A STUDY OF BIAS AND VARIANCE IN LANDSAT DATA BASED REGRESSION ESTIMATES FOR CROP SURVEYS USING SIMULATED DATA

by

James C. Lundgren*

Lockheed Engineering and Management Services Company, Inc.
2400 NASA Road 1
Houston, Texas

December 1984

*Research supported under NASA Contract NAS 9-15800

LEMSCO-20239

ACKNOWLEDGEMENTS

The author wishes to express his thanks to Richard Sigman and Raj S. Chhikara for their many valuable comments and suggestions throughout the course of this research and the preparation of this report. He is also indebted to Robert R. J. Mohler for his extensive contribution to the editing of this document.

This work has been facilitated by the United States Department of Agriculture, Statistical Reporting Service, under National Aeronautics and Space Administration contract NAS 9-15800.

TABLE III

DIFFERENCES BETWEEN THE SAMPLE AND POPULATION REGRESSION SLOPES

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	0.723	0.250	0.430	0.565	-0.229	0.689	0.158
2	25	4	0.605	0.147	-0.137	1.409	0.879	0.109	0.288
3	25	4	0.844	-0.119	0.101	0.941	0.295	0.214	0.211
4	25	4	0.452	-0.013	-0.071	1.273	0.403	0.514	-0.019
5	25	4	0.790	0.077	0.198	1.172	0.183	0.106	0.080
6	25	4	0.628	0.150	0.013	1.447	-0.150	0.349	-0.017
7	25	6	0.636	0.039	-0.091	-0.302	0.199	0.153	-0.073
8	25	6	0.306	-0.017	0.107	1.169	-0.291	0.290	-0.037
9	25	6	1.049	0.130	-0.022	-0.057	0.057	0.210	0.002
10	25	6	0.550	0.042	0.010	0.380	0.059	0.022	0.064
11	25	6	-0.129	-0.003	0.264	0.256	-0.044	0.077	0.186
12	25	6	0.344	0.161	0.152	0.002	-0.056	0.062	0.111
13	50	4	0.605	-0.018	-0.203	1.002	0.349	-0.023	0.203
14	50	4	1.007	-0.001	0.083	1.663	0.267	0.117	0.074
15	50	4	1.540	0.076	-0.340	0.220	0.126	0.464	0.242
16	50	4	0.026	-0.237	0.018	1.760	0.208	0.750	0.131
17	50	4	0.533	0.010	0.149	0.711	0.383	0.567	0.126
18	50	4	0.123	-0.064	0.069	1.328	0.221	0.121	0.142
19	50	6	0.322	-0.023	0.184	0.724	-0.091	0.296	0.098
20	50	6	0.144	0.044	-0.019	1.279	0.235	0.351	-0.202
21	50	6	-0.217	0.006	-0.067	0.613	0.381	0.203	0.006
22	50	6	0.192	0.008	-0.027	0.545	0.088	0.066	0.240
23	50	6	-0.009	0.045	-0.189	0.492	0.057	0.171	0.200
24	50	6	0.209	-0.164	0.120	0.367	0.389	0.069	0.166
25	50	10	0.455	0.158	0.128	1.407	0.110	0.079	0.066
26	50	10	0.249	0.283	0.125	0.495	0.209	0.233	0.073
27	50	10	-0.022	0.030	-0.004	0.557	0.373	0.194	0.018
28	50	10	0.143	-0.048	0.060	1.104	-0.024	0.172	0.121
29	50	10	0.455	0.033	-0.076	0.207	0.199	0.120	0.126
30	50	10	0.052	-0.031	0.041	1.562	0.127	0.174	0.013
31	100	4	-0.123	-0.028	-0.046	1.488	0.748	0.147	-0.057
32	100	4	-0.090	-0.078	0.076	1.048	0.584	0.071	0.217
33	100	6	0.161	0.110	-0.004	1.774	0.196	0.240	0.067
34	100	6	0.290	-0.030	0.069	0.782	0.383	0.372	0.183
35	100	10	0.169	-0.066	-0.018	1.292	0.243	0.162	0.031
36	100	10	0.519	0.015	0.143	0.235	0.160	0.215	-0.004
37	100	15	0.397	-0.024	-0.019	0.822	0.141	0.155	0.122
38	100	15	0.139	-0.015	-0.081	0.185	0.015	0.134	0.070
NO. NEGATIVE DIFFERENCES			6*	18	17	2*	7*	1*	7*
MEAN DIFFERENCE			0.370	0.022	0.030	0.840	0.194	0.221	0.090
PERCENT OF MEAN SLOPE			22.6	2.3	3.1	49.7	24.6	108.6	26.5

* THE DIFFERENCE IS SIGNIFICANT AT THE FIVE PERCENT LEVEL.

TABLE IV

SLOPE OF POPULATION REGRESSION LINE

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	1.87	0.82	0.65	0.79	1.18	0.06	0.18
2	25	4	0.88	0.81	1.19	1.86	1.04	0.42	0.30
3	25	4	2.00	1.09	1.07	1.28	0.82	0.59	0.11
4	25	4	1.58	1.20	1.01	2.05	0.63	0.27	0.61
5	25	4	1.38	1.07	0.89	1.46	0.87	0.14	0.17
6	25	4	1.54	0.82	1.16	1.90	0.77	0.26	0.57
7	25	6	0.98	0.82	1.27	2.57	0.95	0.46	0.35
8	25	6	1.98	0.80	1.17	2.04	0.97	0.25	0.45
9	25	6	1.09	0.81	1.24	2.29	1.01	0.51	0.35
10	25	6	1.01	0.82	1.24	2.01	1.03	0.42	0.32
11	25	6	1.82	1.03	0.81	2.27	0.77	0.14	0.08
12	25	6	2.08	0.76	0.65	1.03	1.35	0.13	0.19
13	50	4	1.31	1.00	0.89	1.49	0.84	0.41	0.25
14	50	4	1.87	0.97	0.86	1.04	0.85	0.30	0.18
15	50	4	1.49	0.97	1.11	1.82	0.63	0.20	0.28
16	50	4	1.77	1.68	1.08	1.17	0.52	0.13	0.92
17	50	4	0.97	0.71	0.96	1.42	0.85	0.33	0.43
18	50	4	1.54	0.88	0.78	1.01	0.69	0.19	0.17
19	50	6	1.68	0.86	0.81	1.20	0.84	0.23	0.19
20	50	6	1.85	1.21	1.05	1.66	0.55	0.17	0.78
21	50	6	2.15	1.06	1.13	1.74	0.58	-0.05	0.43
22	50	6	1.72	0.97	0.85	1.34	0.86	0.31	0.20
23	50	6	1.70	0.98	1.02	2.09	0.68	0.20	0.26
24	50	6	1.61	0.98	0.89	2.24	0.80	0.37	0.33
25	50	10	1.28	0.73	1.05	2.09	0.98	0.35	0.41
26	50	10	1.32	0.92	0.83	1.56	0.60	0.02	0.32
27	50	10	2.05	1.06	1.09	1.77	0.61	-0.02	0.41
28	50	10	1.68	0.97	0.92	1.55	0.93	0.35	0.19
29	50	10	1.82	0.97	1.00	1.97	0.74	0.23	0.27
30	50	10	1.89	1.21	1.05	2.11	0.56	0.30	0.65
31	100	4	1.77	1.01	1.04	1.38	0.43	0.11	0.21
32	100	4	1.85	0.92	0.90	0.91	0.85	0.26	0.35
33	100	6	1.78	0.91	0.83	1.12	0.92	0.31	0.40
34	100	6	1.95	1.00	0.98	1.50	0.46	0.12	0.29
35	100	10	1.55	0.98	0.99	2.53	0.90	0.29	0.27
36	100	10	1.35	0.88	0.82	1.79	0.64	0.16	0.31
37	100	15	1.92	1.07	1.01	2.11	0.64	0.21	0.44
38	100	15	2.28	0.99	1.03	1.96	0.59	0.11	0.28
MEAN			1.64	0.97	0.98	1.69	0.79	0.24	0.34

The mean difference for all ground covers is positive. This suggests that the slope of the sample tends to be larger than that of the population (see Table IV for population slopes). Chhikara and Houston (1984) have shown that the relationship between Y (the ground truth segment proportion) and X (the classified segment proportion) can be described in terms of ϕ_1 and ϕ_0 (the proportion of incorrectly classified among those classified pixels into class 1 and 0). An examination of their equation:

$$Y = \phi_0 + (1 - \phi_1 - \phi_0)X$$

suggests why the sample slopes are generally larger than the population. The term $(1 - \phi_1 - \phi_0)$ which is analogous to the slope will tend to be larger whenever the sum of the two error terms is smaller. One could argue that since the classifier is trained and thus optimalized on the same sample segments used in computing the regression equation, their errors will be less than the other segments of the population.

Another source of bias in the regression estimates is that the mean of the number of pixels classified into each crop for a training sample is not always an unbiased estimate of the population mean. Shown in Table V are the mean differences per segment in these numbers for the training sample and the population. The same three ground covers which had significant biases in their regression estimates also have significant biases in the number of pixels classified. Since the number of pixels classified is equal to the number of pixels in the population, the biases tend to compensate for each other and can be either positive or negative. These biases give further evidence that the training sample is different from the rest of the population and, therefore, its use in the regression analysis may tend to bias the estimates.

Table VI has been included to demonstrate the appearance of random but non-significant differences. It contains analogous differences between the same training samples and populations for the mean number of pixels per segment by ground truth. Because the ground truth is not affected by the classification process, Table VI contains only a simple random

TABLE V

DIFFERENCES BETWEEN THE CLASSIFIED NUMBER OF PIXELS: SAMPLE AND POPULATION

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	9.925	4.062	2.490	-10.855	3.638	-2.470	0.183
2	25	4	2.815	-2.230	10.927	-9.470	3.072	1.220	-9.000
3	25	4	-2.245	-14.358	4.820	-3.870	-1.913	2.887	2.562
4	25	4	1.327	5.472	9.090	-9.943	-8.318	2.788	-4.875
5	25	4	3.855	-13.530	8.742	-4.528	3.407	-4.933	2.450
6	25	4	5.952	0.905	8.770	-7.365	1.240	-5.550	0.443
7	25	6	-0.463	8.833	-0.590	2.322	-4.008	3.375	-4.257
8	25	6	1.113	-5.090	1.672	-0.288	1.312	-0.638	1.365
9	25	6	-1.412	10.127	3.888	-0.958	-8.815	4.477	-0.295
10	25	6	2.168	-1.508	5.607	-0.013	-1.927	1.440	-3.355
11	25	6	6.525	-2.268	4.015	-2.725	-5.393	-4.007	2.458
12	25	6	9.152	-7.662	1.445	-3.885	0.225	-0.813	3.380
13	50	4	10.672	-6.578	15.650	-12.003	2.642	-11.755	-8.310
14	50	4	-1.383	-1.823	5.132	-10.110	0.852	2.425	-1.985
15	50	4	-5.105	9.812	7.412	-13.445	-3.095	-9.810	3.130
16	50	4	10.507	-5.915	6.232	-6.790	-4.200	-0.630	10.207
17	50	4	4.490	-12.648	29.277	-15.188	-1.033	1.292	-3.545
18	50	4	16.672	4.330	7.500	-13.663	-3.295	0.352	1.367
19	50	6	10.863	-6.132	6.838	-3.005	-2.098	-0.625	1.858
20	50	6	4.787	0.038	8.992	-4.750	-1.292	-2.502	-2.780
21	50	6	6.837	-5.633	13.458	-7.505	-0.093	-4.900	-2.223
22	50	6	7.447	-10.663	-0.820	-6.007	0.017	9.595	-0.597
23	50	6	2.010	-1.822	-5.793	-5.103	-1.802	-0.415	1.638
24	50	6	1.570	6.350	1.142	-2.082	-5.578	-6.682	1.335
25	50	10	3.915	0.390	1.640	-4.275	-0.795	0.690	-0.515
26	50	10	0.645	0.090	5.560	-0.510	-5.860	-6.285	1.545
27	50	10	-0.325	8.675	2.905	-3.395	-5.195	-1.415	-1.040
28	50	10	6.250	5.655	2.765	-3.170	-5.555	-0.820	-2.355
29	50	10	3.940	-1.780	-8.480	-0.350	-1.545	2.680	0.325
30	50	10	-3.025	6.345	-0.610	-1.450	-3.495	-4.535	-3.035
31	100	4	6.992	-8.110	6.387	0.445	-2.278	-7.600	-1.620
32	100	4	0.422	-17.210	12.982	-0.355	3.920	-0.183	-6.768
33	100	6	2.478	4.513	-0.165	-3.255	2.490	1.950	0.567
34	100	6	6.792	5.392	11.962	-6.825	-5.270	-7.665	-2.330
35	100	10	-0.400	-0.490	3.480	-0.555	1.535	-1.570	-1.290
36	100	10	3.355	3.440	4.045	-3.295	-7.020	0.040	2.640
37	100	15	1.387	6.613	9.217	-0.937	1.613	-3.563	-5.160
38	100	15	1.870	-5.420	0.673	-0.153	-1.350	1.813	2.693
NO. NEGATIVE DIFF.			8*	20	6*	36*	25	23	20
MEAN DIFFERENCE			3.747	-1.048	5.480	-4.719	-1.717	-1.377	-0.663
PERCENT OF CROP AREA			1.638	-0.544	5.963	-4.066	-2.343	-2.382	-2.430

* THE DIFFERENCE IS SIGNIFICANT AT THE FIVE PERCENT LEVEL

TABLE VI

DIFFERENCES BETWEEN THE ACTUAL NUMBER OF PIXELS: SAMPLE AND POPULATION

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	-5.398	0.227	-0.528	10.492	-1.787	7.250	-3.093
2	25	4	-1.058	-6.873	2.405	-6.745	9.865	-0.455	0.945
3	25	4	-0.380	-11.260	4.730	-4.768	-0.397	3.112	-2.385
4	25	4	0.307	11.257	5.982	-9.038	-2.018	-11.450	2.380
5	25	4	-22.005	-15.005	-3.375	22.032	2.988	8.202	1.693
6	25	4	-2.265	-0.838	8.690	-9.293	0.352	8.180	0.675
7	25	6	-12.187	11.948	-4.728	9.647	0.815	-0.505	-0.063
8	25	6	-10.582	-0.908	9.257	15.803	-9.102	-4.445	-0.800
9	25	6	9.397	-3.493	-0.020	2.947	-6.485	3.645	1.445
10	25	6	-0.512	-3.927	3.355	9.105	-2.743	-2.630	-1.813
11	25	6	15.087	-5.863	-6.350	-1.547	-1.058	-0.952	-1.420
12	25	6	-3.935	-1.827	-3.615	8.180	5.683	-3.517	0.112
13	50	4	-0.635	2.960	9.522	-22.365	5.932	-3.100	-2.965
14	50	4	-14.273	-9.315	8.990	-2.635	0.210	8.447	3.027
15	50	4	-8.733	8.925	-3.680	-23.470	8.322	3.725	4.705
16	50	4	15.572	-14.380	-7.773	4.990	6.408	0.495	5.310
17	50	4	12.912	-8.463	15.822	-2.663	-6.288	-0.490	-5.945
18	50	4	20.877	-0.708	8.142	-11.093	-3.538	-1.172	2.630
19	50	6	-12.235	1.722	0.688	7.178	-0.875	9.040	-0.753
20	50	6	15.652	-6.780	-4.068	0.965	3.070	-4.438	-2.307
21	50	6	-7.847	-4.997	5.305	-3.750	3.272	10.285	-1.980
22	50	6	9.057	-13.248	-3.752	8.482	-2.215	-1.582	0.365
23	50	6	15.847	-10.908	-10.547	2.155	0.060	-10.583	0.838
24	50	6	-21.027	15.935	-3.873	6.652	-2.347	-1.775	1.818
25	50	10	3.585	9.105	-10.640	-4.330	-1.190	3.500	-0.420
26	50	10	-11.175	0.830	5.800	1.495	-4.595	4.385	-0.310
27	50	10	-3.250	6.965	-5.035	-2.345	1.160	3.600	-0.365
28	50	10	9.215	10.520	-0.360	-0.600	-14.325	0.105	-1.015
29	50	10	7.175	-5.325	-12.620	5.180	-0.700	0.955	-0.730
30	50	10	-16.775	5.495	0.420	0.255	1.195	-2.190	1.185
31	100	4	-20.433	-6.275	0.097	20.407	4.267	0.950	-3.055
32	100	4	-1.683	-25.313	2.255	17.420	9.545	-5.008	-6.640
33	100	6	-14.112	6.708	-3.820	16.445	3.170	-0.028	-0.590
34	100	6	-5.253	10.208	7.918	-14.380	-1.745	7.775	-1.538
35	100	10	-0.150	0.200	-1.980	9.165	-1.320	-4.300	0.010
36	100	10	3.200	-0.805	0.945	-3.805	2.600	-1.685	1.150
37	100	15	6.147	2.020	3.873	0.680	1.963	-4.777	-0.277
38	100	15	2.143	-6.693	-0.020	-3.647	-0.640	5.997	2.363
NO. NEGATIVE DIFFERENCES*			23	22	19	17	19	20	21
MEAN DIFFERENCE			-1.309	-1.531	0.458	1.400	0.198	0.646	-0.206
PERCENT OF CROP AREA			-0.572	-0.794	0.498	1.206	0.270	1.117	-0.755

*NO SIGNIFICANT DIFFERENCES WERE FOUND AT THE FIVE PERCENT LEVEL.

phenomenon with no significant differences. The actual mean number of pixels per segment in each crop for all of the 38 analyses is shown in Table VII.

The estimates of bias in the ratio estimator are depicted in Table VIII. It is similar to Table II in both computation and results. Both estimators have significant biases in the same three ground cover classes.

Illustrated in Figure I is the similarity between the regression and ratio estimators. Depicted is a plot of each bias estimate from Table II (regression estimator) versus its respective value from Table VIII (ratio estimator). A reference line is drawn to indicate where the points would fall if the two estimates were always equal.

There is some concern that the effect of using the training sample to estimate the regression line is confounded with the inherent bias of the regression estimator (Cochran, 1977). Because of the above concern a small study was conducted. Six segments were repeatedly selected at random from a population of 25, after an independent sample of six segments had been used to train the classifier. A regression estimate was computed for all seven crops in each of 20 samples, for each of 20 different classifications of the single population (2800 regression estimates).

Table IX contains the mean biases per crop detected in this analysis. Both the number of significant biases and their magnitudes were unexpected. Only one population was used in the analysis (thus the bias may be population specific). The results suggest a need for further study.

Two results of this study were not surprising. First, there is not a significant overall positive bias in the slope of the sample regression lines. Second, the mean number of classified pixels per segment in the sample is not significantly different from that of the population (for any of the ground covers). These results can be traced to the independence of the samples used to train the classifier and calculate the regression equations. Because the regression sample is selected after classification there is no reason to suspect that the classification

TABLE VII

MEAN NUMBER OF PIXELS PER SEGMENT (GROUND TRUTH: Y)

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	295.96	167.76	82.24	84.52	79.60	60.00	18.48
2	25	4	229.12	177.96	87.72	137.72	73.36	43.48	32.68
3	25	4	237.28	183.16	85.12	124.28	66.96	64.60	23.96
4	25	4	187.68	209.88	77.48	127.60	74.68	55.80	43.92
5	25	4	216.08	201.28	106.20	105.88	66.20	75.16	24.12
6	25	4	182.84	213.20	110.96	132.48	63.76	51.92	24.20
7	25	6	229.12	177.96	87.72	137.72	73.36	43.48	32.68
8	25	6	182.84	213.20	110.96	132.48	63.76	51.92	24.20
9	25	6	229.12	177.96	87.72	137.72	73.36	43.48	32.68
10	25	6	229.12	177.96	87.72	137.72	73.36	43.48	32.68
11	25	6	216.08	201.28	106.20	105.88	66.20	75.16	24.12
12	25	6	295.96	167.76	82.24	84.52	79.60	60.00	18.48
13	50	4	224.26	193.54	96.34	111.44	72.18	73.00	28.24
14	50	4	210.66	207.24	92.16	107.66	79.54	67.14	29.06
15	50	4	244.92	198.60	76.88	120.02	70.54	54.90	24.52
16	50	4	217.94	194.48	94.66	117.06	70.78	54.78	34.04
17	50	4	215.00	176.80	100.04	119.90	81.40	60.14	29.72
18	50	4	256.06	183.12	85.12	100.38	76.40	57.36	21.62
19	50	6	256.06	183.12	85.12	100.38	76.40	57.36	21.62
20	50	6	217.94	194.48	94.66	117.06	70.78	54.78	34.04
21	50	6	234.18	187.68	103.22	115.80	74.22	58.84	22.18
[illegible]	50	6	210.66	207.24	92.16	107.66	79.54	67.14	29.06
[illegible]	50	6	244.92	198.60	76.88	120.02	70.54	54.90	24.52
24	50	6	224.26	193.54	96.34	111.44	72.18	73.00	28.24
25	50	10	215.00	176.80	100.04	119.90	81.40	60.14	29.72
26	50	10	238.92	185.72	89.44	109.14	74.46	59.50	19.98
27	50	10	234.18	187.68	103.22	115.80	74.22	58.88	22.18
28	50	10	210.66	207.24	92.16	107.66	79.54	67.15	29.06
29	50	10	244.92	198.60	76.88	120.02	70.54	54.90	24.52
30	50	10	217.94	194.48	94.66	117.06	70.78	54.78	34.04
31	100	4	237.22	192.45	95.89	118.93	71.07	55.75	27.53
32	100	4	218.27	206.95	87.37	104.33	83.78	49.87	28.49
33	100	6	218.27	206.95	87.37	104.33	83.78	49.87	28.49
34	100	6	237.22	192.45	95.89	118.93	71.07	55.75	27.53
35	100	10	234.06	208.50	89.20	126.68	64.22	57.57	25.71
36	100	10	227.14	188.29	92.64	113.23	69.24	60.24	23.42
37	100	15	231.43	196.54	85.77	118.54	70.66	54.84	29.28
38	100	15	237.22	192.45	95.89	118.93	71.07	55.75	27.53
MEAN			228.70	192.71	91.90	116.07	73.28	57.81	27.28
PERCENT OF TOTAL			29.03	24.46	11.67	14.73	9.30	7.34	3.46

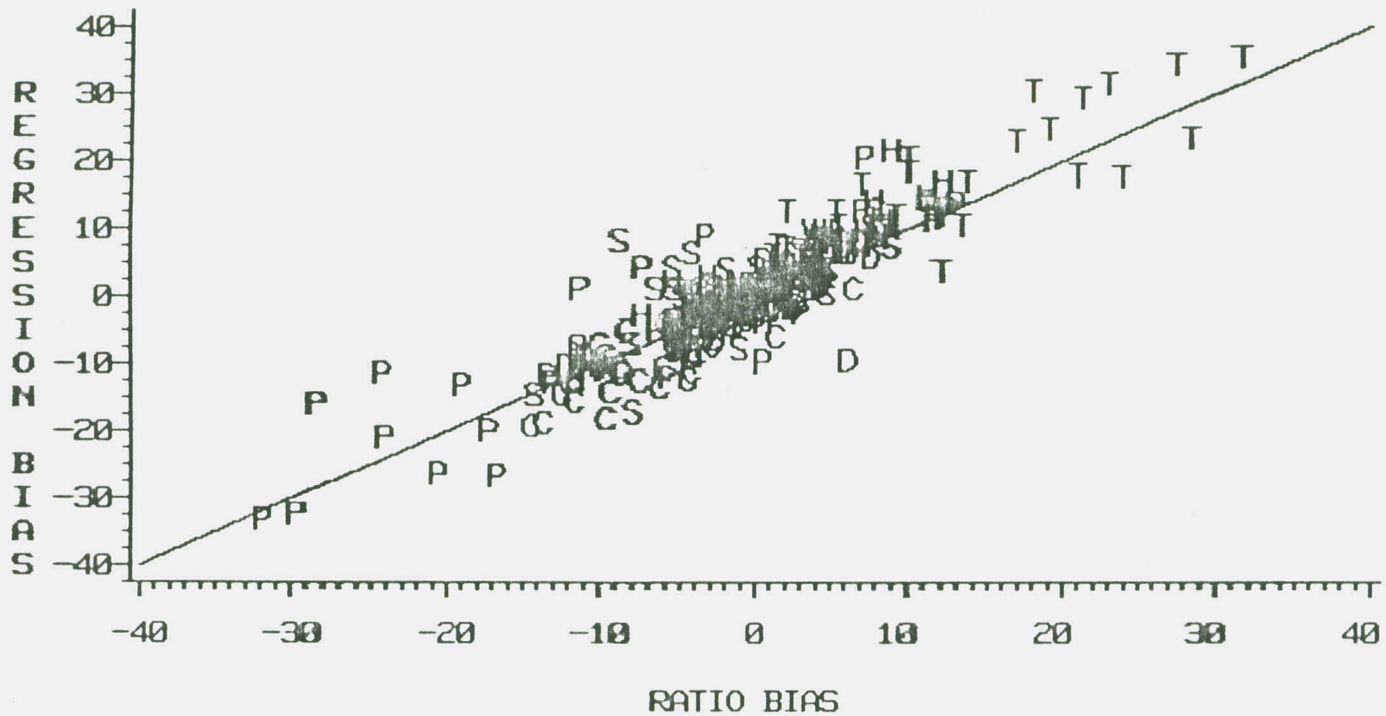
TABLE VIII

DIFFERENCES BETWEEN RATIO ESTIMATE AND POPULATION MEAN Y (BIAS)

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	-28.439	-3.533	-4.596	28.4452	-4.835	8.362	-3.2656
2	25	4	-5.844	-5.400	-7.627	10.5856	6.170	-2.000	4.3412
3	25	4	7.248	2.922	0.042	1.4448	0.474	1.762	-3.4524
4	25	4	-4.420	4.125	-4.054	9.9372	2.569	-11.428	2.1972
5	25	4	-29.666	-4.386	-11.959	31.4876	-0.484	11.567	1.2300
6	25	4	-11.496	-1.139	1.288	5.2580	-3.248	8.934	0.4192
7	25	6	-11.334	2.089	-5.205	3.6436	0.593	-3.210	1.7876
8	25	6	-13.052	4.658	6.412	16.8340	-12.142	-3.330	-0.9456
9	25	6	12.358	-14.301	-3.919	2.9820	-0.526	-0.816	0.9016
10	25	6	-5.407	-3.754	-2.314	9.1772	-1.806	-5.200	-1.0504
11	25	6	2.903	-6.719	-9.711	4.6748	2.393	0.342	-1.5036
12	25	6	-24.206	4.724	-5.418	11.9592	3.829	-2.002	-0.8012
13	50	4	-16.755	8.674	-3.905	1.9896	3.843	3.677	0.5044
14	50	4	-9.420	-8.929	4.760	17.9824	-2.599	3.683	4.0576
15	50	4	-3.474	-2.078	-14.540	-1.4254	6.166	7.823	3.8478
16	50	4	-7.506	-5.044	-13.697	19.1046	7.633	0.426	-3.1024
17	50	4	7.960	2.857	-9.633	20.9530	-8.590	-1.843	-4.9632
18	50	4	-11.336	-5.662	0.857	13.6600	-4.190	-5.642	2.5792
19	50	6	-31.883	7.621	-4.338	13.3042	1.254	8.048	-0.8654
20	50	6	6.999	-8.060	-11.657	9.9382	2.009	-2.016	-2.3282
21	50	6	-20.568	-0.701	-6.148	11.7956	3.790	12.017	-1.7958
22	50	6	-3.416	-2.796	-4.807	23.8326	-3.949	-7.496	-0.7020
23	50	6	12.820	-11.383	-10.113	6.9914	0.966	-10.590	-0.2666
24	50	6	-24.420	7.946	-6.042	11.0752	-0.109	1.329	0.8308
25	50	10	-0.659	8.930	-12.485	5.4476	-2.795	1.966	0.1240
26	50	10	-10.051	0.027	1.312	1.1042	-0.529	7.121	-0.7800
27	50	10	-3.939	-2.789	-8.925	4.7750	4.830	5.168	-0.3188
28	50	10	-4.013	3.927	-3.114	6.5482	-10.946	1.560	-0.0880
29	50	10	-1.358	-4.093	-8.457	1.3750	-0.652	-1.230	-1.0208
30	50	10	-13.354	-1.894	-0.381	3.8008	2.821	-0.728	1.6788
31	100	4	-33.983	2.770	-6.354	22.9473	5.787	4.459	-3.2588
32	100	4	0.497	-7.924	-9.365	21.2584	4.325	-6.547	-5.1429
33	100	6	-18.998	0.150	-4.945	27.4240	1.112	-0.640	-1.5042
34	100	6	-17.272	4.170	-1.571	0.9026	0.949	11.156	-0.7307
35	100	10	-0.457	-0.021	-4.736	11.1143	-4.304	-3.732	0.4858
36	100	10	-3.765	-5.241	-2.304	4.5032	6.926	-2.338	-0.3738
37	100	15	3.423	-5.528	-2.967	2.0477	0.552	-2.546	1.4478
38	100	15	-4.329	-1.701	-0.502	-4.0534	0.075	5.601	0.3531
NO. NEGATIVE DIFFERENCES			30*	23	32*	2*	16	19	22
MEAN DIFFERENCE			-8.44	-1.25	-5.03	10.39	0.19	0.83	-0.30
PERCENT OF CROP AREA			-3.69	-0.65	-5.47	8.95	0.26	1.44	-1.10

* THE BIAS IS SIGNIFICANT AT THE FIVE PERCENT LEVEL.

FIGURE I. BIAS OF REGRESSION VERSUS BIAS OF RATIO ESTIMATOR PER SEGMENT



LEGEND   *P=PASTURE   S=SOYBEANS   C=CORN   T=WASTE*
D=WOODS   H=HAY   W=WINTER WHEAT   A=ALFALFA

TABLE IX

Mean Bias of Regression Estimates for Number of Pixels per Segment
Using Independent Samples for Classification and Regression
Population Size = 25 Segments
Sample Size = 6 Segments

	Pasture	Soybeans	Corn	Waste	Woods	Hay	Wheat
Bias	-0.91	-4.68 *	-1.46	3.84 *	2.83 *	5.07 *	-0.50
t	-0.79	-5.22	-1.38	2.20	2.57	4.80	-1.64
Area (%)	-0.50	-2.20	-1.32	2.90	4.44	8.77	-2.07

* The bias is significantly different from zero at the five percent level of significance.

accuracy for the sample is different from that of the population. Thus, both the slope and the mean of the number of pixels classified should not be different.

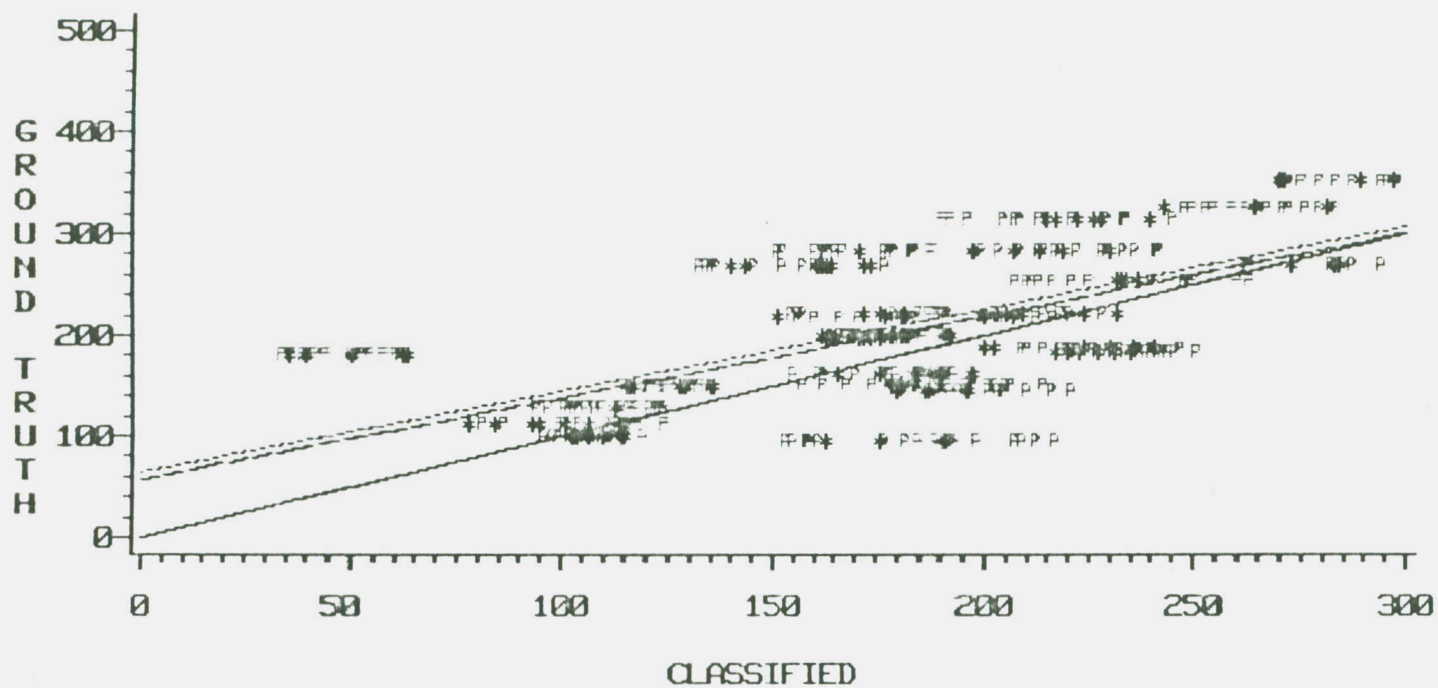
Figures II-VIII contain plots of the number of pixels per segment from ground observation versus the number classified for each ground cover. The data are from the above special study of independent samples for classification and regression. The plots seem to have points forming rows across the page. This effect is the result of using data from all 20 different classifications of the population. Each segment has only one ground truth value but has up to 20 unique classification values which vary over a wide range.

Three lines are drawn on each plot. The solid line represents Y = X. The line with long dashes represents the mean population regression line, and the line with short dashes represents the mean of the training sample regression lines. The sample regression lines are computed from the training segments and usually have a noticeably higher slope.

The bias of an individual regression estimate is equal to the distance between the sample and population regression lines as measured along the line defined by the mean population classification value. Therefore, the bias of using the training sample is affected by the difference in slope of these two lines, the relationship of the population classification mean to the intersection of these lines, and the difference between the sample and population classification means.

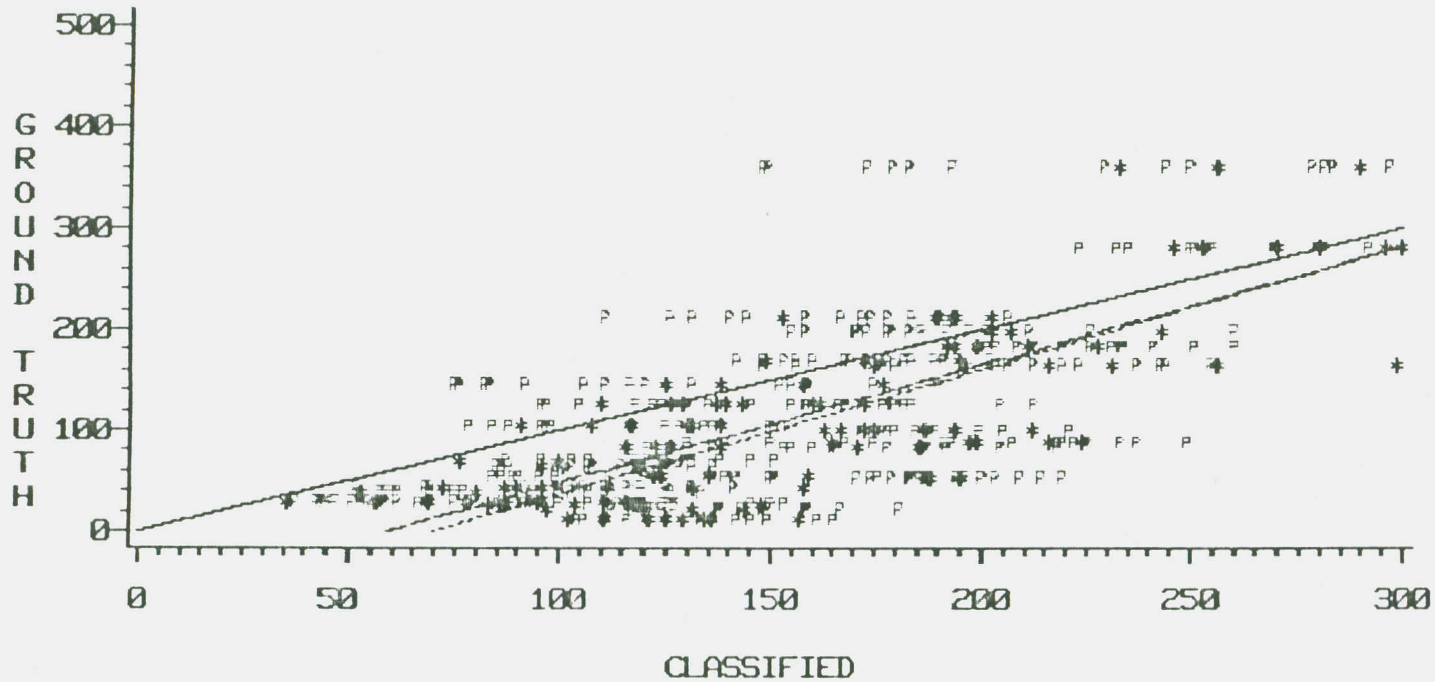
These plots are also useful in developing a clearer concept of the magnitude of the variability in the classification process. This variability appears to be quite high for both training sample segments (*) and the rest of the population (P).

FIGURE III. NUMBER OF GROUND TRUTH PIXELS VERSUS CLASSIFIED PIXELS PER SEGMENT
(SOYBEANS)



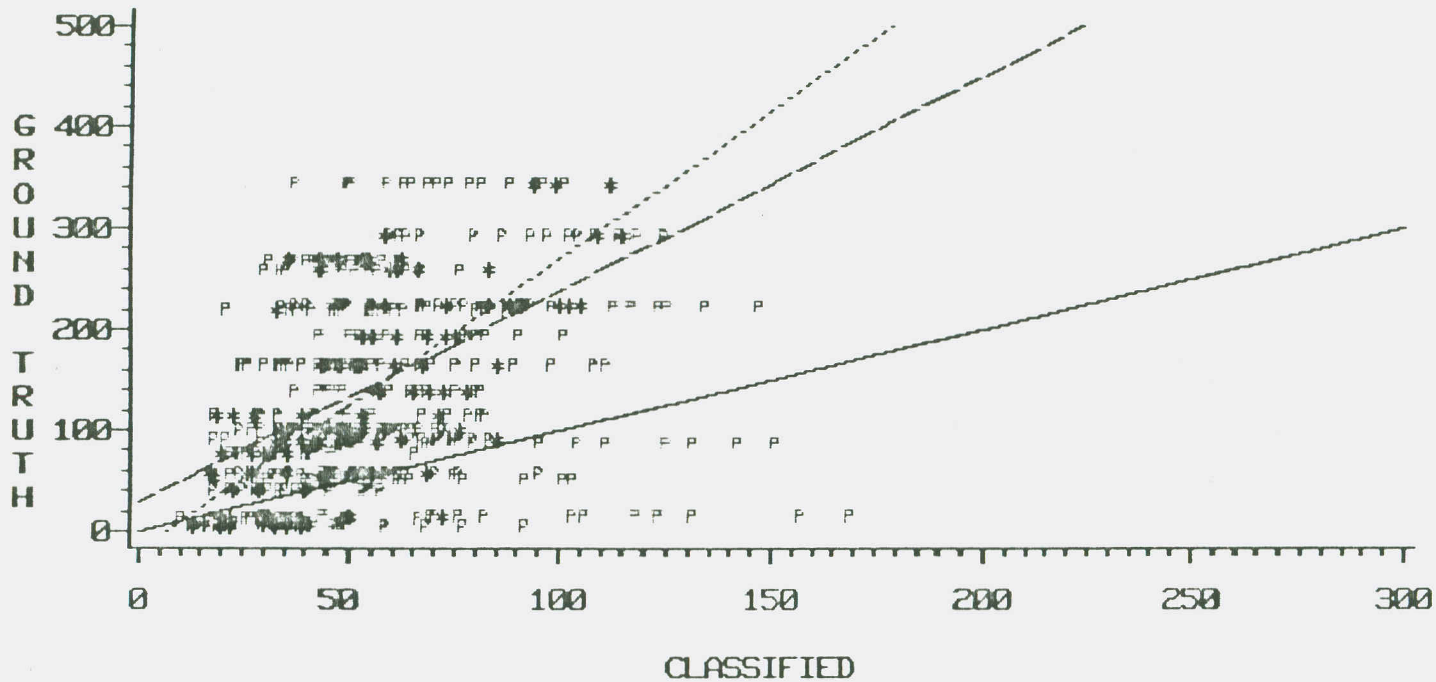
LEGEND *=TRAINING SEGMENTS P=REMAINING SEGMENTS

FIGURE IV. NUMBER OF GROUND TRUTH PIXELS VERSUS CLASSIFIED PIXELS PER SEGMENT (CORN)



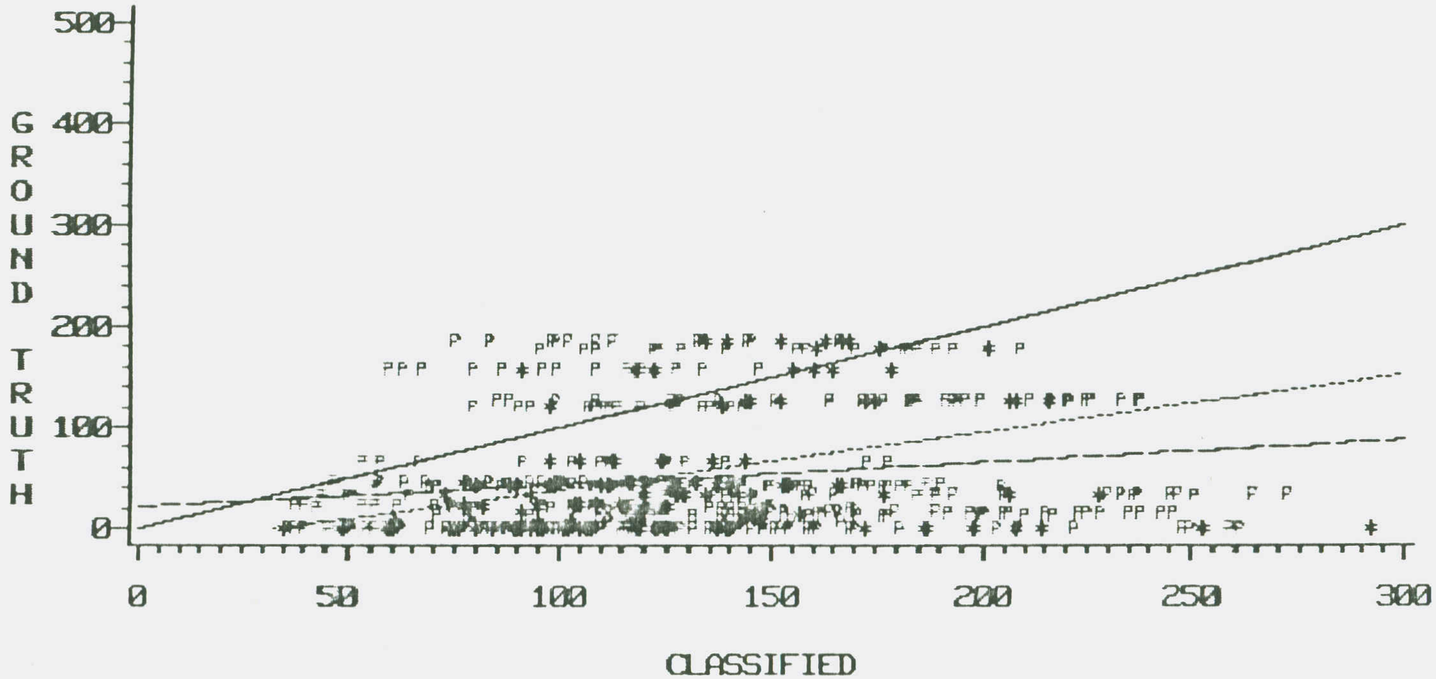
*LEGEND *=TRAINING SEGMENTS P=REMAINING SEGMENTS*

FIGURE V. NUMBER OF GROUND TRUTH PIXELS VERSUS CLASSIFIED PIXELS PER SEGMENT
(WASTE)



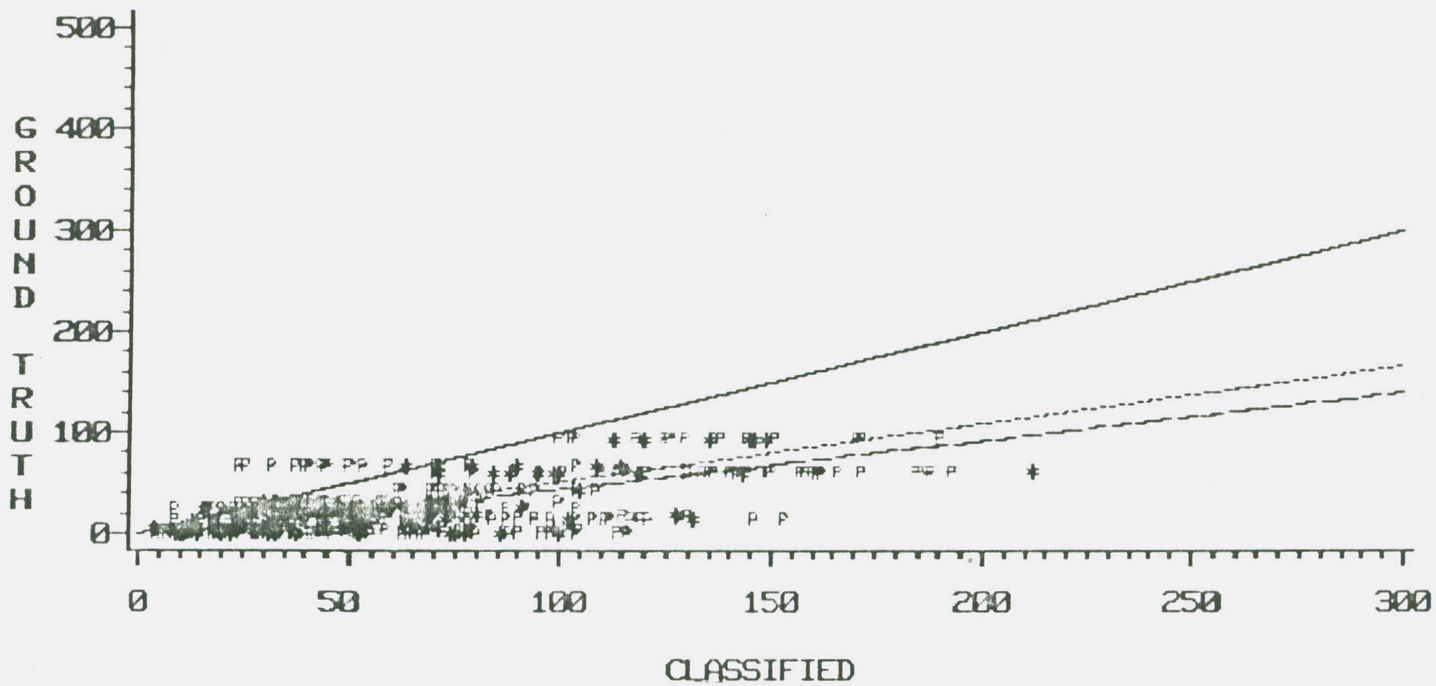
LEGEND *=TRAINING SEGMENTS P=REMAINING SEGMENTS

FIGURE VI. NUMBER OF GROUND TRUTH PIXELS VERSUS CLASSIFIED PIXELS PER SEGMENT
(HAY)



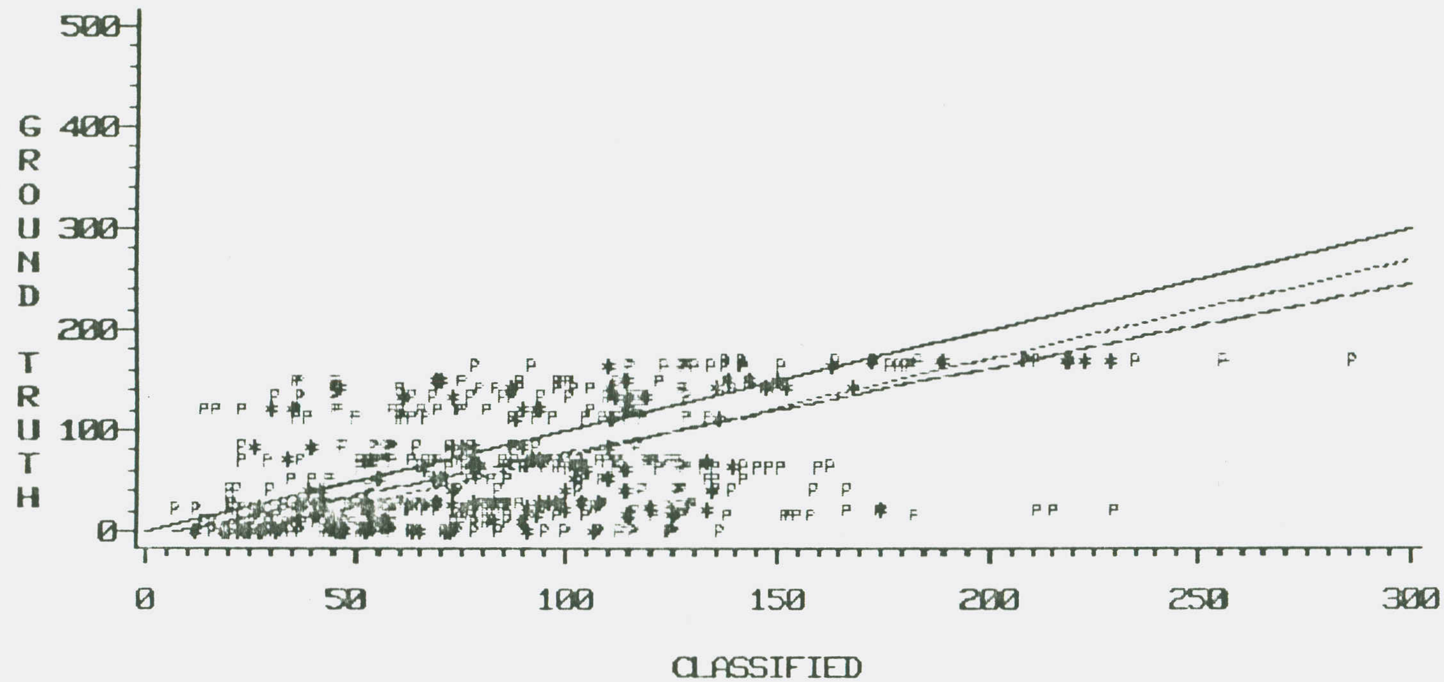
LEGEND *=TRAINING SEGMENTS P=REMAINING SEGMENTS

FIGURE VII. NUMBER OF GROUND TRUTH PIXELS VERSUS CLASSIFIED PIXELS PER SEGMENT
(WINTER WHEAT)



LEGEND *=TRAINING SEGMENTS P=REMAINING SEGMENTS

FIGURE VIII. NUMBER OF GROUND TRUTH PIXELS VERSUS CLASSIFIED PIXELS PER SEGMENT (*WOODS*)



*LEGEND *=TRAINING SEGMENTS P=REMAINING SEGMENTS*

VARIANCE

The simulation study provided an opportunity to evaluate different variance formulas for the regression estimator. Table X contains estimates of the true variance for the regression estimates of mean crop area per segment. These values were computed from the variability of the regression estimates among 20 replicates. The ratio of the mean Cochran (1977) variance to this true variance is provided in Table XI. Thus, if the numerator and denominator are independent and have identical distributions, the mean ratio could be significantly higher than one.[1] The median however would still be one.[1] A sign test was used to test the null hypothesis that the median ratio is equal to one. This null hypothesis was rejected for only two crops: waste and wheat.

The Cochran estimate of variance is based upon the variation of the observations around a fitted regression line. The replicate variance is based upon the variability of the regression estimate from one replicate to another. Therefore, if there is a variation in the regression line which cannot be accounted for by the simple randomness of the observations around that line, the Cochran formula would underestimate the variance.

Typically, the maximum likelihood procedure classifies a pixel into one of the classes whose means are closest to its spectral values. However, because of the relatively large total spectral variability of waste, values of the likelihood function are largest only in the tails of its distribution. This creates a tendency to collect the outliers from the other classes and also causes waste to have the lowest classification accuracy. It is sensitive to any shift in its own distribution as well as the distributions of other crops. Many shifts could cause large differences in the classifier relative to waste, leading to a corresponding large change in the regression line, and finally, a large variance of the regression estimates between replicates.

---

[1]Findings were supported by a computer simulation of 500,000 random ratios.

TABLE X

VARIANCE OF REGRESSION ESTIMATES ACROSS REPLICATES

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	3734.7	850.35	1233.91	3041.56	832.29	935.05	119.508
2	25	4	2121.7	1074.94	420.40	5846.97	3730.89	520.40	163.368
3	25	4	4697.4	1696.24	816.45	2085.53	1131.89	788.77	50.909
4	25	4	1955.6	1423.12	354.54	2908.25	976.43	1089.16	429.082
5	25	4	2636.0	1408.18	954.70	1936.40	531.92	563.80	122.032
6	25	4	2686.3	493.22	691.22	4042.38	523.36	2812.34	138.248
7	50	4	4000.0	3004.85	1093.50	4320.54	827.63	1546.35	236.944
8	50	4	4426.6	1585.30	495.35	2849.89	978.93	579.70	243.642
9	50	4	10670.5	1595.37	376.36	3981.20	1090.03	1963.69	379.874
10	50	4	5376.2	1197.66	627.95	3367.76	869.41	1842.90	128.635
11	50	4	2058.8	1212.29	800.63	3932.90	1956.30	846.29	204.175
12	50	4	3534.0	1476.74	587.71	3902.80	1358.10	710.02	189.954
13	100	4	5434.3	905.43	237.60	3923.39	1605.27	1844.48	372.562
14	100	4	1887.0	782.83	259.36	3237.72	1360.17	637.97	156.030
15	25	6	1448.2	513.68	475.74	893.64	586.13	272.71	51.869
16	25	6	1399.0	390.72	241.31	1299.55	212.31	765.39	93.067
17	25	6	1466.1	411.76	301.55	965.33	285.71	196.33	62.651
18	25	6	888.9	553.89	124.63	950.75	393.30	372.08	82.338
19	25	6	1781.1	1546.07	175.54	656.65	450.00	326.04	133.593
20	25	6	1834.7	694.70	442.74	1446.54	611.41	356.93	38.238
21	50	6	1361.3	526.59	579.75	539.87	879.02	599.10	103.813
2	50	6	4487.1	564.57	633.93	1815.59	334.17	788.92	197.450
23	50	6	1921.1	631.33	584.92	1585.15	912.05	1053.26	73.218
24	50	6	1704.6	396.13	297.91	1128.14	339.78	410.54	149.250
25	50	6	2272.0	822.31	285.42	1870.06	419.77	580.55	96.267
26	50	6	3910.4	712.23	336.27	1465.28	420.44	634.60	124.617
27	100	6	1432.0	491.80	286.70	1900.94	234.80	245.85	121.502
28	100	6	1048.7	595.30	464.30	1805.02	684.21	585.30	93.194
29	50	10	728.0	302.24	249.18	871.42	267.03	195.02	60.084
30	50	10	1467.0	222.31	191.25	469.69	627.17	125.42	27.394
31	50	10	1435.9	287.05	249.77	301.74	423.38	442.37	21.252
32	50	10	732.2	356.04	181.02	369.69	157.94	332.78	29.418
33	50	10	852.0	257.35	244.31	597.27	204.24	188.55	62.702
34	50	10	1401.8	486.15	180.97	734.58	206.58	310.28	63.170
35	100	10	1111.8	446.40	152.05	638.52	571.12	224.09	37.222
36	100	10	1399.3	354.37	330.58	969.80	296.42	268.14	50.610
37	100	15	855.2	415.96	199.60	104.76	186.33	165.94	40.410
38	100	15	1187.1	292.91	175.39	365.93	131.97	352.55	57.855

TABLE XI

RATIO OF COCHRAN VARIANCE TO REPLICATE VARIANCE

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	1.11593	0.81410	0.76541	0.45590	1.02787	0.96581	0.35976
2	25	4	1.17671	1.58575	1.34983	0.30240	0.17591	1.41620	0.77403
3	25	4	0.69880	1.01033	0.92243	0.55138	0.42216	1.19436	3.36248
4	25	4	1.71374	1.36953	1.89145	0.54071	0.93789	0.82203	0.77945
5	25	4	1.55445	1.43960	1.05309	1.00755	1.47366	1.29403	2.54413
6	25	4	0.96423	2.73436	1.61403	0.32700	2.08435	0.70220	0.59353
7	50	4	1.22660	1.20900	0.99751	0.21570	1.43359	0.95546	1.14703
8	50	4	0.68097	0.59151	1.24875	0.40815	0.54210	0.52547	0.31724
9	50	4	0.12900	0.84996	2.29871	0.32472	1.24229	1.55706	0.86828
10	50	4	0.77212	0.96868	1.08168	0.48944	1.28445	0.81498	0.59749
11	50	4	1.69996	0.87559	1.75036	0.37708	0.53421	1.02092	0.59723
12	50	4	0.88754	0.50436	1.20119	0.38092	0.84852	0.87521	1.28231
13	100	4	0.64259	0.94175	3.99819	0.50089	0.50735	0.70527	0.77906
14	100	4	1.45489	1.30137	2.28113	0.80123	0.62467	2.48460	1.34129
15	25	6	0.83871	1.29626	0.43008	0.84026	0.70395	0.43161	0.48062
16	25	6	0.53167	0.91728	1.17778	0.36272	1.09251	0.55134	2.76826
17	25	6	0.67893	1.30603	0.92961	0.72436	1.31920	1.12209	0.48122
18	25	6	1.43692	0.99702	1.89467	0.64021	1.07196	1.50799	1.23409
19	25	6	1.02667	0.32901	2.62305	0.91918	0.81863	0.99846	0.75273
20	25	6	1.08230	0.46301	1.04021	0.40109	0.72749	0.48264	0.47529
	50	6	1.01951	0.98590	0.73700	1.28404	0.55700	0.69718	0.60995
	50	6	0.39016	0.87924	0.43728	0.52895	1.32991	0.88338	0.59777
23	50	6	0.95594	0.82813	0.74925	0.42676	0.60715	0.58866	0.72112
24	50	6	0.79444	1.11459	0.73372	0.46362	1.49375	0.61044	0.79790
25	50	6	0.68560	0.70589	1.00469	0.24099	1.15578	1.23589	0.55292
26	50	6	0.47711	1.58481	1.28504	0.60481	1.16108	1.96440	1.43594
27	100	6	0.81033	0.89569	1.16258	0.57057	1.99968	0.67785	1.33456
28	100	6	1.41094	0.88022	0.75552	0.32090	0.41743	0.58387	1.51654
29	50	10	0.76599	0.99976	0.49932	0.35572	0.72559	1.19364	0.69850
30	50	10	0.54835	1.23083	1.00383	0.65170	0.29365	0.62053	0.96713
31	50	10	0.63263	1.01825	0.80784	0.98057	0.54726	0.55242	0.63338
32	50	10	1.09297	0.68694	0.72384	0.61841	1.49931	0.78106	0.97839
33	50	10	0.96996	1.08056	0.58538	0.58325	0.96046	1.38678	0.97119
34	50	10	0.63962	0.60666	0.89487	0.56332	1.06459	0.99223	0.80346
35	100	10	0.93330	0.96278	1.27434	0.70282	0.46392	1.10329	1.31507
36	100	10	0.65292	0.90593	0.54322	0.37194	0.74285	0.72417	0.78337
37	100	15	0.62110	0.42464	0.52512	2.29301	0.70784	0.63331	0.79904
38	100	15	0.54415	0.62705	0.88425	0.65600	1.11234	0.47871	0.57384
NO. RATIOS GT 1.0			13	14	20	3*	17	13	11*
MEDIAN			0.825	0.952	1.004	0.535	0.983	0.848	0.781

* THE MEDIAN RATIO IS SIGNIFICANTLY DIFFERENT FROM 1.0 AT THE FIVE PERCENT LEVEL.

Winter wheat is the other crop with a significantly different ratio of the Cochran to replicate variance. It has the fewest number of pixels of any crop on which a regression estimate was computed. Several training samples had very few or even zero pure wheat pixels available for training the classifier. One would expect the regression equation from these replicates to be quite different from the others and thus cause a large replicate variance.

Although the Cochran formula leads to biased estimates of variance for waste and wheat, the null hypothesis is accepted for the other crops which are not as susceptible to the types of problems discussed above. One could conclude the Cochran formula is a suitable estimate of variance when there are enough pixels to train the classifier and where the classifier is relatively robust.

Table XII displays a similar ratio of the large sample (L.S.) variance to the replicate variance. The null hypothesis used in testing the Cochran formula would be rejected for each of the crops of the large sample variance.

The Cochran formula for the variance of a regression estimate is equal to the large sample formula multiplied by the factor $(1 + 1/(n-3) + 2\gamma^2/n^2)$ where n is the sample size and γ is the measure of skewness. Since the value of skewness is usually relatively small for simulated data, the $1/(n-3)$ term is the main difference in the Cochran formula.

If one assumes that the skewness is zero and that the Cochran formula is an unbiased estimate of the true variance, then one can estimate what the true median of the large sample variance ratio should be. For example, in a sample size of four:

$$\begin{aligned} \mathrm{Var(Cochran)} &= \mathrm{Var(L.S.)}(1 + 1/(n-3)) \\ &= 2\ \mathrm{Var(L.S.)} \end{aligned}$$

TABLE XII

RATIO OF LARGE SAMPLE VARIANCE TO REPLICATE VARIANCE

	POP	SAMPLE	PASTURE	SOYBEANS	CORN	WASTE	WOODS	HAY	WHEAT
1	25	4	0.54125	0.39803	0.37756	0.22350	0.50347	0.47244	0.17749
2	25	4	0.58255	0.78162	0.66499	0.14746	0.08679	0.70121	0.37927
3	25	4	0.34382	0.49807	0.45245	0.27020	0.20829	0.58526	1.65625
4	25	4	0.84728	0.67041	0.92705	0.26504	0.45998	0.40179	0.38493
5	25	4	0.75829	0.70081	0.52117	0.49633	0.72365	0.63475	1.25130
6	25	4	0.46966	1.35369	0.78757	0.16004	1.02606	0.34347	0.28984
7	50	4	0.59985	0.59651	0.49163	0.10621	0.69996	0.34912	0.38195
8	50	4	0.33441	0.28925	0.61213	0.19907	0.26292	1.23244	0.66184
9	50	4	0.06375	0.41949	1.13777	0.15928	0.60943	0.21358	0.23724
10	50	4	0.37734	0.47569	0.53129	0.24023	0.62974	0.27281	1.35524
11	50	4	0.82902	0.43090	0.86133	0.18288	0.26062	0.55321	0.23715
12	50	4	0.43599	0.25002	0.58979	0.18732	0.41377	0.73676	0.60141
13	100	4	0.31639	0.46642	1.96806	0.24762	0.24933	0.27159	0.31006
14	100	4	0.72149	0.64468	1.11371	0.39462	0.31087	0.38222	0.48403
MEDIAN		4	0.505	0.487	0.639	0.211	0.437	0.437	0.383
15	25	6	0.62297	0.96484	0.31871	0.61598	0.52382	0.70897	0.84486
16	25	6	0.51999	0.90022	1.15655	0.35900	1.05603	0.51945	0.31065
17	25	6	0.50432	0.96977	0.67634	0.52840	0.97866	1.15794	0.64103
18	25	6	1.06589	0.74000	1.40283	0.46390	0.78868	0.60444	0.43737
19	25	6	0.75522	0.23852	1.95108	0.67690	0.59926	0.75656	0.44490
20	25	6	0.79772	0.34397	0.77159	0.29341	0.52796	0.64519	0.94756
21	50	6	0.75145	0.72224	0.54635	0.94919	0.41356	0.73423	0.55312
22	50	6	0.28990	0.64893	0.32281	0.38250	0.98226	0.35729	0.34758
23	50	6	0.70236	0.61055	0.55189	0.31507	0.44907	0.51700	0.45158
24	50	6	0.58219	0.83050	0.54173	0.34184	1.09570	0.65249	0.44152
25	50	6	0.50834	0.51543	0.74305	0.17469	0.84765	0.43809	0.53160
26	50	6	0.35237	1.17746	0.95335	0.44462	0.86354	0.45154	0.58905
27	100	6	0.59950	0.66535	0.85839	0.41986	1.47304	1.01219	0.72174
28	100	6	1.35897	0.86946	0.73491	0.31151	0.40650	0.97918	0.79136
MEDIAN		6	0.611	0.731	0.739	0.401	0.818	0.649	0.542
29	50	10	0.76295	0.99665	0.48980	0.35148	0.72253	1.22763	0.54570
30	50	10	0.47099	1.06991	0.86956	0.55993	0.25424	1.71321	1.25045
31	50	10	0.54298	0.88510	0.69551	0.84404	0.47445	0.58945	1.14450
32	50	10	0.93600	0.59554	0.62913	0.53749	1.29365	0.50551	1.30784
33	50	10	0.84210	0.93792	0.50543	0.49603	0.83232	1.04005	0.60476
34	50	10	0.55757	0.52858	0.77668	0.49041	0.92809	0.54143	0.83915
35	100	10	0.80868	0.83403	1.10598	0.60985	0.40271	0.96165	1.14331
36	100	10	0.64220	0.90033	0.53888	0.36560	0.73503	0.71542	0.77764
MEDIAN		10	0.703	0.893	0.662	0.517	0.729	0.839	0.991
37	100	15	0.61701	0.42334	0.52194	2.25214	0.70547	0.63012	0.79620
38	100	15	0.49729	0.57418	0.81478	0.59594	1.02290	0.44028	0.52419

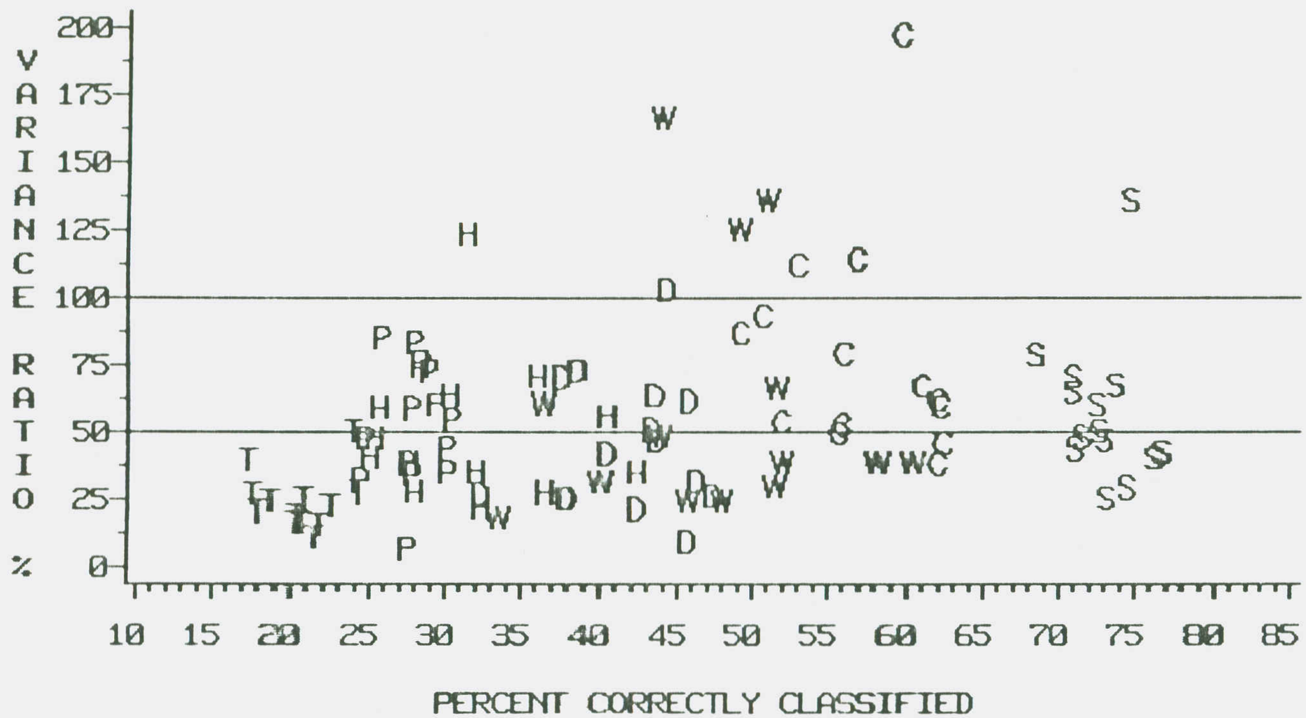
Thus, the median large sample variance ratio should be 0.5. Since there are too few estimates in each sample size to draw a meaningful conclusion in a test of hypothesis, the values of the large sample variance ratio for each sample size are simply displayed in Figures IX - XI. Contained in these plots are the crop labels each point is associated with, reference lines drawn at Y = 1, and other lines depicting the true median (given that the Cochran formula is correct: Y = .5 for sample size = 4). The plots also show the percent of the pixels which were correctly classified for each crop and analysis. Casual observation of the data suggests that the variance estimate is not substantially affected by the classification accuracy.

Similar plots were also made for the Cochran variance ratio (Figures XII-XIV). Casual observation of these plots suggests the suitability of the Cochran formula for variance estimation.

Relative Efficiency

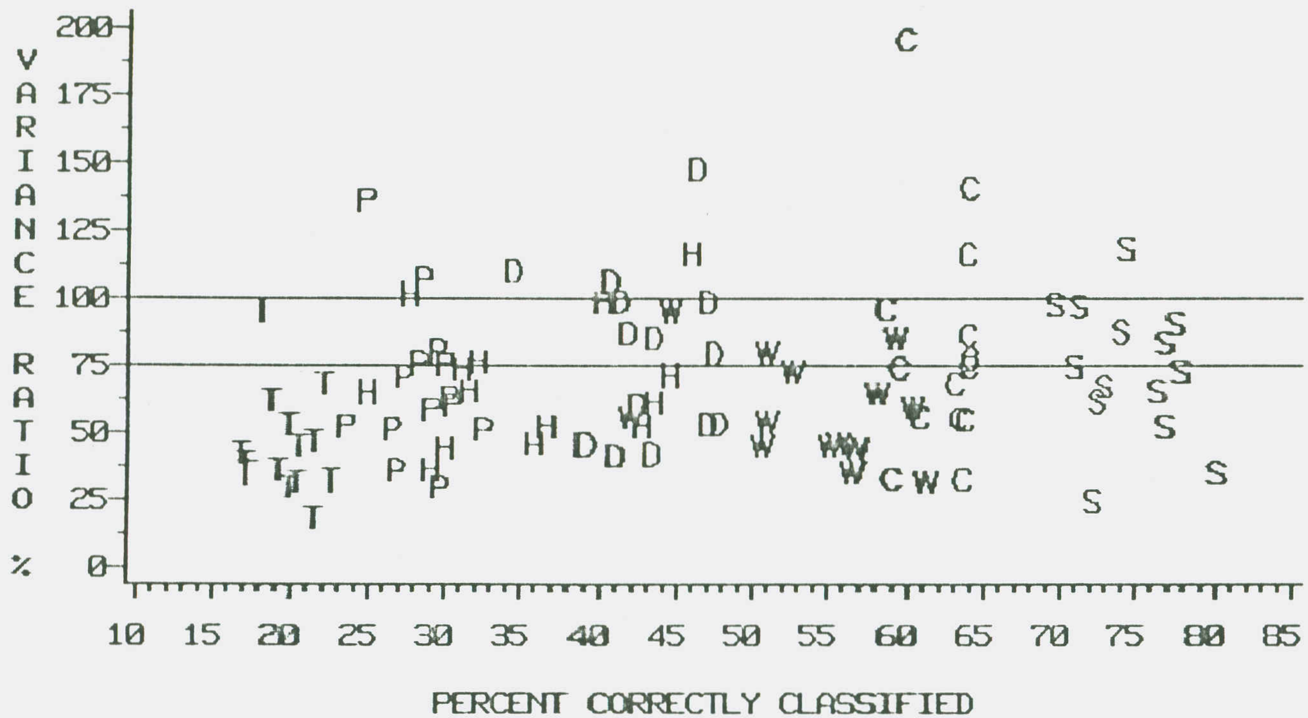
A question which often arrises in using Landsat data for crop estimation is, "How does one know if the data will really improve the efficiency of the estimate?" One approach to answering this question is to relate the relative efficiency of the regression estimate to some measure of separability. If the expected relative efficiency could be predicted from some estimate of separability in the training sample, an early decision could be made as to whether or not to classify the entire region. It is assumed that better separability would lead to more accurate classification, more information upon which to base the estimate, and thus more efficient estimates. Depicted in Figure XV is the relationship between the relative efficiency of the regression estimates and the percent of the pixels correctly classified. Similarly, the relationship between relative efficiency and a more direct measure of separability between classes is characterized in Figure XVI. This measure of separability is the harmonic mean of the generalized squared distances from a given crop to each of the other crops. The generalized squared distances are the pairwise squared generalized distances between groups, as calculated in the SAS (Statistical Analysis System) discriminant analysis procedure (1982). It

FIGURE IX. RATIO OF LARGE SAMPLE VARIANCE TO TRUE VARIANCE VERSUS PERCENT CORRECTLY CLASSIFIED (SAMPLE SIZE EQUALS FOUR)



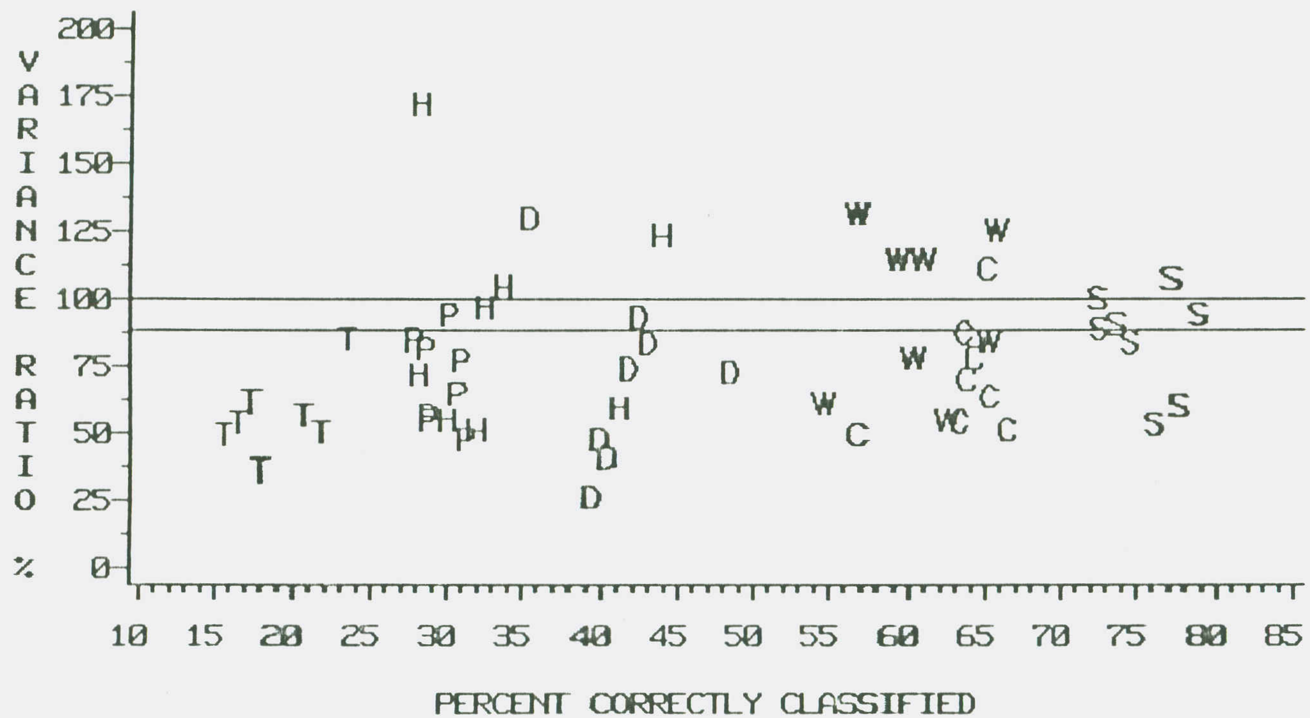
LEGEND *P=PASTURE S=SOYBEANS C=CORN T=WASTE D=WOODS H=HAY W=WINTER WHEAT A=ALFALFA*

FIGURE X. RATIO OF LARGE SAMPLE VARIANCE TO TRUE VARIANCE VERSUS PERCENT CORRECTLY CLASSIFIED (SAMPLE SIZE EQUALS SIX)



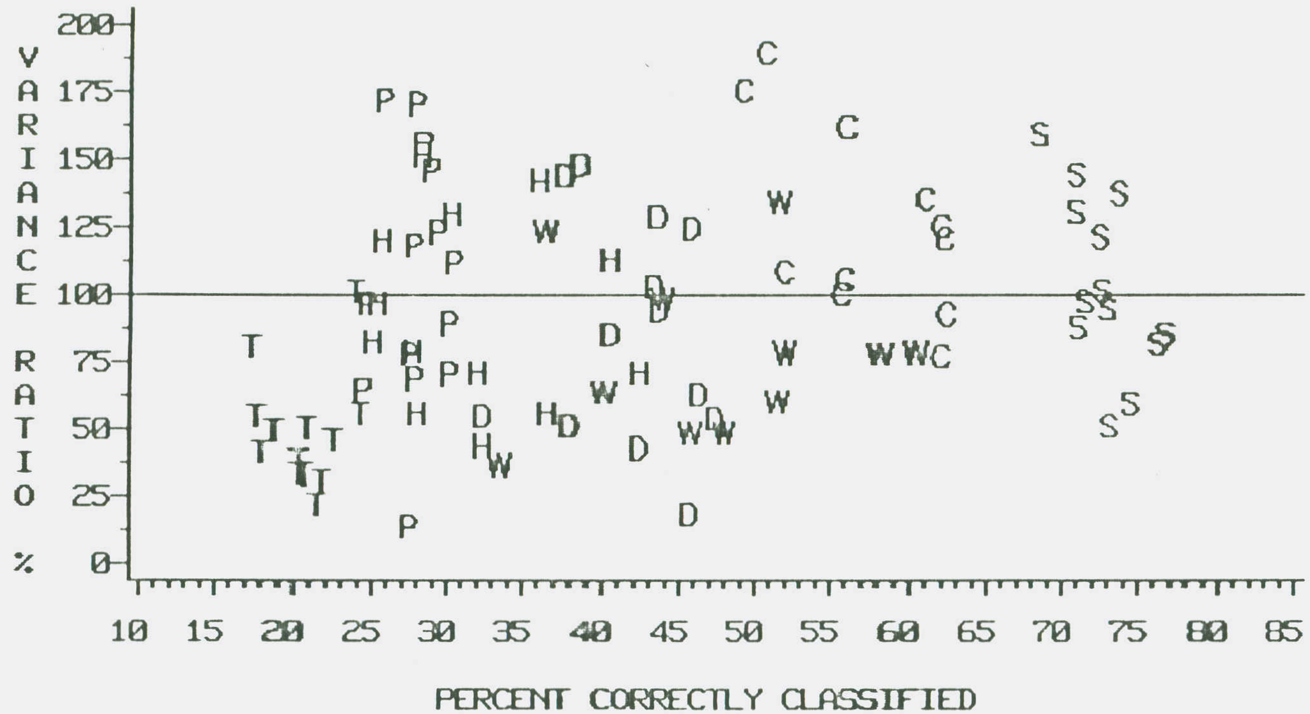
LEGEND *P=PASTURE S=SOYBEANS C=CORN T=WASTE*
D=WOODS H=HAY W=WINTER WHEAT A=ALFALFA

FIGURE XI. RATIO OF LARGE SAMPLE VARIANCE TO TRUE VARIANCE VERSUS PERCENT CORRECTLY CLASSIFIED (SAMPLE SIZE EQUALS TEN)



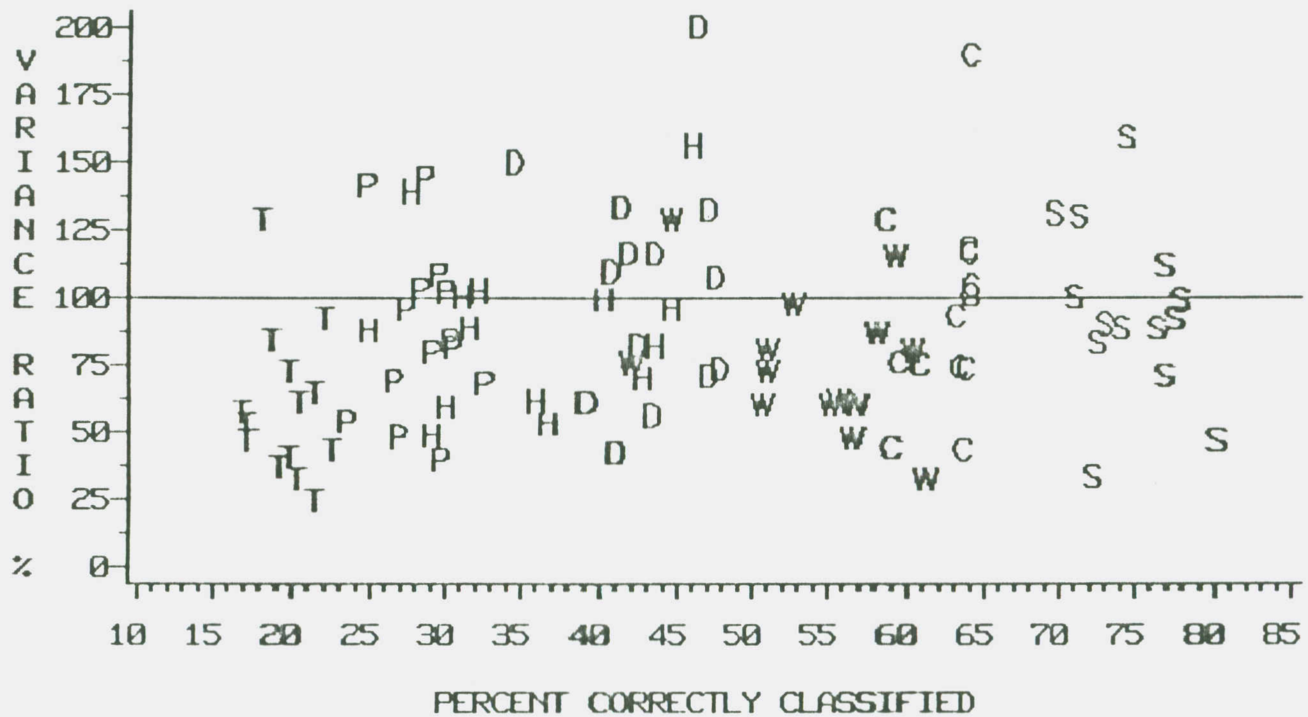
LEGEND P=PASTURE S=SOYBEANS C=CORN T=WASTE D=WOODS H=HAY W=WINTER WHEAT A=ALFALFA

FIGURE XII. RATIO OF COCHRAN VARIANCE TO TRUE VARIANCE VERSUS PERCENT CORRECTLY CLASSIFIED (SAMPLE SIZE EQUALS FOUR)



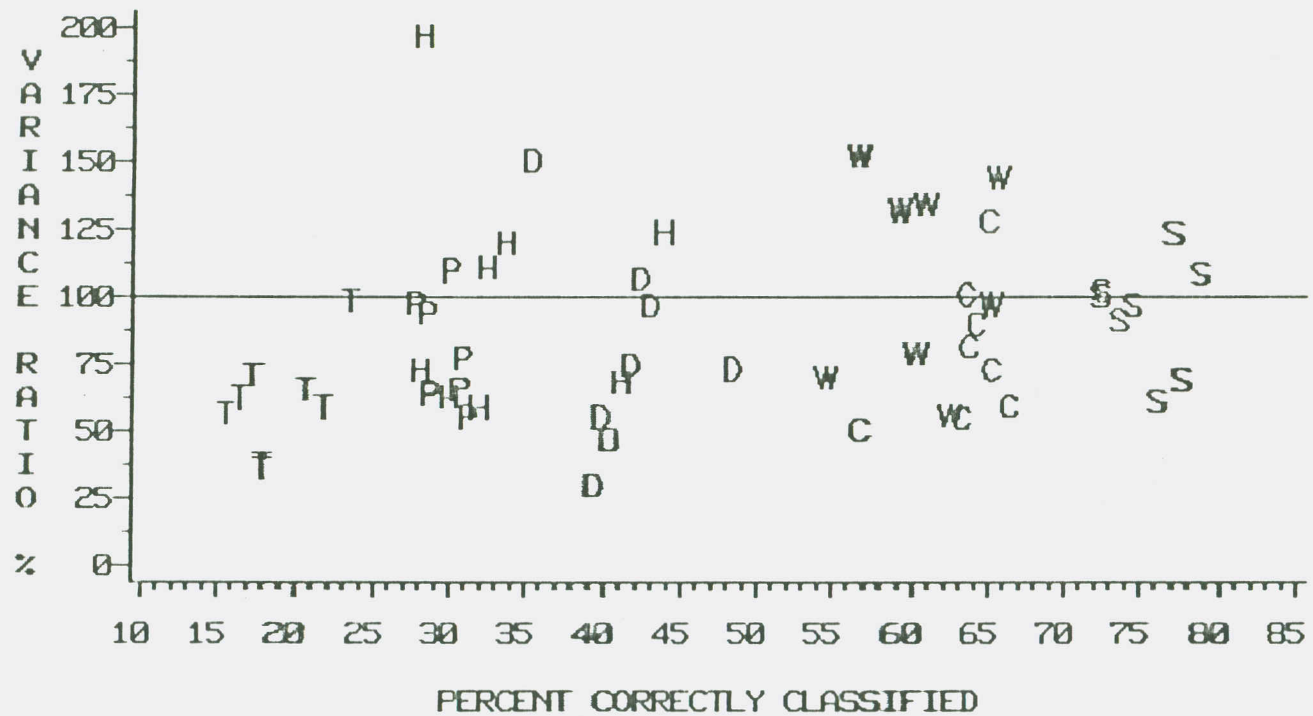
LEGEND *P=PASTURE S=SOYBEANS C=CORN T=WASTE D=WOODS H=HAY W=WINTER WHEAT A=ALFALFA*

FIGURE XIII. RATIO OF COCHRAN VARIANCE TO TRUE VARIANCE VERSUS PERCENT CORRECTLY CLASSIFIED (SAMPLE SIZE EQUALS SIX)



LEGEND *P=PASTURE S=SOYBEANS C=CORN T=WASTE D=WOODS H=HAY W=WINTER WHEAT A=ALFALFA*

FIGURE XIV. RATIO OF COCHRAN VARIANCE TO TRUE VARIANCE VERSUS PERCENT CORRECTLY CLASSIFIED (SAMPLE SIZE EQUALS TEN)



LEGEND *P=PASTURE S=SOYBEANS C=CORN T=WASTE*
D=WOODS H=HAY W=WINTER WHEAT A=ALFALFA

FIGURE XV. RELATIVE EFFICIENCY VERSUS PERCENT CORRECTLY CLASSIFIED

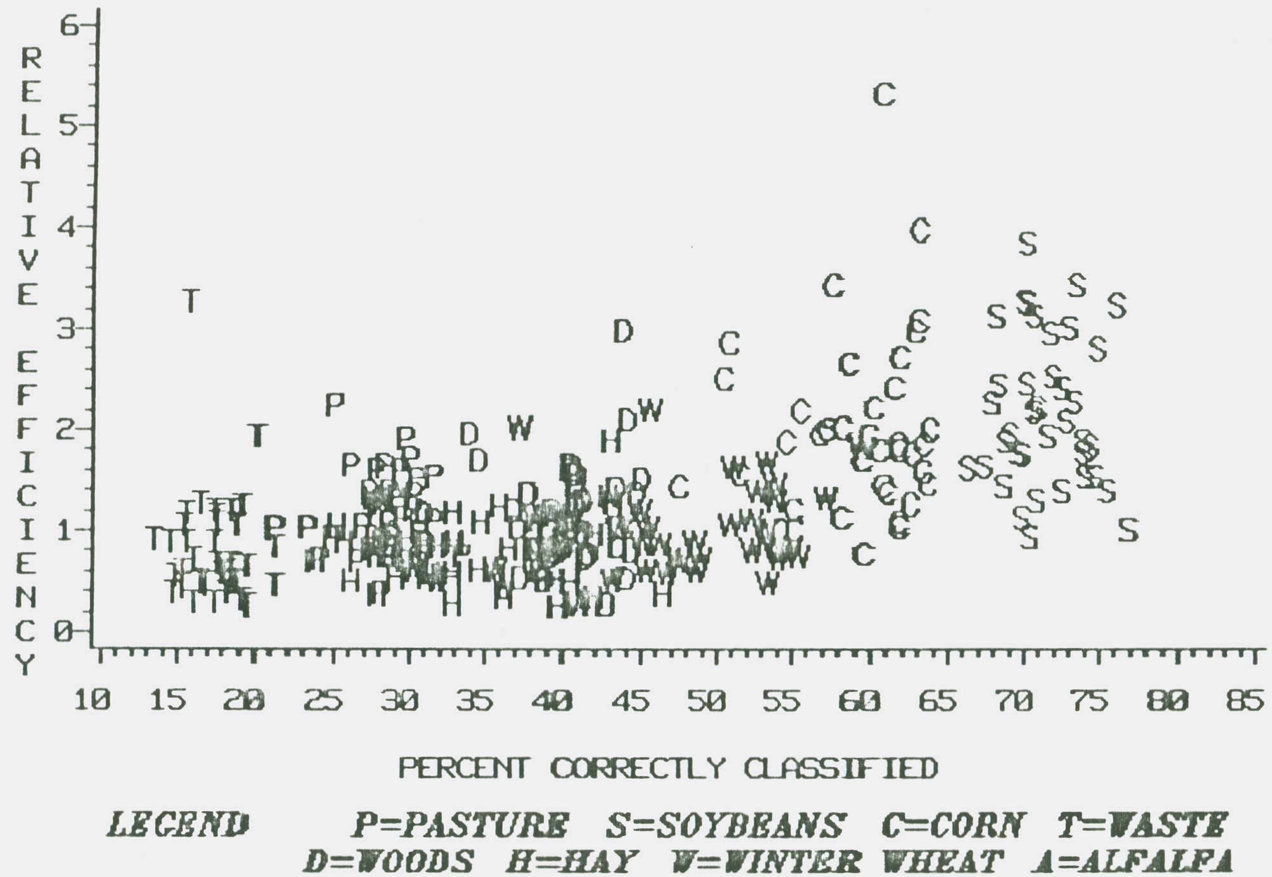
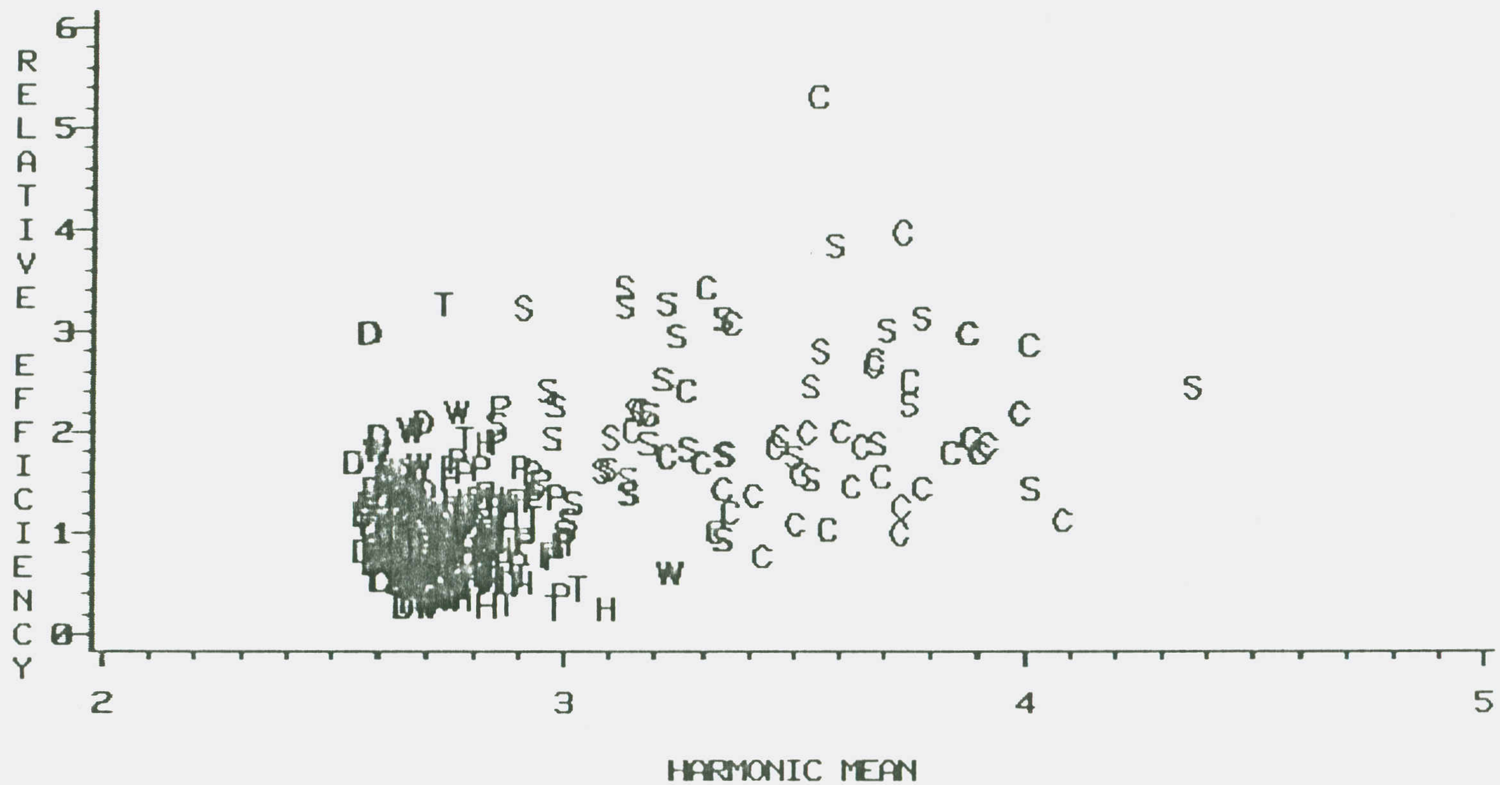


FIGURE XVI. RELATIVE EFFICIENCY VERSUS HARMONIC MEAN OF GENERALIZED SQUARED DISTANCES



LEGEND P=PASTURE S=SOYBEANS C=CORN T=WASTE D=WOODS H=HAY W=WINTER WHEAT A=ALFALFA is expected that classification error rates are greatest between groups with the smallest distances. Thus, the harmonic mean was used in the analysis because it assigns larger weights to smaller values.

Although, there is a trend for the relative efficiency to be larger for higher values of the percent correctly classified and of the harmonic mean, there does not appear to be a point above which one could safely predict high relative efficiency. On the other hand, it appears that if the percent correctly classified is less than 50 or if the harmonic mean distance is less than 3, the use of Landsat imagery will not improve the efficiency of the estimate. It should be emphasized that these plots may portray a substantially different relationship if more segments were used or if this analysis were applied to real rather than simulated data.

REFERENCES

Chhikara, R.S. and Houston, A.G., "Estimation with Classifier as Auxiliary Variable." Presented at the American Statistical Association Meeting in Philadelphia, August 19, 1984.

Cochran, W.G., Sampling Techniques, John Wiley, New York, 1977.

Lundgren, J.C., "Landsat Data Based Regression Estimates in Crop Surveys Simulation Study." Technical report prepared for the USDA by Lockheed-EMSCO, Houston, Texas, September 27, 1984.

SAS Institute Inc. SAS Users Guide: Statistics, 1982 Edition. Cary, N.C.: SAS Institute Inc., 1982