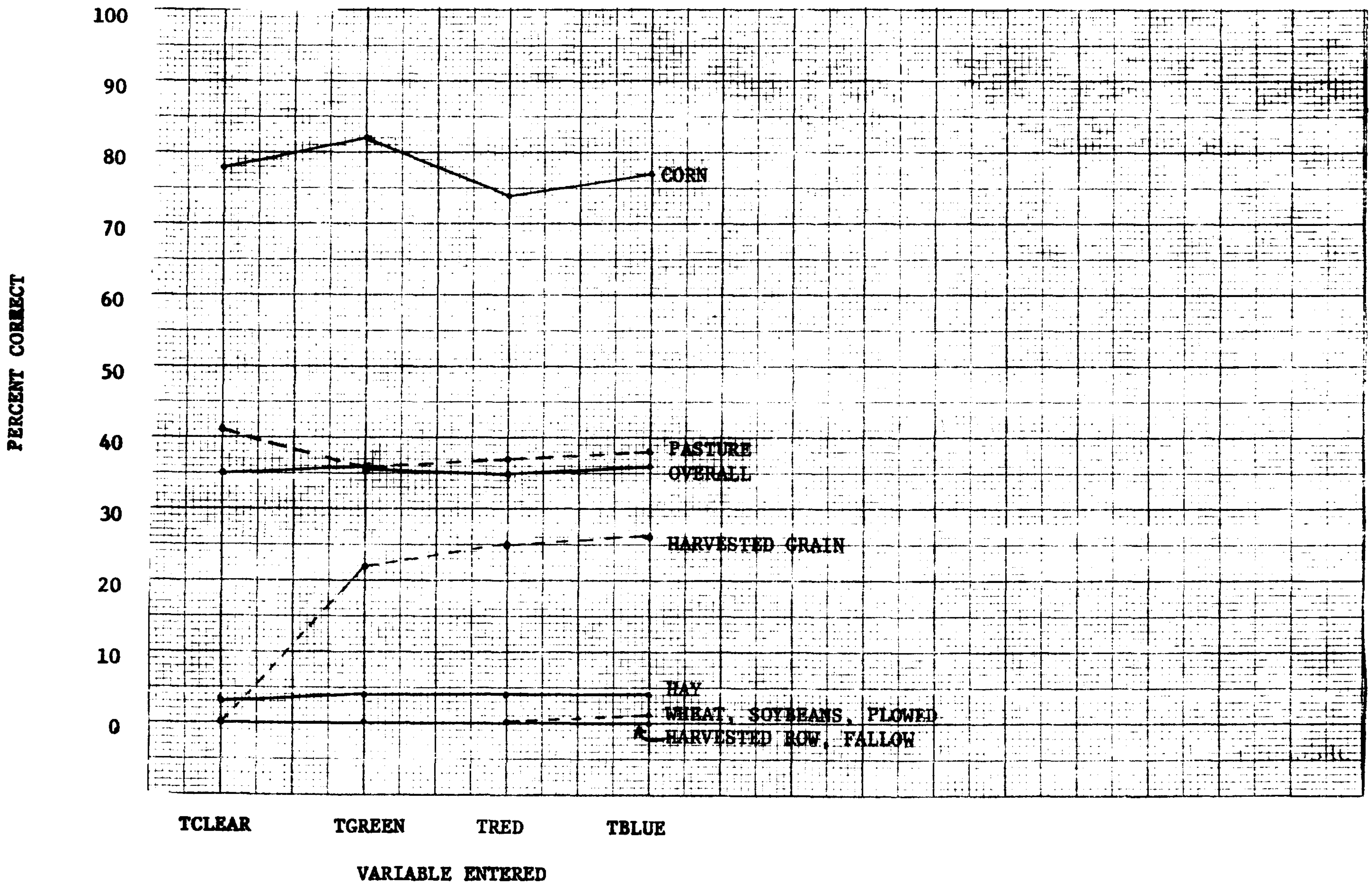


FIGURE 22

Stepwise discriminant analysis, classification into nine groups, density scanning mode, South Dakota, 1972.

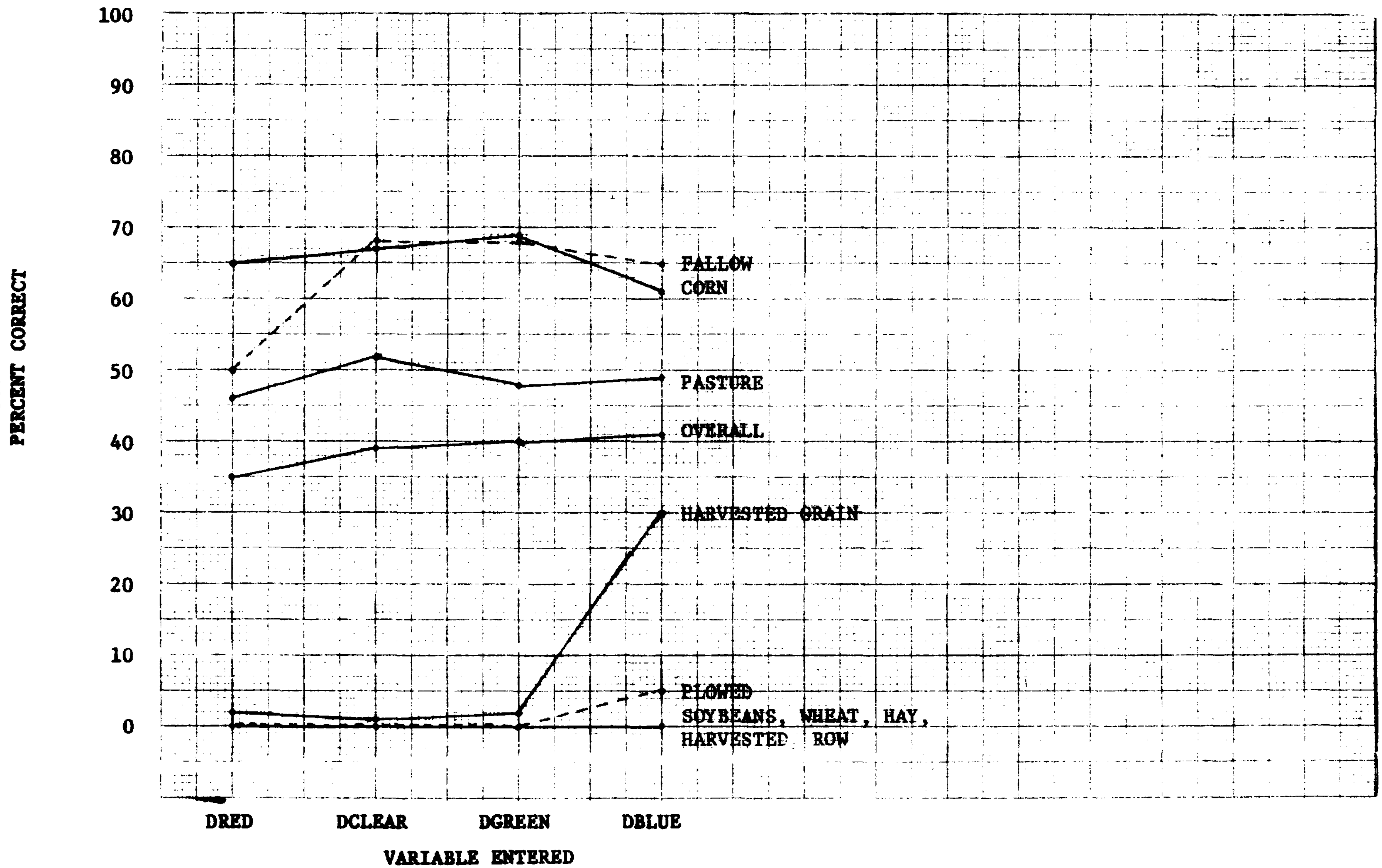


FIGURE 23

Stepwise discriminant analysis, classification into eight groups, all variables, South Dakota, 1972

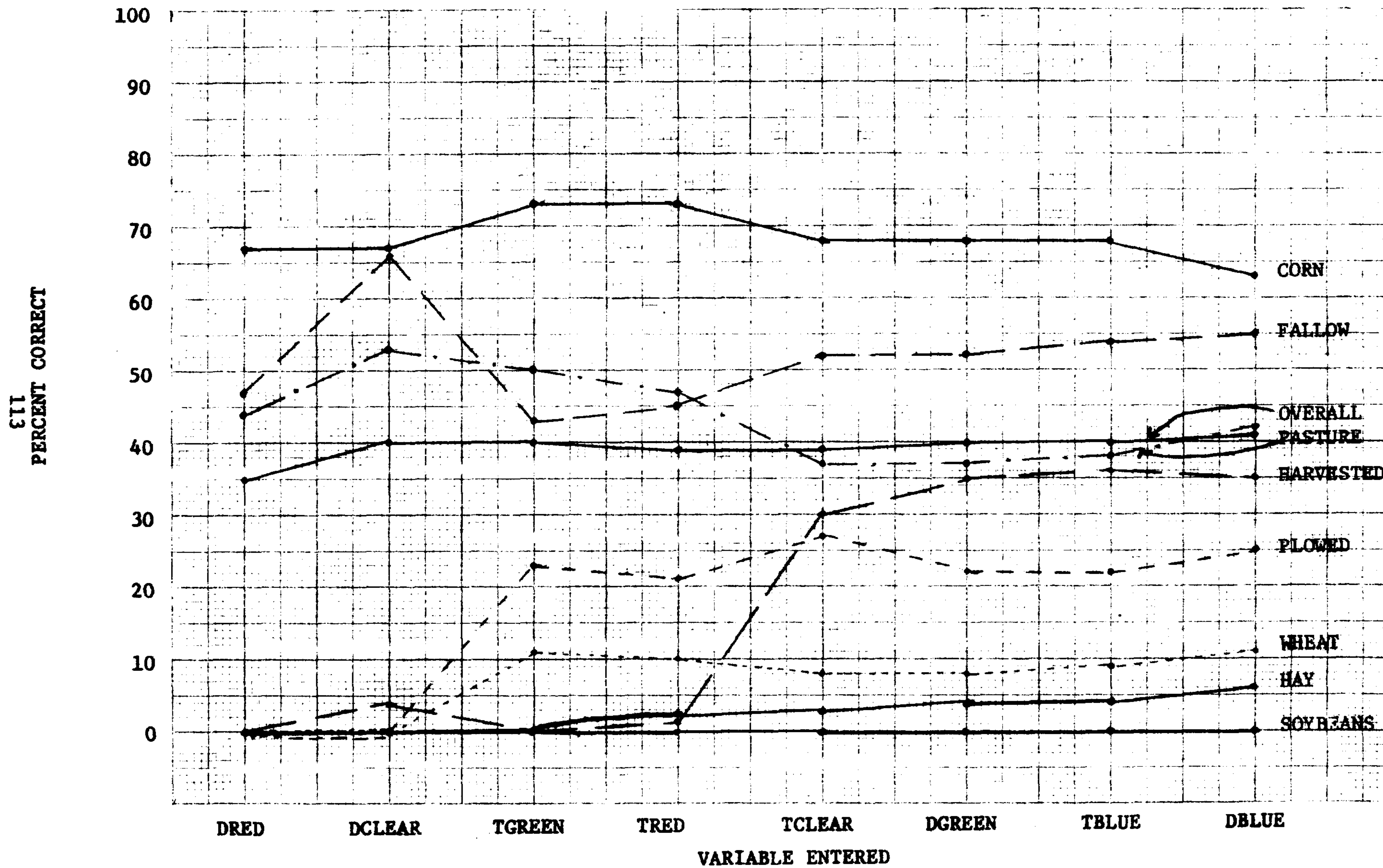


FIGURE 24

Stepwise discriminant analysis, classification into eight groups, transmission mode, South Dakota, 1972.

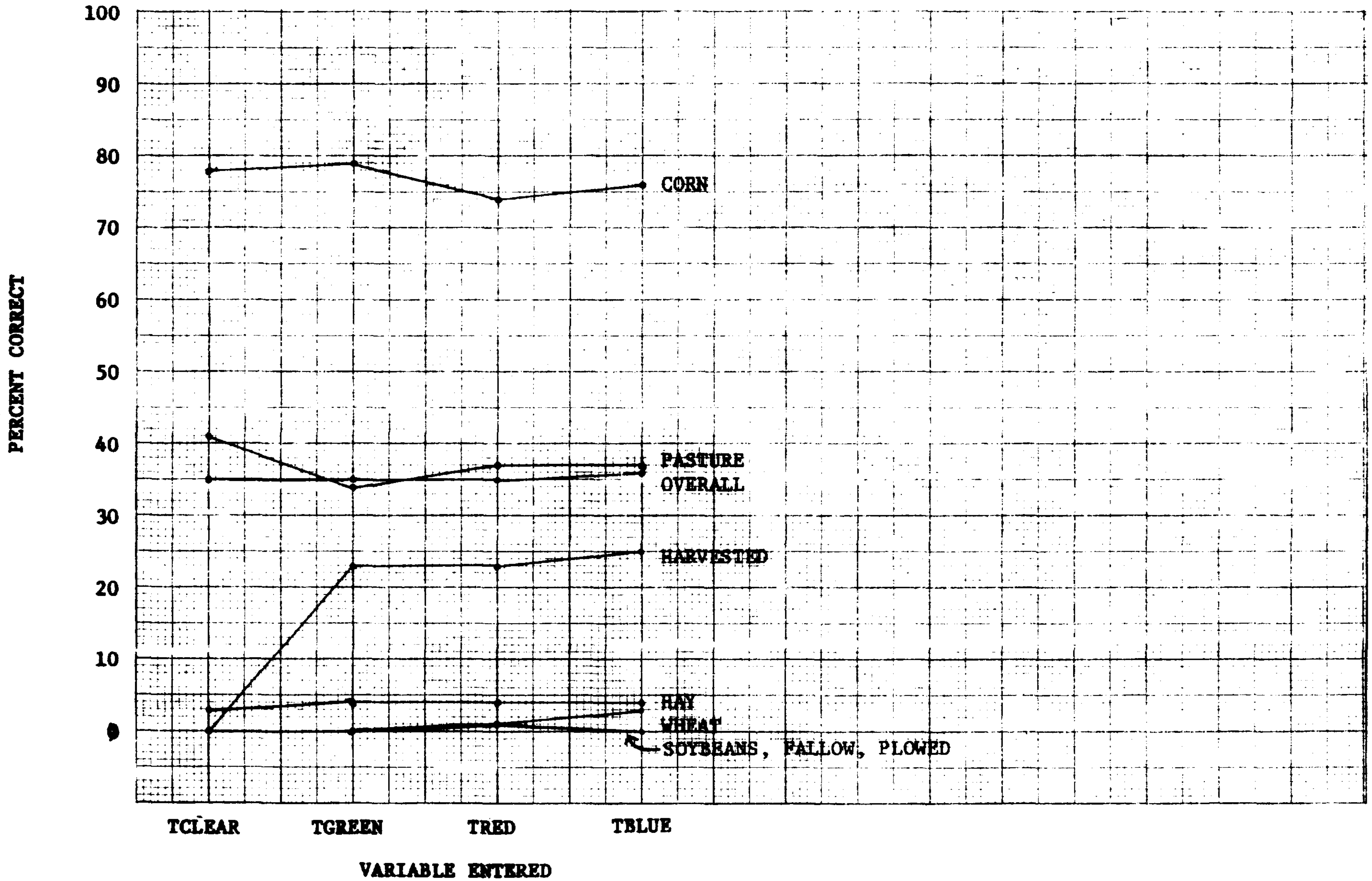


FIGURE 25

Stepwise discriminant analysis, classification into eight groups, density mode, South Dakota, 1972.

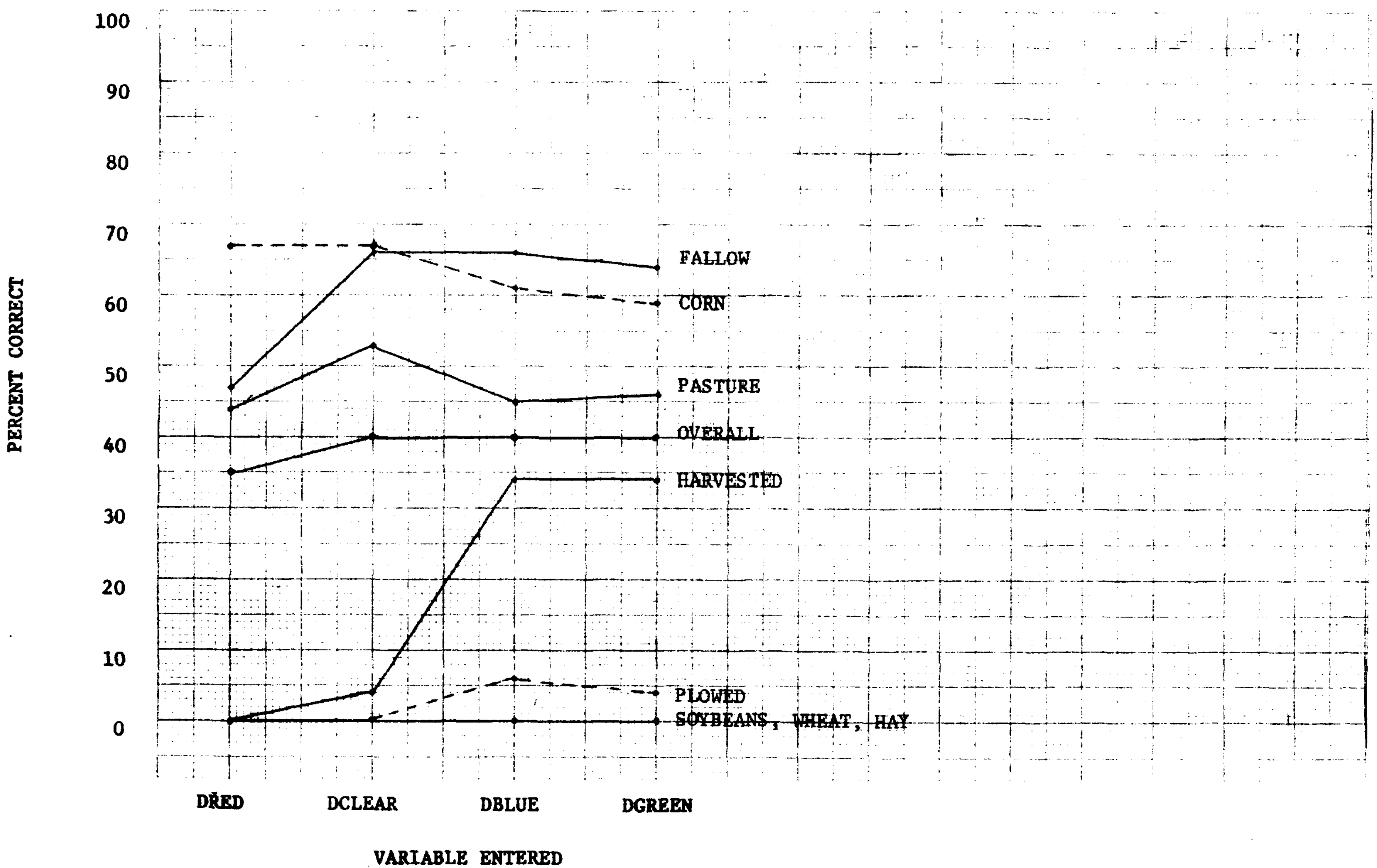


FIGURE 26

Stepwise discriminant analysis, classification into six groups, density and transmission scanning mode, South Dakota, 1972.

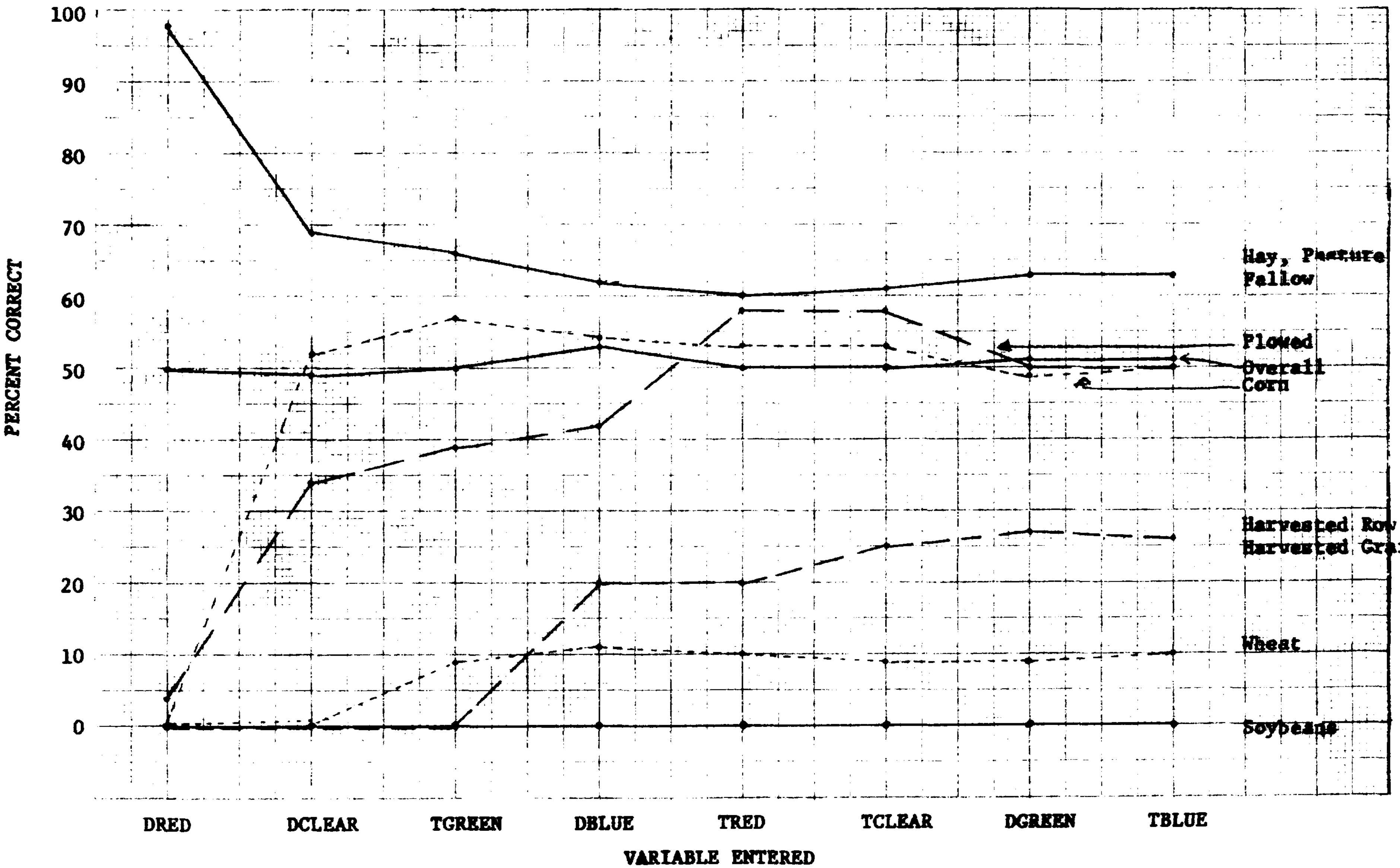


FIGURE 27

Stepwise discriminant analysis, classification into six groups, transmission scanning mode, South Dakota, 1972.

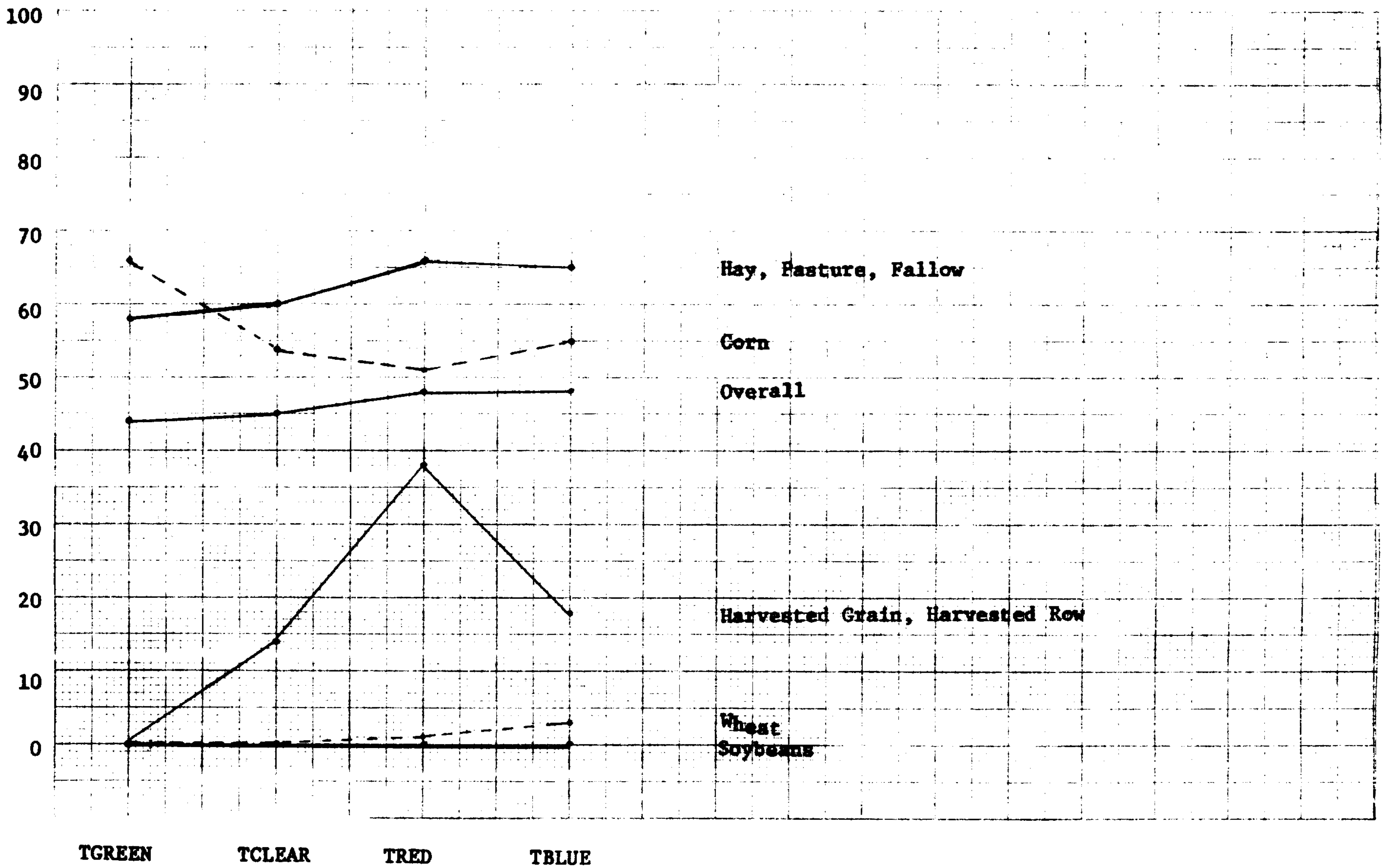


FIGURE 28

Stepwise discriminant analysis, classification into six groups, density scanning mode, South Dakota, 1972.

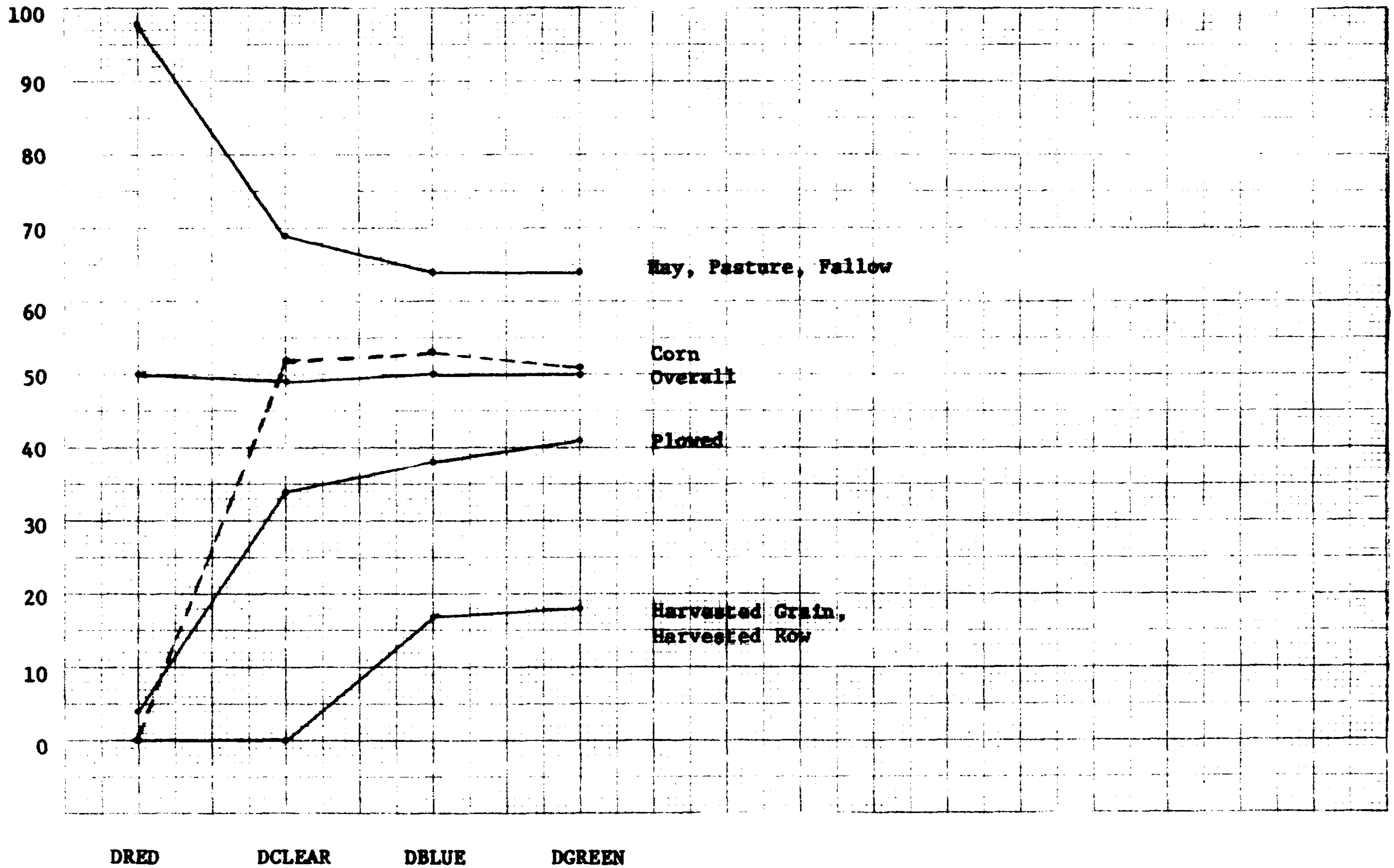


FIGURE 29

Stepwise discriminant analysis, classification into four groups, density and transmission scanning mode, South Dakota, 1972.

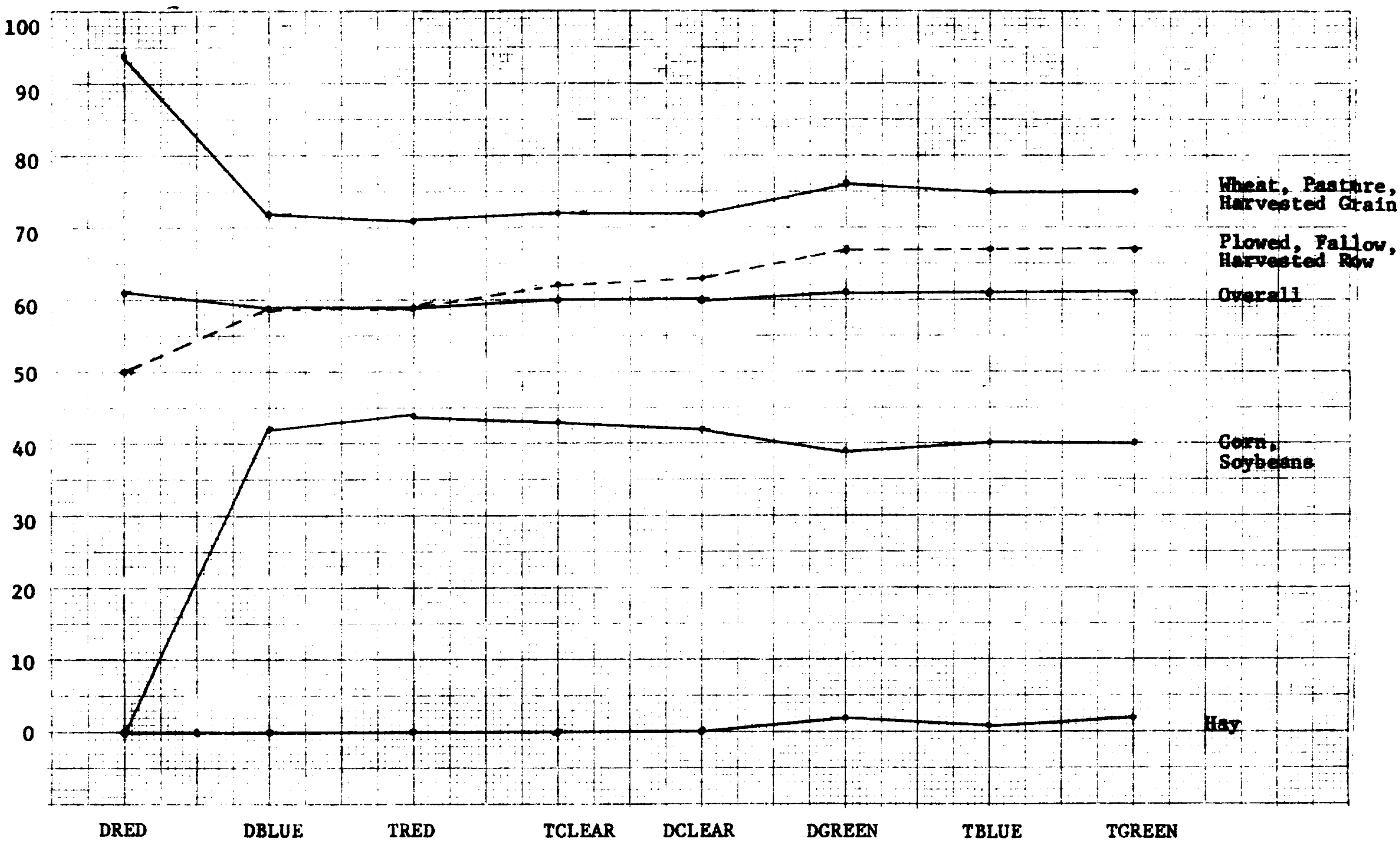


FIGURE 30

Stepwise discriminant analysis, classification into four groups, transmission scanning mode, South Dakota, 1972.

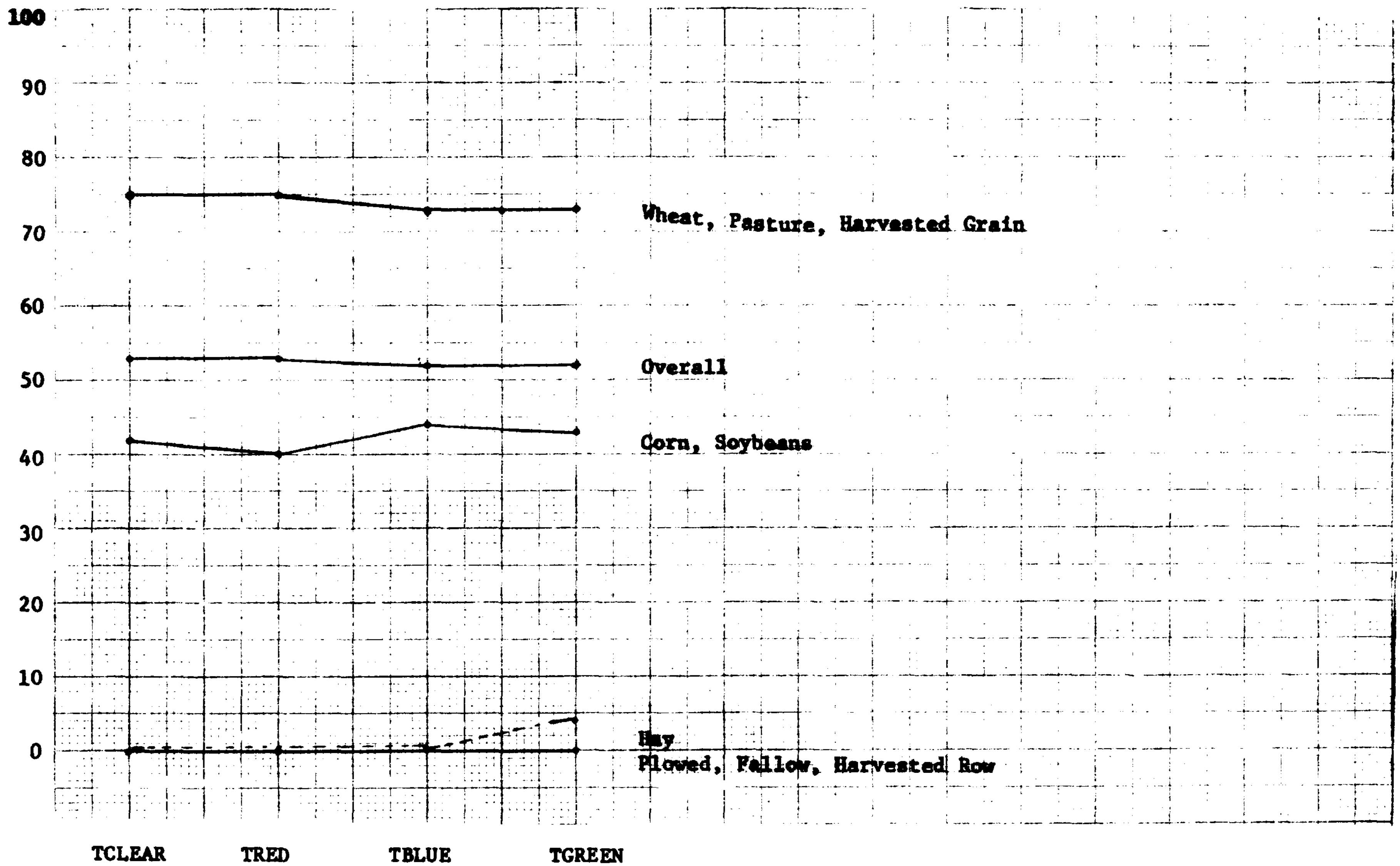


FIGURE 31

Stepwise discriminant analysis, classification into four groups, density scanning mode, South Dakota, 1972.

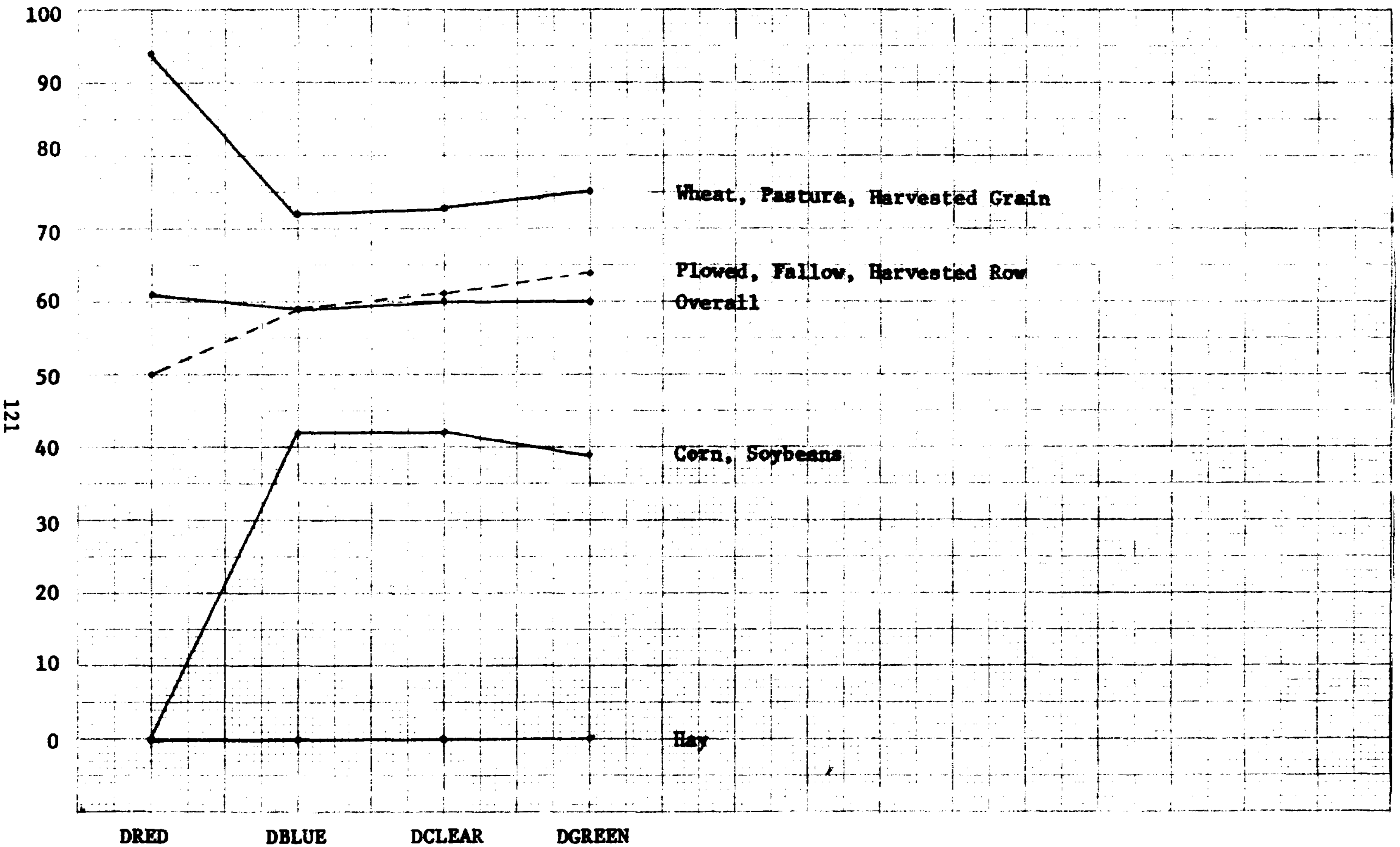
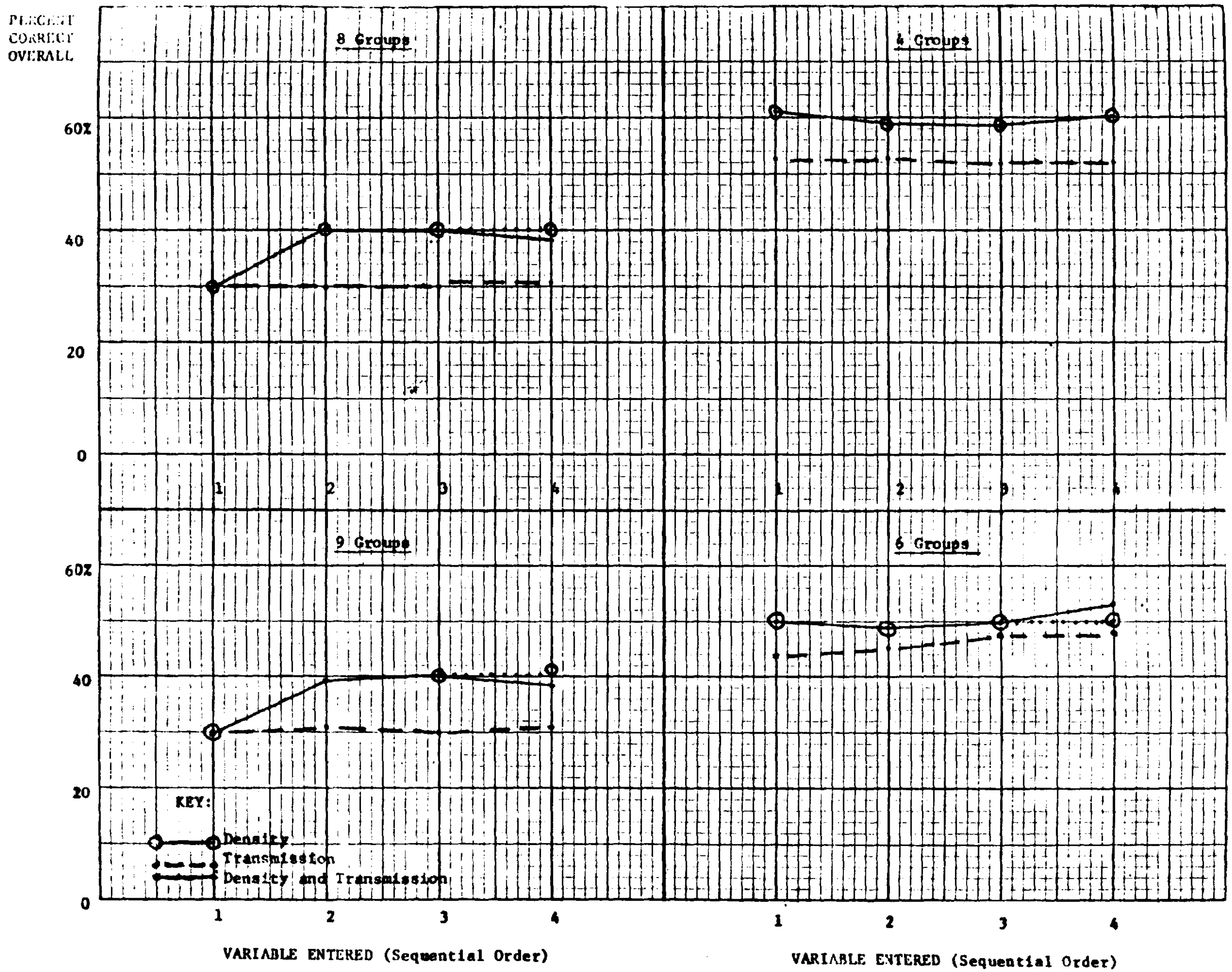


Figure 32--Overall Classification Accuracy By Number of Groups and Measurement Mode for Four Variables



4.2 Crop Acreage Estimation

The objective of this section is to present a procedure that will use classification results to produce an area acreage estimate. The regression technique presented may not be appropriate for users with different ground data. This technique requires that a random subsample of the total of all segments be selected for ground observations.

It is assumed that classification errors will be substantial, that is, perfect classification is not possible, and unbiased classification is not probable. Unbiased classification means more than that the classification errors simply balance. It means that the prior probabilities used are correct and the data are multivariate normal.

If unbiased classification were possible, we could use pixel counting techniques as estimators.

We know that the prior information was not exact and further that the data are not multivariate normal. Some delicate adjustments are necessary to produce an unbiased estimator and in order to make this adjustment, we will use the fact that a random subsample of segments has been selected for ground observations.

The first step is to estimate the linear relationship between total crop acres and total crop pixels inside the segment. This information must come from the ground truth segments and the relationship must be applied to the segments that were not selected for ground observations. An example of how the procedure would work follows. It turns out to be illuminating, but the estimates are poor because the relationships that are established in the ground observation segments do not represent the population that is being estimated.

This data came from the Southwest Crop Reporting District in Kansas.

The correlation coefficients squared (r^2) between the items of interest are presented in Table 66.

The relationship between acres on the ground and points classified corresponding to the same on the ground area can be established on a per segment basis.

Table 66--Source, r^2, $\bar{Y}$, $\bar{X}$, Var(Y), Cov(XY), and Var(X).

Source	r^2	$\bar{Y}$	$\bar{X}$	Var(Y)	Cov(XY)	Var(X)
Total acres (Y) versus total pixels (X)	.95	1843	1841	2,401,627	2,716,190	3,242,228
Alfalfa acres (Y) versus alfalfa pixels (X)	.01	39	223	7,187	-2,417	9,302
Pasture acres (Y) versus pasture pixels (X)	.89	728	890	1,467,689	1,325,965	1,348,245
Corn acres (Y) versus corn pixels (X)	.76	145	69	61,931	23,668	11,850
G. Sorghum acres (Y) versus G. Sorghum pixels (X)	.53	171	404	70,505	115,948	656,917

The model that will be used to represent the relationship is:

$$\hat{y}_i = \bar{y}_i + b_i (\bar{X}_{total\ i} - \bar{X}_{sample\ i})$$

where $\hat{y}_i$ is the adjusted acreage estimate for the i^{th} crop.

$\bar{y}_i$ is the average number of acres of the i^{th} crop in the selected segments.

b_i is the regression coefficient for the i^{th} crop estimated by:

$$\frac{\sum_{j=1}^{N} x_{ij}\, y_{ij}}{\sum_{j=1}^{n} x_{ij}^2} = \frac{\mathrm{cov}(xy)}{\mathrm{var}\ (x)}$$

where $\bar{X}_{total\ i}$ is average number of pixels of i^{th} crop in all segments in a county.

$\bar{x}_{sample\ i}$ is the average number of pixels in the selected sample for the i^{th} crop.

The estimator y_i is the adjusted average number of acres in the average segment. To get an estimate of the total y_i, would be multiplied by the total number of segments in the population (N).

The error of the regression estimator is written as:

$$Var(\hat{Y}) = \frac{S^2_{yi} \quad (1-r^2)}{n}$$

where $Var(\hat{Y}_i)$ is the variance of the final adjusted estimator of the average segment of the i^{th} crop.

S^2_{yi} is the adjusted between segment sums of squares for the i^{th} crop.

r^2 is the correlation coefficient squared between the number of acres in the segment and the computer classified number of pixels in the segments for the ith crop.

n is the number of degrees of freedom in the estimator.

Since the estimator for the total number of acres in the county is $N(\hat{y}_i)$, the variance of the total is N^2 times $Var(\hat{y})$.

The regression estimator above is the best in terms of lowest bias and smallest variance. Other estimators of the regression type such as, ratio estimators and difference estimators may be quite good in special cases. The regression estimator has definite advantages over the other two types of estimators just mentioned.

In Stevens County, Kansas, each pixel was classified. There were 410,505 pixels in the county and 468,000 acres. Each pixel represents 1.1401 acres. Actually, the county boundaries were approximated and this introduces a small amount of error. Out of the total of 410,505 pixels, the following pixels were classified as:

1.)	Alfalfa	5,362	5.) Other	37,567
2.)	Pasture	172,021		
3.)	Corn	30,448		
4.)	Grain Sorghum	165,107		

The first step is to put these pixels into a per segment basis. There were 280 segments in the county so the average segment contains 1,466 pixels for all land uses. The other averages were:

1.	Alfalfa	19.2
2.	Pasture	614.1
3.	Corn	108.7
4.	Grain Sorghum	590.0
5.	Other	134.0

Since the relationship between alfalfa acres and alfalfa pixels is quite poor, we shall demonstrate the procedure using pasture data.

The pasture acreage estimate for Stevens County using ERTS data is:

$$y_{pasture} = 430 + .9835(614 - 714) = 332$$

$$\hat{Y} = (280)(332) = 92,960 \text{ acres for Stevens County.}$$

$$Var(\hat{Y}_{acres}) = \frac{(1467,689)(4)(1-.89)(280)^2}{4(5)} = 3,164,337,484.$$

Standard Error = 56,252.4 C.V. = 60.5

The estimate and variance without using LANDSAT data are 120,400, and 23,013,363,520, respectively:

$$\text{where } V(\hat{y}) = \frac{1,467,689}{5}(280)^2 = 23,013,363,520$$

$$\text{and C.V.} = \frac{151,702}{120,400} = 126\%$$

Table 67 shows acreage estimates with variance and coefficients of variation for various crops with the aid of LANDSAT data.

Table 68 shows acreage estimates, variances, and C.V.'s for Stevens County, disregarding LANDSAT data.

The first point is that the variances of the estimates that use LANDSAT depend on the variance of the ground observations, the correlation of LANDSAT data with ground observations and the sample size. If the correlation is very high as with pasture, it is possible to produce an accurate estimate only if the ground observation is accurate. For example, no alfalfa was observed in the ground truth segments. Even though the computer was trained with alfalfa from outside the county and 5262 pixels were classified into the alfalfa category for Stevens County, the relationship was bad, and the ground observations were poor, and therefore, the estimate is bad and the C.V. very large.

These estimates and estimates of the variance were computed for two sample sizes. There were really three segments in Stevens County, and one of those was not used because of location problems. These numbers used the two segments left in Stevens County, the relationship for all 17 segments, and the total Stevens Company classification data. However, variances and C.V.'s were figured for samples of size 5 and 10.

If total aircraft classification were available for the same area, the model would be as follows:

$$\hat{y} = \bar{y} + b_1 (\bar{X}_1 - \bar{x}_1) + b_2 (\bar{X}_2 - x_2)$$

The variance would be similar to the previous formula:

$$Var(\hat{y}) = \frac{S_y^2 (1-R^2)}{n}$$

where R^2 is the multiple correlation coefficient squared and n is the number of degrees of freedom left in the estimator.

Table 67--Acreage estimates, variances, coefficients of variation for sample sizes of 5 and 10, using LANDSAT data.

Crop	Acreage Estimate	Sample of 5 segments		Sample of 10 segments	
		Variance	Coefficients of Variation	Variance	Coefficients of Variation
Alfalfa........	0	111,565,238	∞	55,782,619	∞
Pasture........	92,960	2,531,469,978	54.1%	1,265,734,994	38.3%
Corn...........	78,764	223,058,739	19.4%	116,529,370	13.7%
Grain Sorghum..	150,689	519,593,648	15.1%	259,796,824	10.7%

Table 68--Acreage estimates, variances, coefficients of variation for sample segments of size 5 and 10, without the aid of LANDSAT data.

Crop	Acreage Estimate	Sample of 5 segments		Sample of 10 segments	
		Variance	Coefficients of Variation	Variance	Coefficients of Variation
Alfalfa........	0	112,692,160	∞	56,346,080	∞
Pasture........	120,400	23,013,363,520	126.0%	11,506,681,760	89.1%
Corn...........	65,520	971,078,080	47.6%	485,539,040	33.6%
Grain Sorghum..	321,840	1,105,518,400	14.3%	552,759,200	10.1%

V. Cost Analysis

This section is presented to provide cost information relative to various sources of data collection. It is documented so that as technology is improved, the cost of developing an integrated data collection system can be realistically evaluated.

However, the cost data cited reflects only the conditions under which this project was completed. It is to be expected that new technology will change some of these costs in the future.

Cost of Ground Data

The cost of ground data can be broken into collection costs and summarization costs. The data collection costs include:

a. pre-survey planning and materials preparations,

b. enumerator training schools, and

c. enumerator fieldwork.

The summarization costs include:

a. collection, edit, and keypunch time for Washington, D.C. and State Statistical Office personnel, and

b. programming and summarization costs. These costs pertain to all four test sites and are as follows:

5.1 Data Collection

1) Survey Planning and Materials Preparation

Research and Development		
Salaries	$1,342.00	
Travel costs (map preparation salaries)	263.67	
Programming Costs		
Salaries	849.59	
Computer costs	1,259.11	
		$ 3,714.37

2) Enumerator Training Schools

Instructors	901.84	
Salaries	477.44	
Travel		
Enumerators		
Salaries	530.00	
Travel	210.00	
		$ 2,119.28

3) Enumerators Fieldwork

Salaries	$4,931.65	
Travel	3,044.00	
		7,975.65
Total Data Collection Costs		$13,809.30

5.2 Data Summarization

1) Collection Edit and Keypunch Costs		
SSO Salaries	2,524.05	
Research and Development Salaries	6,816.68	
		$ 9,340.73
2) Programming and Summarization Costs		
Salaries	2,632.80	
Computer Costs	1,281.49	
		$ 3,914.29
Total Data Summarization Costs		$13,255.02
Total LANDSAT Ground Truth		$27,064.32

It should be noted that the above cost data are for the update work conducted in August, September, and October. The costs of the regular June Enumeration Survey (JES) are not comparable since in addition to observing and recording ground cover, the JES records crop intentions and livestock numbers. Estimates of these costs can be derived, however, by using enumerator time and mileage costs. Mileage rates and hourly wages applied against the miles driven and hours worked together give a total cost estimate by segment. This comparison follows:

5.3 JES Fieldwork Costs

A. Time

District	State	Time	#Segs.	$/hour	
9	Missouri	6.42 hr/seg.	52	$3.30	$1,101.67
6	S. Dakota	4.80 hr/seg.	50	3.30	792.00
7	Kansas	8.93.hr/seg.	48	3.30	1,414.51
2	Idaho	5.75 hr/seg.	44	3.30	834.90
	Total		194		$4,143.08

Time cost per segment = $21.36

B. Mileage

District	State	Miles	#Segs.	$/mile	
9	Missouri	99.98 m/seg.	52	.11	$ 571.89
6	S. Dakota	80.86 m/seg.	50	.11	444.73
7	Kansas	136.81 m/seg.	48	.11	722.36
2	Idaho	82.85 m/seg.	44	.11	400.99
	Total Mileage Cost				$2,139.97
					$6,283.05
	Mileage cost per segment		$11.03		

C. Total Time and Mileage

Total time and mileage cost/segment $32.39

5.4 Update Fieldwork costs (3 visits)

A.	Salaries	4,931.65	
B.	Travel	$3,044.00	
C.	Total Time and Mileage ($7,975.65/3=$2,658.55)		$7,975.65

Total update time and mileage costs per segment $41.11

Total update time and mileage costs per segment per visit $13.70

The difference between $6,283.00 and $2,658.55 represents the additional costs of $3,624.50 needed to locate the June Segment Operators, secure livestock data and farm labor data. This LANDSAT update fieldwork only included locating the segments and recording the crops present and their conditions. The operators were not contacted unless the enumerator could not view the fields from the road.

Tables 69 through 76 show detailed time and mileage data for the study sites.

Table 69--Missouri 1972 JES Time and Mileage Data

Dist	Number Segs	Visits/ seg	Hours/ seg	Miles/ seg
10	60*	1.43	4.90	70.63
20	51*	1.75	4.70	81.08
30	49	2.04	4.91	82.78
40	50	1.94	5.30	86.78
50	60*	1.93	5.34	80.90
60	39*	1.82	6.95	85.08
70	46	1.52	5.19	61.04
80	42	1.71	5.55	94.52
90	52	2.23	6.42	99.98

*Not all segments in this district had cost data reported.

Table 70--South Dakota 1972 JES Time and Mileage Data

Dist	Number Segs	Visits/ seg	Hours/ seg	Miles/ seg
10	31	1.87	7.93	107.16
20	46	1.67	5.33	93.20
30	42	1.83	5.74	88.19
40	31	1.90	8.19	122.23
50	42	1.90	6.96	99.90
60	50	k.92	4.80	80.86
70	22	1.86	8.32	131.23
80	35	1.63	7.23	101.14
90	51	1.82	4.65	75.98

Table 71--Kansas 1972 JES Time and Mileage Data

Dist	Number Segs	Visits/ seg	Hours/ seg	Miles/ seg
10	42	2.26	6.99	130.52
20	54	2.04	5.67	94.43
30	50*	2.58	5.74	88.94
40	40	2.25	8.16	141.90
50	56	2.46	6.81	127.21
60	53*	1.87	4.92	69.00
70	48	2.79	8.93	136.81
80	60*	1.93	6.68	105.83
90	53*	1.81	4.69	82.34

Table 72--Idaho 1972 JES Time and Mileage Data

Dist	Number Segs	Visits/ seg	Hours/ seg	Miles/ seg
2	54	1.54	5.75	82.85

Table 73--Time and mileage data for Idaho by enumerator.

Enumerator identification	JES segments completed	Average Number visits per segment	Average hours per segment	Average miles per segment
6	3	3.00	3.78	104.67
12	8	3.00	8.83	125.00
18	10	2.10	3.44	56.70
19	14	3.29	7.41	80.07
30	13	1.54	4.19	80.23
33	6	2.83	5.93	71.00
Totals:	54	2.54	5.75	82.80

Table 74--Time and Mileage data for Missouri by enumerator.

Enumerator identification	JES segments completed	Average number: visits per segment	Average hours per segment	Average miles per segment
1	15	1.33	4.89	58.53
2	15	1.47	4.19	80.93
3	6	2.50	3.38	87.5
4	17	1.76	3.66	59.18
5*	6	1.83	7.77	118.33
6	9	1.89	7.09	93.78
7*	12.5	1.68	4.69	66.64
8	13	1.77	7.06	78.92
9	10	1.40	4.99	78.10
10	11	1.73	4.44	109.91
11	16	2.12	4.64	78.56
12	15.7	1.27	5.06	63.76
13	17	1.41	4.50	57.53
14	13	2.15	6.27	109.23
15	13	1.08	4.94	56.23
16	11	1.45	5.11	81.18
17	15.5	2.58	5.97	115.10
18	14	2.43	7.01	93.21
19*	14	1.21	3.93	51.00
20	11	1.55	6.66	90.09
21*	13	2.00	5.58	78.46
22	7	3.57	9.18	99.71
23*	13	1.92	5.47	87.62
24	11	3.27	10.27	163.91
25	14	2.07	4.79	90.36
26	12	1.67	4.60	64.17
27	14	2.43	6.70	102.57
28	14.3	2.10	4.70	103.08
29	9	1.67	8.78	67.67
30	16	1.94	5.07	78.00
31	9	2.56	6.13	114.33
32	18	1.28	4.56	46.50
33	9	1.22	4.28	62.11
34	10	1.50	5.38	77.80
35	17	1.41	3.76	65.29
36	8	1.75	6.12	130.25
TOTALS:	449	1.82	5.42	82.22

* Supervisors

Table 75--Time and mileage data for South Dakota by enumerator.

Enumerator identification	JES segments completed	Average number visits per segment	Average Hours per segment	Average miles per segment
1*	2.0	1.03	1.04	37.59
2*	3	1.33	5.50	111.33
3*	8	1.75	4.35	73.12
4	13.8	1.88	6.96	151.52
5	8.6	1.74	6.25	87.79
6	7	1.43	8.36	80.14
7	14	2.07	5.91	117.03
8	6	1.50	4.79	46.50
9	19	2.16	4.13	80.21
10	15	1.93	5.51	95.93
11	9	1.33	7.19	69.78
12	7	1.86	5.17	63.43
13*	13	1.85	3.87	79.54
14	11	1.45	5.18	71.91
15*	7.4	1.08	2.90	62.03
16	10.5	1.43	5.09	76.86
17	13	2.38	9.29	144.69
18	12	2.50	1	159.00
19	25.4	2.24	4.72	82.72
20	15	1.53	7.01	90.07
21	11	1.73	4.67	71.55
22	14	1.64	4.90	133.29
23	15	1.60	5.61	70.33
24*	8.7	1.84	8.33	126.78
25	13.1	2.14	7.36	129.01
26	8.3	2.29	7.14	135.54
27	13	1.54	7.81	83.85
28	15	2.00	7.48	95.33
29	15	1.80	5.38	77.67
30	5	1.40	6.02	33.00
31	9	1.56	5.81	113.67
32	2.3	2.17	5.20	149.13
TOTALS:	350	1.83	6.27	95.91

* Supervisors

Table 76--Time and mileage data for Kansas by enumerator.

Enumerator identification	JES segments Completed	Average number visits per segment	Average hours per segment	Average miles per segment
1	8.9	2.70	7.15	105.73
2	18	2.00	5.21	105.11
3	15	2.40	5.58	89.67
4	12	2.33	4.63	78.17
5	20	1.90	5.99	111.00
7	14	2.71	5.29	98.07
8	9.8	1.63	9.29	134.69
9	11	2.09	6.33	69.18
10	9	1.56	7.96	135.56
11	4	1.75	7.96	80.00
12*	3.4	2.94	11.06	272.94
13	19	1.95	4.82	74.37
15	16.9	2.84	9.09	180.65
16	14	3.5	8.78	142.36
17*	4	1.25	6.69	116.75
18	11	1.73	4.83	58.64
19	12	2.42	8.38	162.08
20	12	3.25	8.68	132.92
21	14	2.64	5.08	108.07
22	3	1.67	5.67	98.67
23	12	2.25	7.85	122.00
24	14	2.71	4.67	90.00
25	16	2.44	4.60	99.69
26	13	1.85	6.50	108.77
27*	5	3.00	15.92	260.20
28	16	2.12	5.43	89.12
29	12	1.75	5.62	69.25
30	12	2.08	7.25	117.25
31	15	1.60	4.76	65.73
32	14	1.21	4.23	69.14
33	9	3.11	7.62	114.67
34	7	3.29	10.86	173.86
35	11	2.82	9.20	151.09
36	14	1.64	7.02	93.57
37	11.5	1.48	4.85	81.04
38	17.5	2.46	5.15	107.43
39	15	1.53	5.29	91.40
40	11	1.73	5.93	54.91
TOTALS:	456	2.21	6.44	107.10

* Supervisors

5.2 Aircraft Cost Analysis

NASA provided the following estimates for aircraft costs:

U-2 operational costs are $2,150 per hour with coverage of about 400 nautical miles per hour. Coverage is 14.8 nautical miles on a side per scene.

Scenes per hour $= \frac{400}{14.8} = 27.03$

Cost per scene $= \frac{\$2,150}{27} = \$79.63 = \$80$

For the study areas, the acquisition costs average about $60 per segment.

The activities and the approximate time and costs required to prepare the aircraft data for crop classification are:

		Average time per segment
Sketch segment and record field boundaries		37 min.
Microdensitometer scanning		33 min.
Recording and keypunching input data for field extraction		
	Total man hours	1.83 hours
	Cost/man hours	$4.50
	Average cost/segment	$8.23

ADP Costs

PDSCMS data conversion

Field extraction

Total ADP costs/segment

The average cost per segment for data preparation $29.23

The costs of crop classification varies with the size of segment, but in order to have a comparable cost with ground observations, it is presented on a per segment basis. The average cost per segment for crop classification was about $81 segment. The average cost per data point is about 3 cents per point.

The total aircraft survey costs were about $170 per segment. This compares with $47 per segment per visit for the ground observations.

This analysis deals primarily with the time and costs required for scanning the aerial photography and converting the data into a form suitable for crop classification by discriminant analysis in the Statistical Analysis System (SAS).

Time and cost data were collected as follows:

1) Pre-scan setup: the time (man minutes) required to locate the segment on the microdensitometer, sketch the segment, record field boundary coordinates and define the microdensitometer scanning parameters.

2) Scanning: the time (man minutes) required for system analog calibration and microdensitometer scanning with each of the four filters in density and transmission units.

3) Data preparation for field extraction: the time required to record and keypunch input data for the field extraction program.

4) PDSCMS data conversion: ADP costs for converting the microdensitometer output data to SAS compatible data.

5) Field extraction: ADP costs for assigning crop classes, tract and field identifiers to individual pixels on the basis of ground observations utilizing pixel coordinate information.

Several factors contributed to the substantial differences between states for the average cost per segment. The differences for pre-scan setup times can be attributed to two primary factors:

1) different microdensitometer operators. A new operator was in training while scanning South Dakota, and had gained in experience when Missouri was scanned.

2) the relative difficulty recording field boundary coordinates for each state. South Dakota and Missouri were most difficult because of many small field sizes, followed by Idaho, with Kansas least difficult.

New field boundary coordinate recording procedures were implemented near the end of the Idaho scanning and were subsequently employed while scanning the Missouri photography. Due to operator differences, it is difficult to objectively assess the effectiveness of the new procedures. Subjectively, it is believed the new procedures will reduce pre-scan setup time by 10-20% and data preparation for field extraction by 25-40%.

Scanning time remains fairly constant between states (the large difference in South Dakota is attributable to a new operator in training on the microdensitometer). Small differences are a function of the number of segments and average size of each segment.

Between state differences in automated data processing costs are a function of the number of segments, average size of each segment, and the number of tracts and fields within each segment.

5.5 Computer Costs

Processing LANDSAT data and digitized aircraft data requires enormous amounts of computer time. The following table shows the cost of computer time for processing at the Washington Computer Center for various broad classes of processing.

DEVELOPMENT	$ 6,631
GROUND DATA	$ 2,915
MAPS FOR SEGMENT LOCATION	$ 9,227
AIRCRAFT DATA ANALYSIS	$30,142
TOTAL	$48,915

Development costs include converting software to run at WCC, maintenance, developing original programs and overhead.

Ground data costs are for building and maintaining, and summarizing of ground data.

The MAPS were grey level maps of LANDSAT CCT's for segment location.

The aircraft analysis cost includes charges for conversion of microdensitometer data, building the data, files, and runs used to determine the best analysis procedure, the discriminant analysis, and the combination of the satellite results, aircraft results, and ground data.

APPENDIX A

ERTS ENUMERATORS INSTRUCTIONS

ERTS[1/] ENUMERATORS INSTRUCTIONS

1.1 What you will do:

You are one of about 16 enumerators in four states (Kansas, Missouri, South Dakota, and Idaho) employed to obtain "ground truth" about crop species, acres and crop condition. Briefly, your job is to update information from the June Enumerative Survey (JES) by verifying crop species and acres and observing crop condition during July, August, September, and October. Your field verifications and observations are to be recorded on the Earth Resource Technology Satellite (ERTS) Ground Truth Printouts.

1.2 Equipment and Supplies

USDA identification card
aerial photos
aerial photo mailing boxes
county maps
CEF-201's
ERTS Ground Truth Printout
large envelopes for mailing completed forms
motor vehicle accident report kit
ball point pen
lead pencil, plus red, orange, and yellow colored pencils
clipboard
highway maps
Julian dates.

1.3 Mailing and survey dates

After each survey is completed you will mail the updated printout and your CEF-201's to the SSO in the envelopes provided. All other materials used during the survey will be retained until the final survey period Your final mailing will include the updated printout, CEF-201, aerial photos and county maps, plus any other surplus materials.

The survey periods and mailing dates are as listed:

Survey period	Enumerators mailing date
August 7-11	on or before August 11
September 11-15	on or before September 15
October 10-13	on or before October 13

[1/] At the time this manual was written, LANDSAT was called ERTS, acronym for Earth Resources Technology Satellite. ERTS was never changed to LANDSAT because this manual was never used after 1972.

1.4 Terms and definitions

The regular enumerative survey definitions hold for this survey

A. Segment - land area outlined in red on aerial photos and county maps. Each segment is identified by a permanently assigned 4-digit number. See the "Survey Enumerators Handbook" for discussion on use of aerial photos and locating segments.

B. Tract - an area of land inside the segment which is under one management, Each tract is identified by a letter code A, B, C, etc. on the aerial maps and by the corresponding numeric code on the form printout (i.e. A = 01, B = 02, C - 03, etc.). Tract boundaries and letter codes are drawn in blue pencil inside the segment on the aerial photo.

C. Field - a continuous area of land inside a tract which is devoted to one crop or land use. Each tract on the aerial photo is divided into fields during the enumeration of segment acreage in late May or early June. Fields are numbered and their boundaries outlined in red pencil.

D. Farm Operator - the person who is responsible for the day-to-day decisions for a tract.

Part II - The Survey

2.1 Purpose of the survey

The purpose of the Earth Resource Technology Satellite (ERTS) program is to:

a) investigate and evaluate the use of space imagery to identify crop species.

b) investigate ways of using space imagery to improve agricultural statistics.

Through ground truth obtained during July, August, September, and October we will be able to check and verify the accuracy of satellite imagery (500 miles) and high altitude photography (60,000 feet) as a method of measuring crop acres which in turn will be used to generate an expanded estimate of crop acreage.

Ground, high altitude, and satellite estimates of acreage will be obtained and compared against collection costs to indicate a cost-information ratio. Trained enumerators as yourselves will collect the ground information from the field. Trained photo interpreters will record species and acreages for high altitude photography in the Washington, D.C. office. Computers will be used to analyze satellite imagery in the Washington, D.C. office. Cost information for each method of collection will be retained and compared versus the accuracy of reliability of each method of data collection.

2.2 The sample

The segments selected for this survey were selected to provide different crops in different locations. A different mix of crops will be found in Idaho versus Kansas versus Missouri versus South Dakota. How do sugar beets compare with potatoes in Idaho or grain sorghum in Kansas? Will spring wheat be distinguishable versus winter wheat? Does corn in Missouri look the same as corn in South Dakota? The information collected will provide answers to these types of questions. Additionally, with the distant geographic areas, inclement weather should not cover all the test sites and limit the quality of all imagery on a particular survey.

We use the JES segments since they represent 100% coverage of the areas in question. If bad weather renders some of the aerial photography or satellite images useless, we will attempt to develop reliable estimates for the other areas based on the ERTS ground truth. This may become a multiple frame model for acreage estimation.

2.3 Survey forms

For the second visit you will be provided a printout listing in segment and tract order fields, acreages and crops from the JES (first visit). On this visit you will note the condition of the crop on the printout. On succeeding visits the printouts will show fields, acreages, crop and conditions for each earlier visit.

The name and addresses of the operators from the JES will be provided on separate sheets of paper grouped by segments. These will be for use when it is necessary to locate the tract operator for permission to view fields not observable from public roads.

Part III - Field Observations

3.1 Locating the segment

Locate the segments you will be visiting on the county maps, then plot them on a highway map. Plan your journey to observe these segments with minimum mileage and travel time for each day's journey.

3.2 Recording observations for segment

When the printout and maps are updated on the monthly visit, use the color codes listed below for field boundaries and numbers.

Second visit	- red dashes
Third visit	orange dashes
Fourth visit	yellow dashes

Note: We will only mark corrections on the map. Incorrectly drawn fields will not be erased. There may be no new dates for the survey duration if fields are drawn correctly from the JES and there are no acreage changes.

For this survey we are not interested in transfer of ownership etc., except to know whom to contact for enumeration purposes. Our concern is enumerating the land use and crop development condition of the segment.

Step 1. Verify that you are looking at the correct photograph(s) and the correct printout by locating landmarks on the map and locating the segment and tract number on the printout. Record the Julian date on all N of N pages of the tract printout.

Step 2. Verify that the field is drawn correctly on the maps by a) looking at the field defined by the map and b) deciding whether the map accurately shows the field with respect to common landmarks. If the field cannot be observed from public roads contact the tract operators and request permission to observe the fields in question, then write on the bottom of the printout whether permission was secured or refused. (By default, the printout will write unasked unless permission is noted as secured or refused).

If permission is refused, record observations for fields observable from public roads. Enter refused (code - RFSD) for the fields not observable from public roads in acres, crop and condition columns. If the map is correct go on to Step 3. If the map is incorrect, redraw the fields using the correct color scheme before beginning Step 3. Do not erase any previous survey boundary lines.

Step 3. For the given field number, check the acres listed on the printout versus the map and your own best estimate of the actual field acreage in whole numbers. If the acres are the same as the previous visit check (✓) the space for acres. Where a correction is necessary (i.e. --- an error has beem made or an obvious change in acreage has occurred since the previous visit) check with the operator for the corrected acres or record your own acreage observations where checking is not possible and write a note explaining the change. See Figure 1 for examples of corrections and changes.

Step 4. If the crop is the same as the previous visit check (✓) the crop code. If a change has occurred record the corrected crop code for that field.

Step 5. Using the guide from Section 3.3, write the condition of the crop on the printout in the space provided. Write a note to explain any situation or our condition codes do not accuractly describe.

Step 6. Repeat steps 1-5 until all fields are completed, then check that all N of N pages for a tract listing are present and complete.

Step 7. Repeat steps 1-6 until all tracts and segments are completed.

3.3 Crop codes and conditions

Since we will be looking at aerial photography and satellite imagery, we need to know the crop species and the condition code that best described and coded appropriately on the printout. In order to code the condition properly you must observe the total area in the field which would be covered by the crop and then give a subjective evaluation of the crop development as well as a recommendation for action to be taken.

Crop	Situation	What to do
Alfalfa	part down, part cut	condition = cut or down, whichever portion is the larger.
All crops	very poor stand, dry etc.	condition - what a normal healthy crop would look like with a note.
Applies to most	two similar species but different planting dates	draw in new fields boundaries and properly number and classify the new field.
Grass waterways	located on natural boundaries or can accurately be drawn on the map	draw in new field and classify as OTHR and specify on printout with a note.
Drowned out areas	located near natural boundaries or can accurately be drawn on the map	draw in new field and classify as OTHR and specify on printout and a note.

ERTS Editing Instructions

1. Edit the Julian date to correspond to the actual field visit.

2. Check the acres for a given field number versus the previous recorded acres.

   A. Do not edit where column is checked (✓).

   B. Dn case the acres differ from the previous visit:

      1) If a new field is created or acres for given fields are adjusted, the acres should be adjusted to total the previous acreage total or a note shoudl explain the total acres change. Check that the new field is correctly numbered.

      2) Where an error occurred on the previous enumeration, an enumerators note should explain the correction. With an explanation, the correction will be punched. With no explanation talking to the enumerator or statistics judgement will appropriate to edit in corrected acres.

   C. Check (✓) the acres column where the tract was a refusal and the field not observable.

3. Check the crop codes for correctness.

   A. Do not edit where the column is checked (✓).

   B. Edit out where they are unrecognizable.

   C. Correct the code when it is a change from the previous month and incorrectly written.

   D. Enter RFSD where the tract was a refusal and the field not observable.

4. Check the condition code against nearby fields.

   A. Edit to compare with fields in the tract or segment where the condition is not entered.

   B. Check with the enumerator where the condition is not entered and there are no comparable fields in the tract or segment.

   C. Edit cut where condition is the sure as the previous month.
   NOTE: On the first visit there must be an entry for every field.

D. Enter RFSD where the tract was a refusal and the field is not observable.

5. Code the permission

   A. Unasked = 0

   B. Secured = 1

   C. Refusal = 2

ERTS Mailing Instructions

Send the edited printout and the punched cards Air Mail Special Delivery in "Special C" envelopes and "Special C" card mailer.

Each envelope and card mailer should be marked in the lower left hand corner as follows:

REPORT: ERTS Ground Truth

STATE: Your State (99) (1 of 2)

Secure each envelope and card mailer with a strand of filament tape each way around the envelope and card mailer.

Send the aerial photos as follows:

Research and Development Branch
SRS of USDA
Room 4837 South Bldg.
Washington, D.C. 20250

APPENDIX B

State Office Instruments

TERMS AND DEFINITIONS USED FOR JUNE ENUMERATIVE SURVEY (JES)

AND ERTS FIELDWORK

A. SAMPLE:

Information for the ERTS Survey is obtained from a small sampling of the total land area in four States. Small areas of land have been selected at random for this survey. Each area to be enumerated has been outlined in red on the county highway maps and aerial photographs which you are supplied. Every acre has one and only one chance to be selected in the sample.

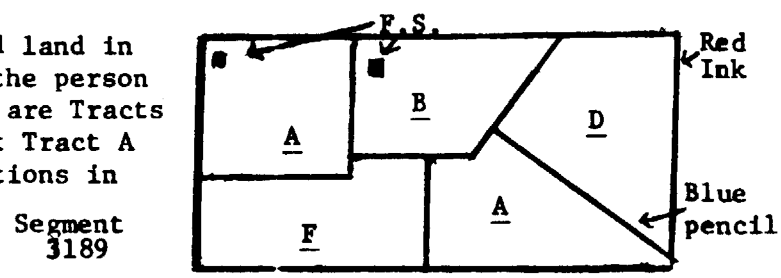
B. SEGMENT:

Segments are land areas outlined in red on aerial photos. Segments generally range in size from one-half square mile to three square miles. A few are larger or smaller depending on locations. Segments are identified by a permanently assigned number.

C. TRACT:

A TRACT is an area of land inside the segment which is under one operation. This tract may consist of agricultural land, non-agricultural land, residential areas, or some other land use. Examples of tracts are as follows:

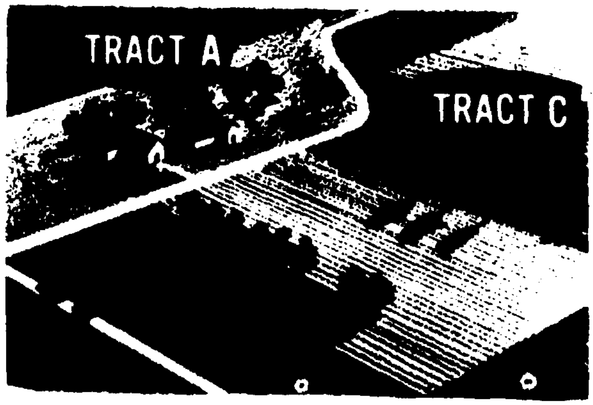
(1) An occupied house and land in segment operated by the person in charge. Examples are Tracts A and B. Notice that Tract A has land at two locations in segment.



(2) A farm operator living in the segment on a dwelling where he is not the person in charge, and who has no other land in the segment. See F.S. above. If all the land he operates is outside the segment, he is still a resident farm operator and should be assigned 1.0 acres of land for the tract. He lives in this dwelling.

(3) Any area of land in the segment under one operator who does not reside inside the segment. Tracts D and F are examples.

The boundaries of each tract will be outlined in blue pencil and each tract will be identified by a code letter. If a tract consists of more than one separate parcel of land, all parcels will be identified with the same letter; i.e., all of the land inside the segment that is operated by one person will be reported under one tract code.



D. FIELD:

A field is a continuous area of land inside a tract which is devoted to one crop or land use. Each tract will be divided into fields by you during this Survey. Each field will be outlined in red pencil and assigned a number.

E. FARM:

A farm consists of the area or areas of land both inside and outside of the segment boundaries under one management on which there were crops, livestock, poultry, or some sales of agricultural products at some time in 1971 or 1972.

F. OPERATOR:

The OPERATOR is the person who is responsible for the day-to-day decisions for the tract and total land operated.

If the tract contains a farming operation, the operator could be the owner, hired manager, cash tenant, or sharetenant. If a person operates farmland as a hired manager or partner and also operates land for himself as a separate farm, the managed or partnership land should be separated and assigned another tract code.

If the land is rented to others or worked by others on shares, the tenant or renter is considered the operator of the rented land.

UNITED STATES DEPARTMENT OF AGRICULTURE
Statistical Reporting Service
O. M. B. NUMBER 40-R2766
Approval Expires 4/30/74 480

State	District	Segment	Tract

PART A - 3 (Mo.)

JUNE 1972 ACREAGE, LIVESTOCK & LABOR

ENUMERATIVE SURVEY

Use this questionnaire only if operator lives INSIDE the SEGMENT.

Facts about your farm or ranch will be kept CONFIDENTIAL and used only in combination with similar reports from other producers.

SEGMENT NUMBER ______________________ TRACT CODE LETTER: ______________________

NAME OF RESIDENT OPERATOR: ____

ADDRESS: ____
(Route or Street)

____
(City) (State) (Zip)

TELEPHONE NUMBER: ______________________ COUNTY: ______________________

NAME OF FARM: ____

DATE: ______________________

Name and Address of PARTNERS: ____
(Record partnership operations as a separate tract.)

1. How many acres are inside these boundaries drawn on the photo (or map)? .... [ ]

2. Will any acres INSIDE these boundaries be IRRIGATED during 1972?

YES ( ) Ask irrigation questions
NO ( ) Skip irrigation questions

Now I would like to ask you about each field in the tract, the total acres in each field, and the crop or land use in 1972. For crops, I will ask for acres planted and to be planted for harvest this year. If two crops will be harvested from the same field, we should list both of them.

SECTION A - ACREAGES OF FIELDS AND CROPS IN TRACT

A-3

		FIELD NUMBER	1	2	3	4
1.	TOTAL ACRES IN FIELD					
2	CROP OR LAND USE (Specify)					
3.	ARE CROPS HARVESTED FROM THIS FIELD?		No ( ) Yes ( )	No ( ) Yes ( )	No ( ) Yes ( )	No ( ) Yes ( )
3a.	Has this field been planted?		No ( ) Yes ( )	No ( ) Yes ( )	No ( ) Yes ( )	No ( ) Yes ( )
4.	ACRES IRRIGATED AND TO BE IRRIGATED?					
5.	FARMSTEAD, ROADS, DITCHES, WOODS, WASTE, ETC.					
6.	PASTURE	Permanent - Not in Crop Rotation	842	842	842	842
7.		Cropland - Used Only for Pasture	845	845	845	845
10.	WINTER WHEAT	Planted	540	540	540	540
11.		For Grain	541	541	541	541
12.	RYE	Planted	547	547	547	547
13.		For Grain	548	548	548	548
14.	OATS	Planted	533	533	533	533
15.		For Grain	534	534	534	534
16.	BARLEY	Planted	535	535	535	535
17.		For Grain	536	536	536	536
18.	CORN	Planted	530	530	530	530
19.		For Grain	531	531	531	531
20.	SORGHUM	Planted	603	603	603	603
21.		For Grain	604	604	604	604
22.	OTHER USES OF GRAINS PLANTED — Acres abandoned, cut for hay, etc.	Use				
		Acres				
23.	HAY Cut and to be cut	ALFALFA AND ALFALFA MIXTURES	653	653	653	653
		OTHER HAY Kind				
24.		Acres	65—	65—	65—	65—
25.	SOYBEANS		600	600	600	600
26.	TOBACCO		67—	67—	67—	67—
27.	PEANUTS		690	690	690	690
28.	RICE		605	605	605	605
29.	COTTON, UPLAND	Planted	524	524	524	524
30.		Abandoned	523	523	523	523
35.	IRISH POTATOES		552	552	552	552
36.	OTHER CROPS	Name				
		Acres planted or in use				
38.	CROPS PLANTED FOR SOIL IMPROVEMENT ONLY — No other use during 1972		856	856	856	856
39.	IDLE CROPLAND - Idle all during 1972		857	857	857	857

Special Keypunch Instructions

1. Punch 76 in column 1-2 for all cards.

2. Face page: Punch identification as appears on face page upper right hand corner.

3. Page 2

   a) Punch field numbers as they appear at the top of page.

   b) Item 5: Leave crop code blank and punch acres 'as is.'

      Item 6-9: Punch code and acres 'as is.'

      Item 10-21: Punch code and planted acres only.

      Item 22: Skip.

      Item 23-39: Punch code and acres 'as is.'

4. Page 3 on

   a) Punch field numbers as they appear at the top of page.

   b) Other Land: Leave crop code blank and punch acres 'as is.'

      Permanent Pasture and Cropland Pasture: Punch crop code and acres 'as is.'

      After Cropland Pasture through Sorghum for Grain: Punch crop code and planted acres only.

      Alfalfa Hay through Idle Cropland: Punch crop code and acres 'as is.'

5. Verify.

   Note: a) Punch acres to one decimal without the decimal point.

   b) Right justify and punch lead zeros for all numbers.

   c) There will be only one code and one acreage figure punched per field number. The proper code and acreage will always be the first entry in a column for any field number.

ERTS SSO Keypunch Instructions

1. Do not punch blanks or edited out data fields.
2. Punch only current survey data.
3. Punch permission code only on the first card.
4. Punch the first four alpha characters of the recorded condition.

APPENDIX C

Grey-Scale Map Computer Program

WMAP

This program will:

1. Map directly from LANDSAT MSS Bulk data tapes (either the original) non-labeled tapes or standard label copies).

2. Map from any one of the four MSS response bands (LANDSAT channels 4, 5, 6, or 7).

3. Compute a histogram of a sample of a designated area and compute grey scale boundaries for the mapping from this histogram. 1/ The user may specify as many as 16 grey scale classes. The user may also specify:

   a. as to whether or not the program will assign boundaries so that each grey scale class will contain about the same number of data points,

   b. If the number of data points in a class will be proportional to the square roots of the percentage distributions of the different response levels found, or

   c. that the program use limits which are defined by the user.

4. For very large areas, will map about 14,000 characters a second (CPU time) on the WCC IBM 370-168. For smaller areas, e.g. 100 lines and 100 columns the mapping rate decreases to around 5,000 characters per second.

The USER MUST:

1. Specify the response band to be mapped (default is LANDSAT Band 5.

2. Specify the number (k) of grey scale divisions to be used in the mapping (default is 9).

3. Specify a printable character for each grey scale division.

4. Specify the location of the areas to be sampled for the frequency tabulation and/or to be mapped.

1/ If the total number of data points in the sample area is less than 10,000, then the histogram will include all data points in the designated area.

Control cards required for each run are:

1. A CLASS card which will define:

   a. the response band to be mapped (punch in column 14).

   b. the number ($k \leq 16$) of grey-levels to be used in the mapping (punch in columns 15-16, right justified).

   c. the string of printable characters to be used in the mapping (punch these in consecutive one-column fields starting with column 18. The first character will be used for the lowest level set of response values. Blanks in the string will cause a blank to be printed for that level(s) on the map).

   If mapping in LANDSAT band 7 (MSS channel 4), any data points having values of 1 to 5 (deep water) and 6 to 9 (shallow water) will be assigned the characters punched in columns 18 and 19 of the class card. Therefore, when mapping in band 7, the user should specify (k+2) printable characters on the CLASS card.

2. At least 1 SAMPLE/MAP AREA block card.

   A SAMPLE AREA card defines an area on the tape which is to be sampled for the frequency distribution to be used in determining the class levels for the printout. The first card after the CLASS card will <u>always</u> be treated as a SAMPLE AREA card. Any later card which has a '1' punched in column 20 will also be used as a SAMPLE AREA card. Any SAMPLE AREA card which has a '1' in column 24 will also be treated as a MAP AREA card.

   The format for the SAMPLE AREA card is:

C.C.	
1-4	the number of scan lines to be skipped.
5-8	the number of the last scan line in the desired area.
9-12	the number of data points to the left of the desired area.
13-16	the number[1/] of the last data point to be included in the desired area.

[1/] Columns should always be numbered in conformance with the LARS System whereby data points 1-804 are on tape 1, 805-1614 on tape 2, 1615-2424 on tape 3, and 2425-3228 on tape 4.

20 a '1' (optional if first SAMPLE AREA card).

32 a '1' (optional, to be used only if class limits are to be assigned by means of the square root transformation).

The MAP AREA card defines an area for which a grey-scale printout is to be produced. As with the SAMPLE AREA card, a single MAP AREA card can define an area as large as the tape itself (1/4 of an LANDSAT frame) or anything smaller. However, the output will be in 120 column strips. The format for the MAP AREA card will be the same as for the SAMPLE AREA card EXCEPT that:

1. A '1' is also punched in column 24 (optional unless a '1' has been punched in column 20).

2. A '1' in column 28 indicates that the user has inserted a 'LIMITS'card after that MAP AREA card.

LIMITS Card

The LIMITS card enables the user to specify the class limits to be used for a particular map area, regardless of the values computed by the program. The values established by a LIMITS card will continue to be used for succeeding map areas until the next LIMITS card is read.

The values to be punched on the LIMITS card will be the upper boundaries of the k grey-scale divisions. They are to be punched in consecutive four digit integer fields, from smallest to largest.

JCL and Control Card Sequence

for Program WMAP in USDA Washington Computer Center

Label parameter is for non-labeled LANDSAT tape

```
/*
as many additional SAMPLE and/or MAP area cards as desired
initial SAMPLE area card, may also be a MAP area card
a CLASS card
//GO.SYSIN DD *
//GO.FT10F001 DD SYSOUT=(c,,8431),DCB=RECFM=FBA
//G(.FT09F001 DD SYSOUT=(c,,8431),DCB=RECFM=FBA
//          BLKSIZE=3320),LABEL=(,NL,,IN),BOL=SER=  6   ,DSN=
//GO.FT08F001 DD UNIT=2400,DISP=(OLD,PASS),DCB=(RECFM=U,
//          EXEC RADLGO,P=RADMAP
     job cards
```

APPENDIX D

Detailed Instructions for

Microdensitometer Scanning of Aerial Photography

APPENDIX E

A PROGRAM TO CONVERT PDS MICRODENSITOMETER SCAN LINES INTO SAS COMPATIBLE OBSERVATIONS

This program is designed to convert a Photometric Data System (PDS) microdensitometer scan into a Statistical Analysis System (SAS) compatible multivariate observation. Up to 4 scans of the same area may be included in the SAS observation.

The user controls the number of scans (normally 1 for each filter) to be used in building the multivariate observations. The microdensitometer scans are read in serially and saved on temporary files. After all the data for a given picture section (pisect) has been read in, the temporary files are rewound and read back a line at a time, and a SAS observation produced for each point in the line. Each observation consists of data from corresponding points from all scans used.

The program is divided into 3 phases: (1) parameter phase, (2) read phase, and (3) combine phase. The normal operation of the program is to go from phase 1, to phase 2, to phase 3, and repeat as desired.

Parameter phase:

Allows the user to define the initial settings from all counters, and indicators used during the read and combine phases. If fatal errors occur during the run, control reverts to the parameter phase for an error scan of all remaining control cards, but no data will be processed.

Read phase:

During the read phase,microdensitometer scans are read in and stored on temporary files. During this process, the PDS 9-track format is converted to a 8 bit internal IBM notation. If the data were scanned in a raster or right edge scan, it would be converted to a left edge scan. The user, however, may elect to cancel this option and accept the data in the order scanned. While in read phase, all parameter definition cards are ignored. If an attempt is made to read more than 4 scans, the combine phase is automatically entered.

Combine phase:

This phase combines the results of the read phase. Corresponding points from each read file are included in each SAS observation produced. The data from the reads are put in correspondence with the data items in the SAS observation set. If these are fewer than 4 scans to be combined, the trailing data items are assigned the missing value. The coordinate values and pixel serial numbers are computed and assigned as each observation is produced. At the conclusion of this phase, control reverts to the parameter phase,and new parameter settings will be accepted.

NUMERIC VALUE REPRESENTATION

The microdensitometer output is a digital representation of an analog signal. The amount of light passing through a sample is converted into a voltage by a photo-multiplier tube. If transmissions are being recorded, the voltage is routed to the panel display meter and then to the analog to digital (A/D) converter. If optical densities are being recorded, the voltage is first sent to a logarithmic converter before going to the panel display meter and then to the A/D converter.

The A/D converter produces a positive integer value that represents the voltage. The input range of the A/D converter is 0.00 to 5.12 volts in .005 volt increments. The digital output ranges from 0 to 1024, or 200 times the voltage input. It is important to remember that these values could be either transmission or density depending on the calibration settings.

When the digital output from the A/D converter is stored in the computer (PDP8) it is multiplied by 2 and is now 400 times the value shown on the panel meter. This is done to reduce the effect of noise contamination. Some noise could result from the fact that the microdensitometer actually takes discrete readings from a continuously varying function.

The data values are recorded in a 9-track tape format. The PDP8 computer is a 12 bit word machine with 6 bit bytes and is not directly compatible with the 9-track 8 bit byte tape format. Therefore, 2 zero pad bits are appended to each PDP8 byte as it is written in a 9-track format. Physically, the data on tape has the format shown below:

ppsddddddppdddddn

where p represents the pad bits appended to fill the 9-track tape format,
- s is the PDP8 sign bit and is normally 0,
- d represents one of the 10 data bits from the A/D converter,
- n represents the noise bit position, normally 0.

In reconstructing the microdensitometer data back into a useable form, the program allows the user two choices. By default, values will be produced from storage type data. Optionally, actual panel display values may be generated.

Storage data has been reduced to a form which is suitable for bulk storage Each value is reduced to an 8 bit integer and requires exactly 1 byte of storage. This is the form used by ERTS, LARSYS, and the Penn State Classification System.

The numeric range of the integer valued data is from 0 to 255. Approximate panel values may be derived by multiplying a storage value by .02. At first, it may seem that we are discarding valid data, but this is not so if we consider the accuracy of the microdensitometer.

The microdensitometer specifies linearity of $\pm$.02 density or .5% transmission, and that the drift for a 10 hour period is less than $\pm$.02 density or less than 1% transmission. This means that a recorded value could differ from the true value by as much as .04 density or 1.5% transmission. The stored values will resolve density to the nearest .02 units and transmission to the nearest .4% (.3921569), which is within the limits of the equipment.

The PanelData option allows the reconstruction of exact panel readings as shown by the panel display meter. The data accuracy implied is beyond the capability of equipment, but is should be useful in checking machine specifications.

X Y COORDINATE SYSTEM

The program assumes a generalized coordinate reference system. The x,y coordinates are signed integers, with (0,0) as the default origin. The x ordinate is the element index, and the y ordinate is the line index. The program always assigns the algebraically smallest x,y value to the pixel in the northwest corner (upper left). The x ordinate increases as the scan moves to the east (right), and the y ordinate increases as the lines move south (down).

The PDS microdensitometer normally scans lines in a raster (back & forth) with the direction of scan alternating, and can scan lines from top to bottom or bottom to top. The Photometric Data System Conversion to Microdensitometer Scan (PDSCMS) program has the ability to determine the scanning directions, and use this in the coordinate assignment algorithm. Thus, regardless of how the points are scanned, the above defined coordinate reference system is valid.

The program computes the coordinates during the combine phase. The coordinates of the physically first point are computed and assigned to that point. If this point is not the northwest corner point, the coordinate of the northwest corner point are derived. The program prints out the northwest corner coordinates as the first x and y ordinates.

The above described coordinate reference system may seem unduly complicated, but it (1) sets up a reference system that is both hardware and software compatible, and (2) permits full use of the microdensitometer scanning ability.

Display devices such as line printers and CRD devices, display data from left to right and top to bottom. The natural order of computer indexing is from smallest to highest. Thus, after coordinates are assigned. data points may be sorted by coordinate and they will be in the natural order for computer processing regardless of how scanned.

The user may have several scans from a scene with the microdensitometer defining the origin at each pisect. The conversion software would call that point (0,0) by default. Later, the user may wish to restore or assign relative position of pisects by relocation. The user could also move the origins of all pisects from the microdensitometer (0,0) setting to any arbitrary point (n,n).

The user may have the microdensitometer scan several pisects from a scene relative to a common origin. The conversion software will compute initial coordinates for each pisect using the microdensitometer supplied locations. Thus, the resulting pixel coordinate will preserve the relative spatial location of the pisects relative to the scene origin. Later, the user may wish to perform an origin transformation, and spatially locate this scene relative to any other independently scanned scene.

SAS OBSERVATIONS

Each observation produced has 11 items as follows:

SCENE-NAME 1-8 characters left justified with trailing blanks in bytes 5-12.

This name is used to identify a collection of pisects (picture sections). If the user fails to supply a valid name, the program will use the current date in the form mm/dd/yy by default.

PISECT-NAME 1-8 characters left justified with trailing blanks in bytes 13-20.

This name is used to identify a pisect within a scene. A new name is supplied for each pisect processed. If the user fails to supply a valid name, the program will use the current value of the system clock in the form hh·mm·ss by default.

GROUP-NAME 1-8 characters left justified with trailing blanks in bytes 21-28.

This name is used to identify calibration data. A null or 'blank' name indicates unknown data. The discriminant function, uses named groups as training, and classifies unknown data. If the user fails to supply a valid name, the program supplies the null or 'blank' name by default.

IDENT-NAME 1-8 characters left justified with trailing blanks in bytes 29-36.

This name is used to establish user identity of unknown data. A null or 'blank' name indicates that the user does not know or cannot identify the item. Valid ident-names are taken from the set of group names. The discriminate function would use the ident-name to check classification accuracy. Of the user fails to provide a valid name, the program supplies the null or 'blank' name by default.

XORD integer binary in bytes 37-40.

This is the relative position of the SAS observation within a line of data. It always gives relative element position within its own pisect, and depending on user options may be positional relative to an entire scene or group of scenes.

YORD integer binary in bytes 41-44.

This is the relative line position of the SAS observation. It always gives relative line position within its own pisect, and depending on user options may be positional relative to an entire scene or group of scenes.

PSN integer binary in bytes 45-48.

This is the pixel serial number assigned by the program. Pixels are serialized in order processed in the combine phase. Unless directed otherwise, pixels are serialized for the entire run starting with 1. The serial number may be signed.

PIXF1V real binary in bytes 49-52.

This is the microdensitometer value for the first scan read for the current pisect. It will never be assigned the missing value. 1/

PIXF2V real binary in bytes 53-56.

This is the microdensitometer value for the second scan read in for the current pisect. If there was no second scan, it takes on the missing value.

PIXF3V real binary in bytes 57-60.

This is the microdensitometer value for the third scan read in for current pisect. If there was no third scan, it takes on the missing value.

PIXF45V real binary in bytes 61-64.

This is the microdensitometer value for the fourth scan read in for the current pisect. If there was no fourth scan, it takes on the missing value.

The program writes the SAS compatible file in binary (unformatted) variable blocked spanned mode. (RECFM=VBS). Because SAS includes the record description word as part of the record, the byte locations of all items have been offset by 4 bytes in the above description.

1/ The missing value is a floating point -0, or in hexidecimal 80000000.

CONTROL CARDS

The program uses 14 different control cards. Most of them are optional because the program will supply default values when the user does not. Each control card is divided into 3 major fields as follows: (1) Key word or op-code in columns 1-8; (2) parameter field in columns 11-50; and, (3) comments field in columns 51-80.

There are 4 classes of control cards, depending on the kind of action to be performed. Each class is described separately below:

Class 1 - Run Cards

These cards set indicators that remain in effect for the duration of the run or until redefined during the run. All run cards are optional.

EDGE Card

cols 1-8 EDGE

This card causes the program to convert to all scans to a left edge scan. This effectively removes the raster produced by the back and forth microdensitometer scanning motion. All lines are running from right to left are turned around. If an EDGE card is not supplied, it is assumed.

ASIS Card

cols 1-8 ASIS

This card causes the program to accept the data points in the order scanned. However, the x,y coordinate assigned are computed based on line direction. If the pixels are sorted based on the x,y coordinates, a normal picture will be produced. That is, the true northwest corner point has the algebraically smallest coordinates, and the southeast corner has the algebraically largest coordinates. If an ASIS card is not supplied, EDGE is assumed by default.

ABL Card

cols 1-8 ABL

This card causes the program to accept microdensitometer data sets that have identified with blank or first character blank labels. By default such scans are rejected as a fatal error. Note that once turned on this option cannot be rescinded during a computer run.

VALUE Card

cols 1-8 VALUE

cols 11-18 STORAGE
PANEL

This card allows the user to select the type of numeric values to produce for the SAS file. Storage values are normalized floating point integers, range $0 \leq$ value 255. Panel values are also normalized floating point, but is the microdensitometer A/D converter output expressed as a display panel number. The range is $0.000 \leq$ values ≤ 5.115, in increments of .005. A storage value is numerically 50 times the panel value with the decimal fraction truncated.

Class 2 - Scene Cards

These cards set parameters that apply only to the scene about to be processed. They are automatically cleared to default values after a STACK control card. All scene cards are optional.

SCENE Card

cols 1-8 SCENE

cols 11-18 1-8 character name left justified with trailing blanks used to identify a group of pisects. The contents of columns 11-18 are placed in the scene-name field of the SAS compatible record. If the user does not make a scene, the program supplies the current date by default.

PSN Card

cols 1-8 PSN

cols 11-15 signed integer constant starting serial number.

This card can be used to extend the serialization of previous computer runs. If the user does not supply a starting serial number, a value of 1 will be used by default. The STACK control card resets PSN to 1.

ORIGIN Card

cols 1-8 ORIGIN

cols 11-15 signed integer constant x coordinate offset.

cols 16-20 signed integer constant y coordinate offset.

This control card is used to provide origin translation of each pisect processed. The coordinates of the first point are computed and the offset applied. It may be used to relate the pisects from the current scene to those in a previous or subsequent scene. This feature may be useful when the data are from sequential scenes such as aircraft photography.

Class 3 - Pisect Cards

These cards set parameters that apply only to the pisect about to be processed. They are automatically cleared to default values after the COMBINE or STACK control card. All pisect cards are optional.

PISECT Card

cols 1-8 PISECT

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are saved in the pisect name in the SAS Compatible Record. It serves to identify pisects within scenes. If the user does not supply a PISECT card, the program uses the current value of the system clock by default.

GROUP Card

cols 1-8 GROUP

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are placed on the group field in the SAS Compatible Record. A non-blank name indicates that this pisect contains calibration data for a specific groups. If the user does not supply a group name, the program inserts a blank name by default.

IDENT Card

cols 1-8 IDENT

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are placed in the ident-name field of the SAS Compatible Record. A non-blank name indicates that the user has identified the points in this pisect as belonging to the specified group. If the user does not supply an IDENT, the program inserts blanks by default.

RELOCATE Card

cols 1-8 RELOCATE

cols 11-15 signed integer constant representing the northwest x ordinate.

cols 16-20 signed integer constant representing the northwest y ordinate.

The northwest corner pixel will be assigned the coordinates given on this card. All subsequent pixels will be assigned coordinates relative to these. Thus, any pisect can be arbitrarily moved in space. By default, absolute relocation will not be performed.

This card overrides the origin transformation in effect for each pisect for which relocation is performed. The origin transformation will be performed for each pisect not relocated.

Class 4 - File Manipulation Cards

These control cards cause data to be moved from one file to another, and to perform some transformations in the process. These cards are required as specified below.

READ Card

cols 1-8 READ

cols 11-50 1-40 character name left justified with trailing blanks.

This card causes the program to read in 1 PDS microdensitometer scan, stored on a temporary file. One read card is required for each scan to be included in a SAS observation. When a read card is processed, while the program is in the parameter phase, control is switched to the read phase. No more parameter cards will be honored until control reverts back to the parameter phase.

Up to 4 consecutive read cards will be honored. If a 5th read card is encountered, the program will combine the 4 scans already stored on temporary files, and then scan the remaining control cards for errors. No more data will be transferred. Either an end-of-file, a combine card, or a stack card must follow read cards.

The 1-40 character name is used for label checking as follows:

(1) If the name is absent or begins with a blank, the program assumes that no label checking is to be performed, and whatever file it finds is assumed to be correct.

(2) If a name is present, it must match the label put in the scan line by the microdensitometer operator. Label checking is performed up to the first blank character in the supplied name. Thus, if the user has a common prefix for a series of scans, he may use an abbreviated label to verify that the correct scans are being processed. If the label check fails, no more files are processed, but the remaining control cards are checked for errors.

COMBINE Card

cols 1-8 COMBINE

This card causes the program to combine the results of the previous reads and add the results to the SAS compatible data set being built. If n scans are being combined, exactly n-1 combine cards are required. The last combine card in the control card stream is optional as any uncombined reads are automatically combined at end-of-file. At the end of a combine operation, the program returns to the parameter phase and will accept parameter control cards.

STACK Card

cols 1-8 STACK

This card is the same as combine in that the results of the previous reads are combined and concatenated to the SAS compatible data set being built. In addition, the data set is endfiled and the scene and pisect indicators cleared to default values. Any control statements following a STACK control card cause PDSCMS to start a new SAS compatible file. This new file may be stacked or separated, depending on the JCL used for the run.

Both STACK and COMBINE cards may be used in the same run, providing at least 1 read operation is performed between them. If a STACK card would be the physically last control card,it can be omitted.

EXECUTING THE PDSCMS PROGRAM

The PDSCMS program is executed by using the RADLGO procedure. The PDS microdensitometer tape is read from unit 8, and the converted file is written on unit 9. Program control cards are read from SYSIN.

The microdensitometer output is a series of stacked data sets on magnetic tape. The program reads as many data sets from the stack as directed by READ control cards by incrementing the unit 9 FORTRAN Sequence Number. Each READ control card requires a unit 9 DD JCL statement with an appropriate sequence number. The data set sequence number in the labels parameter points to the particular scan to be processed by the READ command.

```
//FT08F001  DD  LABEL=(i,NL,,IN)       for first READ card
//FT08F002  DD  LABEL=(j,NL,,IN)       for second READ card
//FT08F003  DD  LABEL=(k,NL,,IN)       for third READ card
     :
     :
//FT08Fnnn  DD  LABEL=(n,NL,,IN)       for nnn'th READ card
```

The letters i, j, k, nnn represent the data set sequence number of the tape file to be processed. They point to the i'th, j'th, k'th, and n'th data set respectively.

The converted SAS file is written on Unit 9 in FORTRAN binary (unformatted) mode as either a single data set or a series of stacked (separated) data sets. Stacking is performed by incrementing the Unit 9 FORTRAN Sequence Number. The DD statement parameters determine if stacking or separation is being performed.

```
//FT09F001  DD  LABEL=p                initial output from PDSCMS
//FT09F002  DD  LABEL=q                after the first STACK card
//FT09F003  DD  LABEL=r                after the second STACK card
     :
     :
//FT09Fmmm  DD  LABEL=s                after the (mmm-1)'th STACK card.
```

The letters p, q, r, s represent the data set sequence number on the tape being produced. If the data sets were being written on disk, separated names would be required.

SAMPLE JCL USING TAPE INPUT & OUTPUT

```
//XO   EXEC   RADLGO,
//            P=PDSCMS
//GO.FT08F001  DD  DISP=OLD,UNIT=2400,DCB=(BLKSIZE=6400,RECFM=U,BUFNO=1),
//                 VOL=SER=URxxxx,
//                 LABEL=(i,NL,,IN)
//GO.FT08F002  DD  DISP=)LD,UNIT=2400,DCB=*,FT08F001,VOL=REF=*.FT08F001,
//                 LABEL=(j,NL,,IN)
            .                                                               .
            .  as many ft08fyyy dd statements as required: extra ones do no harm.
            .                                                               .
//GO.FT08Fnnn  DD  DISP=OLD,UNIT=2400,DCB=*,FT08F001,VOL=REF=*.FT08F001
//                 LABEL=(k,NL,,IN)
//GO.FT09F001  DD  DISP(,PASS),UNIT=2400,DCB=(BLKSIZE=6400,LRECL=32000,RECFM=VBS,
                   BUFNO=1),
//                 DSN=dsname,
//                 LABEL=(p,,,OUT)
//GO.FT09F002  DD  DISP=(OLD,PASS),UNIT=2400,DCB=*.FT09F001,VOL=REF=*,FT09F001,
//                 DSN=*.FT09F001,
//                 LABEL=(q,,,OUT)
            .                                                               .
            .  as many FT09Fyyy dd statements as required: extra ones do no harm.
            .                                                               .
//GO.FT09Fmmm  DD  DISP=(OLD,PASS),UNIT=2400,DCB=*.FT09F001,VOL=REF=*.FT09F001,
//                 DSN=*.FT09F001
//                 LABEL=(r,,,OUT)
        PDSCMS     control cards
/*   EOJ.
```

SAS PROCESSING THE COMPATIBLE FILE

JCL Requirements

In order to process the compatible file with the SAS program, an additional DD statement is required by the RADSAS procedure. This statement is required to point to the file to be used. In the following JCL, the PDSFILE [1/] DD statement is used to gain access to the converted PDS data.

```
//S  EXEC  RADSAS
//PDSFILE  DD  DSN=dsname,DISP=OLD,UNIT=2400,VOL=SER=xxxxxx
//SYSIN  DD  *
      .
      .  sas program statements
      .
/*  EOJ.
```

In the above example, the converted file is assumed to reside on magnetic tape as a single unstacked data set. If the file is not on magnetic tape, or is passed from a previous job step, an appropriate alteration in the PDSFILE DD statement will be required.

If the stack option has been used to stack or separate scan pictures, a separate DD statement is required for each stacked data set to be read in during a given SAS run. If the data sets are stacked on tape, extra DD statements may be left in the job stream whether needed or not. The following JCL illustrates the set up the stacked data sets on tape.

```
//ST  EXEC  RADSAS
//STACK1  DD  DISP=OLD,UNIT=2400,
//            DSN=dsname,VOL=(,RETAIN,SER=xxxxxx),
//            LABEL=p
//STACK2  DD  DISP=OLD,UNIT=2400,DSN=*.STACK1,VOL=(,RETAIN,REF=*.STACK1 ,
//            LABEL=q
       :as many stack DD statements as may be needed: extra ones do no harm.
//STACKn  DD  DISP=OLD,UNIT=2400,DSN=*.STACK1,VOL=(,RETAIN,REF=*.STACK1 ,
//            LABEL=r
//SYSIN  DD  *
       :sas program statements
       .
/*   EOJ.
```

1/ The user may substitute any name for PDSFILE, but that name must also be used in the SAS INPUT statement.

If the converted files are separated on disk, a file must exist for each DD statement in the SAS step. If both PDSCMS and SAS are executed in the same job, SAS DD statements may point to PDSCMS DD statements that were not used in the PDSCMS step. However, if the SAS program is run as a separate job, all the converted files referred to by DD statements must actually exist in order to prevent JCL errors.

SAS Program Statements

The SAS program must be directed to use the PDSFILE DD Statement for its input. The model statements given below can be used to read in all the items from the converted file.

```
DATA;
INPUT  DDNAME=PDSFILE  SCENE $ 5-12  PISECT $ 13-20  GROUP $ 21-28
IDENT $ 29-36  XORD  IB 37-40  YORD  IB 41-44  PSN  IB  45-48
PIXF1V  RB  49-52  PIXF2V  RB 53-56  PIXF3V  RB 57-60
```

The user may not wish to read in all the items. Those items not wanted may be omitted from the list in the input statement. The following statement shows how to read in only the data from the first and third read cards.

```
DATA;
INPUT  DDNAME=PDSFILE  PIXF1V  RB 49-52  PIXF3V  RB 57-60; PIXF4V RB 61-64;
```

In order to read stacked or separated data sets in to the SAS system, the user must provide a separate INPUT statement for each separated file. Each data set referred to by the INPUT processor must actually exist. Data sets that do not exist or have never been created cause SAS to abend.

The following example illustrates a simplified method of reading redundant type data sets by using a SAS macro.

```
MACRO  WHATEVER  SCENE  $  5-12  PISECT $ 13-20  GROUP $ 21-28
IDENT $ 29-36  XORD  IB 37-40  YORD  IB 41-44  PSN  IB 45-48
PIXF1V  RB 49-52  PIXF2V  RB 53-56  PIXF3V  RB 57-60  PIXF4V RB 61-64%
```

(other SAS statements could also be included in the macro to perform special transformations, range checks, etc.)

```
DATA  STK1;  INPUT  DDNAME=STACK1  WHATEVER;
DATA  STK2;  INPUT  DDNAME=STACK2  WHATEVER;
     .
     .  as many statements as required:  extra ones must be removed.
     .
DATA  STKn;  INPUT  DDNAME=STACKn  WHATEVER;
```

The PDSCMS program assigns the missing value to the PIXFiV elements for which there was no corresponding read card. The user can do 1 of 4 things with missing value: (1) accept data with missing values and let SAS handle them, (2) do not read in the pixel filter values that are missing, (3) convert the missing value to some neutral value, or (4) identify and take special action for missing items.

Sample Program to Convert Missing Values

```
PIXF2V=PIXF2V+0;
PIXF3V=PIXF3V+0;
PIXF4V=PIXF4V+0;
```

DATA CONVERSION

Microdensitometer data is expected to be used from a storage format which is an 8 bit integer value from 0 to 255 inclusive. Storage data can either represent densities (logarithmic response), or transmission (linear response). Simple linear transformations are required to reduce storage values into the corresponding panel meter value, optical density, or percent transmission.

Storage values can be converted directly into corresponding panel meter values by multiplying by .02. 1/ The resultant is either an optical density or transmission value, depending on the microdensitometer calibration settings when the scan was performed.

When the microdensitometer is calibrated to record densities, the panel value is optical density. Storage values are increments of .02 density units with a valid range from 0.00 to 4.00 inclusive. Density readings larger than 4.00 constitute an overflow condition because they are beyond the specified range of the equipment.

When the microdensitometer is recording transmissions, the stored data represents an incremental percent transmission that is dependent on the gain setting during calibration. Normally, the gain is set at 5.10 to give maximum range and accuracy to the transmission levels. The incremental step is then .3921569% transmission.

In addition, it may be useful to convert the storage data into, from logarithmic densities into linear transmissions and vice versa. In the following relationships, the transmission calibration (Gain) is assumed to be 5.10. The density is always calibrated to 0.

The following symbols are used in the equations that follow.

SD density (logarithmic) storage value	$0 \leq SD \leq 200$
ST Transmission (linear) storage value	$0 \leq ST \leq 255$
G Gain setting for transmission	nominal value 5.10
PT Percent transmission	$0 \leq PT \leq 100$
OD Optical density	$0 \leq OD \leq 4.00$

1/

Described in the numeric representation section.

The relationship between optical density and transmission is:

$$\text{Density} = -\text{Log}_n\ (1/\text{Transmission})$$

If we impose on this basic relationship, the requirement that 100% transmission is 0 density and 0% transmission is 4.00 density, the equation can be rewritten as:

$$OD = 2 = \log_{10}(PT)$$

$$PT = 10 **(2 - OD)$$

Note that the relationship of 0% transmission = 4.00 optical density requires a mathematical impossibility, namely $\text{Log}_{10}(0) = -2$, and $10^{-2} = 0$. These conditions are definitional and are imposed by the resolution limits of the electronic circuiting in the microdensitometer. During computer processing this limiting point requires special handling. Computationally, the valid conversion ranges for percent transmission and optical density are:

$$0 < PT \leq 100$$

$$4.00 > OD \geq 0$$

Also, be aware that 4.00 optical density can be transformed into the computationally valid percent transmission value .01. If storage transmissions are being produced, the minimum storage value is .39% and is larger than 01. An attempt to produce a storage value for .0% transmission will result in a 0 value.

Because in the density to transmission, computations can be performed over the entire density range, it is possible to computationally extend the valid transmission range beyond 2.3 optical density. An image is digitized in densities and the corresponding percent transmission computed. Thus, a percent transmission values less than .30, can be used in computations, but cannot be produced by the microdensitometer, nor stored in standard form.

The equation to convert stored density data into optical density is:

$$OD = SD * .02$$

The equation to convert stored transmission data into percent transmission is:

$$PD = ST * .3921569 \qquad \text{when } G = 5.10$$
$$PT = ST * (2/G) \qquad 0 < G < 5.10$$

The following transformations are used to convert logarithmic values into linear values and vice versa.

To convert stored density into percent transmission use:

$$PT = 10 ** (2 - SD* .02)$$

To convert stored density into stored transmission use:

$$ST = 10 ** (2.40654 - (SD *.02)) \quad G = 5.10 \text{ implied}$$

To convert Optical Density into stored transmission use:

$$ST = 10 ** (2.40654 - OD) \qquad G = 5.10 \text{ implied}$$

To convert stored transmission into optical density use:

$$SD = (2 - \log_{10} (ST *.3921569))*50 \quad G = 5.10$$

$$SD = (2 - \log_{10} (ST *(2/G)))*50 \quad 0 < G < 5.10$$

To convert percent transmission into stored density use:

$$SD = (2 - \log_{10} (PT))*50$$

APPENDIX F

Field Extraction Program

Version 1 and 2

Introduction:

This program generates SAS program statements and control cards for PDSCMS to facilitate conversion and identification of microdensitometer data into final form suitable for discriminant analysis. It is a special purpose program with few options and little in the way of error checking. It is the user's responsibility to make certain the input data is in the correct form, as described in the input section. There are two versions of the program. The major difference between the two versions is the input required for each. Thus, the input section of this paper is divided into two sections, one for version 1, and the other for version 2. Any other differences between versions will be noted in the appropriate sections. The output from the two versions is identical.

JCL Requirements:

```
//jobcard
/*ROUTE PUNCH LOCAL
//  EXEC  RADGO,
//  P=RSFEP1            THIS CARD FOR VERSION 1.
//  P=RSFEP2            THIS CARD FOR VERSION 2.
//GO.FT08F001  DD  SYSOUT=B
//GO.FT09F001  DD  SYSOUT=A
//GO.SYSIN  DD  *

        {input cards}

/*  EOJ.
```

Output: Output is routed through logical units 6, 7, 8, and 9. Logical units 6 and 9 are for printed output, units 7 and 8 for punched output. The printed output on unit 6 consists of job processing information and images of PDSCMS control cards. The PDSCMS control cards are punched from unit 7. SAS program statements for field extraction are punched from unit 8 and printed from unit 9.

PDSCMS Control Cards:

1. SCENE state name (8 character maximum)
2. PISECT segment number (4 characters)
3. READ label for clear filter
4. READ label for red filter
5. READ label for green filter
6. READ label for blue filter