1-1-1976

Evaluation of Classification Procedures for Estimating Wheat Acreage in Kansas

L. M. Flores

D. T. Register

Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 29 - July 1, 1976

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
76CH1103-1 MPRSD

EVALUATION OF CLASSIFICATION

PROCEDURES FOR ESTIMATING WHEAT ACREAGE IN KANSAS*

L. M. Flores and D. T. Register

Lockheed Electronics Company, Inc./Aerospace Systems Division, Houston, Texas

I. ABSTRACT

This report presents the results of experiments which were performed to evaluate procedures for estimating wheat acreage in intensive test sites (ITS's) in Kansas. An analyst/interpreter (AI) selected and labeled fields from Landsat-1 satellite imagery. Statistics were generated for each selected ITS, and the imagery was classified using a maximum likelihood classifier. Various components of the classification process were tested.

II. INTRODUCTION

This experiment was designed to provide some insight into the factors affecting the classification and mensuration processes used for crop identification. Specifically, the effects of the following factors on the acreage estimate and/or the probability of correct classification (PCC) were investigated.

a. The number of training fields selected by the AI for the various classes

b. AI labeling errors and possible bias in the AI's selection of training fields

c. The number of training fields selected (AI or random selection)

d. The method of feature selection (no feature selection or one in which the without-replacement method of feature selection was used)

The experiment encompassed the analysis of five Kansas ITS's where ground truth was available. Three of the five ITS's (in Ellis, Saline, and Rice Counties) had 4.8- by 4.8-kilometer (3- by 3-mile, 9-square-mile, or ≅23-square-kilometer) ground-truth areas; the other two (Morton and Finney Counties) had 8- by 9.6-kilometer (5- by 6-mile, 30-square-mile, or ≅77-square-kilometer) ground-truth areas. Each ITS was interpreted by five AI's working independently of one another, thus providing each site with five sets of training fields. The data contained 16 channels obtained by registering four passes of the Landsat-1 which corresponded to the four biological windows; i.e., emergence, dormancy, maximum growth, and harvest.

The overall experiment was divided into three designs. The purpose of the first design was to study the effects of sites, AI's, and feature selection on classification. The variability in proportion estimation and fraction of training fields labeled wheat were also studied. The second design was to measure the extent proportion estimation was degraded by the mislabeling of training fields. The intent of the third design was to compare the AI's method of selecting training fields with that of a stratified random selection. The scope for each is as follows.

1. The first design called for selecting two 4.8- by 4.8-kilometer ground-truth areas from each of the 8- by 9.6-kilometer ground-truth ITS's in Finney and Morton Counties. Classification was performed on each 4.8- by 4.8-kilometer area separately using the sets of training fields provided by the five AI's. Thus, two estimates of proportion accuracy were available for the Finney and Morton County ITS's. Classification was also performed on the Rice, Ellis, and Saline County ITS's.

2. Only the Morton and Finney County ITS's were used in design 2. This called

*The material for this paper was developed under Contract NAS 9-12200 for the Earth Observations Division, Science and Applications Directorate, Lyndon B. Johnson Space Center, National Aeronautics and Space Administration, Houston, Texas.

for classifying each of the 8- by 9.6-kilometer ground-truth areas using the five sets of training fields provided for each site by the AI's and then correcting the training field sets for labeling and boundary errors and reclassifying the ITS's.

3. The third design was used to compare the classification results from the corrected fields of design 2 with the results of five classifications for each ITS using stratified randomly selected sets of training fields as described in the following paragraph.

For the third design, five sets of training fields were selected from each of the 8- by 9.6-kilometer ground-truth areas of the Morton and Finney County ITS's. An unbiased selection procedure was used, which is defined here to be one in which training fields are selected to represent each major subclass. To place all possible training fields into subclasses, the wheatfields in the ground-truth area were clustered with three as the maximum number of clusters. Using the majority rule (based on the number of pixels assigned to each cluster for that field), each wheatfield was labeled as belonging to one of the three subclasses. All the fields belonging to the second major crop (corn for Finney and fallow for Morton County) were clustered with two as the maximum number of clusters. Again, by majority rule, these fields were divided into two subclasses. All the other crops were represented by one subclass. Four wheatfields were selected randomly from the ground-truth wheatfields of each subclass of wheat. Three fields were randomly selected from the ground-truth fields of each subclass of the second major crop. Three fields (whenever available) were selected randomly to represent each of the other crops.

III. AI, SITE, AND FEATURE SELECTION EFFECTS

The results presented in this section which pertain to design 1 consist of estimated wheat proportions for 4.8- by 4.8-kilometer areas computed using AI training fields and iterative clustering with no feature selection configuration. In this configuration, subclassing of training fields was accomplished by clustering the training fields of wheat and nonwheat separately using an iterative clustering algorithm. Classification was performed using the resulting cluster statistics and the maximum likelihood classifier. Typically, there were between 7 and 9 clusters of wheat and between 14 and 18 clusters of nonwheat. The estimated proportions of wheat and nonwheat were then computed by pixel counting.

The results were analyzed by subjecting them to various statistical tests in an effort to gain some insight into the nature of the different factors that might affect the accuracy of the estimate of the proportion of wheat in a given area. The *analysis of variance* (ANOVA) design was used in analyzing the results whenever applicable. However, the results themselves suggested other tests which proved to be of equal or greater interest.

Tables 1 through 4 show the results of the classifications for the proportion error (PE), $\hat{p} - P_T$, and the PCC when using one configuration. For the no-threshold case, there were sites in which the PE's were, on the average, significantly overestimated and other sites for which they were significantly underestimated. Similarly, the training field sets selected by some AI's produced, on the average, significant overestimates while other sets produced significant underestimates. Notably, the mean of all proportion estimates over all sites and AI training field sets were unbiased. Also, the relative performance of the AI's changed from site to site.

Table 2 presents the results of applying a 1-percent threshold. As mentioned previously, thresholded pixels were considered to be nonwheat; thus, the observed negative overall bias was to be expected. However, the size of the bias (15 percentage points) was larger than expected, and this was attributed to the large number of pixels being thresholded. The explanation for the large amounts of thresholding is either the incompleteness of the training set when three or more passes were used or the use of the wrong statistic (χ^2 instead of F) for threshold decision making, or both.

Also, using a 1-percent threshold, the site and the AI effects become less pronounced and the interaction between them is no longer significant.

Tables 3 and 4 summarize the results when the PCC was used as the response variable. In table 3, the PCC for the no-threshold case showed no significant difference from site to site or from AI to AI. This contrasted with the results obtained when the PE was used as a response variable in the ANOVA. In the latter case, a site effect was observed with a 99.5-percent confidence level and an AI effect with a 97.5-percent confidence level. It should be pointed out, however, that a site or an AI effect would be observed if a consistent overestimate or underestimate of the wheat proportion occurred for certain sites or AI's. One site or AI could have a consistent overestimate and another site or AI a consistent underestimate, and each still could have the same level of PCC. This would produce a site effect in the ANOVA for PE and not in the ANOVA for the PCC.

Table 4, with 1-percent thresholding, shows that some sites had a significantly higher PCC than others. There were no significant differences among AI's and no interaction.

Tables 5 through 8 show the results of running a second configuration; namely, the "iterative clustering without replacement feature selection" configuration. Two values of the ratio parameter of the without-replacement feature-selection program were used, 0.2 and 0.3. A comparison of the tables shows that little, if any, information was lost. The results of the ANOVA are shown in tables 9 and 10. Since thresholding is greatly dependent on the number of channels, it was expected that the feature-selection effect would be statistically significant. Indeed, when thresholding was applied, it was found to be significant at the 95-percent confidence level.

IV. VARIABILITY IN THE PROPORTION ESTIMATION

It was expected that if the PCC was low the PE would exhibit large variance; whereas, if the PCC was high, a small variance for PE would be expected. To determine if this was actually the case the PCC was divided into low, medium, and high categories and the variance for the PE computed for each. The results presented below are for the all-channel/no-threshold case.

PCC	Variance PE
0.450 – 0.599	0.06424
0.600 – 0.749	0.02139
0.750 – 1.000	0.01192

The decrease in PE variance with PCC can be seen in figure 1, which depicts a plot of PE versus PCC for the all-channel/no-threshold case.

Figure 2 presents all 35 classification results and demonstrates that there was very little correlation between the wheat proportion estimates and the ground-truth wheat proportion. This implied that other factors influenced the proportion estimation process more than the actual amount of wheat in the segment. The size of training data may be one of these factors. Omitting all classifications with less than 25 training fields reduced the mean square error 43 percent.

When the requirement was added that the fraction of the training fields labeled wheat be within 0.1 of the ground-truth wheat proportion, the mean square error was 52 percent lower than that for all classifications. For those classifications in which at least 25 fields were used for training and at least 10 of these were wheatfields, the mean square error was reduced by a notable 66 percent from that of all classifications. Thus, about two-thirds of the variability of the PE's was reduced by applying these minimum requirements. This suggests that the variability of PE's can be reduced significantly by applying some simple minimum requirements when the photointerpretation of a segment is performed.

It should be noted that the above-mentioned criteria provided an improvement in the proportion estimation accuracy for the five Kansas ITS's. It is expected that the criteria would be different for other sites under different conditions.

V. EFFECT OF TRAINING FIELD LABELING AND BOUNDARY ERRORS

Tables 11 and 12 show the results obtained from design 2 by correcting AI training field labels and boundaries for the no-threshold and the 1-percent-threshold cases, respectively. It can be seen that, with or without thresholding, there was no significant improvement in proportion estimation when the labels and boundaries were corrected with ground-truth information.

To study improvement a new variable was defined:

$$I_i = \left|\hat{p}_{iu} - p_i\right| - \left|\hat{p}_{ic} - p_i\right|$$

where

I_i = improvement for the *ith* site

$\hat{p}_{iu}$ = estimated proportion for the *ith* site using uncorrected training fields

p_{ic} = estimated proportion for the *ith* site using corrected training fields

p_i = ground-truth proportion for the *ith* site

The improvement is positive if the PE from an uncorrected training sample is larger than the PE from a corrected training sample and is negative for the reverse situation. A t-test was made for the no-threshold and the 1-percent-threshold cases to determine if the average improvement was different from zero.

To determine if an inadequate amount of training data was responsible for the large variability of the PE's, the two criteria which were arbitrarily suggested in section IV were used. The analysis of classifications which met either criterion showed an improvement of 4.6 percentage points for the no-threshold case (table 13). Because of the much lower variability when either criterion was met, this improvement was statistically significant. With a 1-percent threshold, correcting the labels and boundaries seemed to impair the results, although this was not significant (table 14).

VI. EFFECT OF AI BIAS IN TRAINING FIELD SELECTION

Table 15 presents the PE's for the no-threshold case of design 3. The PE's from classifications using AI-selected training fields (corrected for labeling and boundary mistakes) show a much larger variability than those using a stratified random selection of training fields. The PE's for both types of training field selection were positively biased. This was found to be significant for the randomly selected fields but not for the AI-selected fields. A cause-and-effect relationship is not ensured because an effect is declared statistically significant. Factors or variables which are highly correlated with the effect could produce the significance. In this experiment, a low number of training fields was typical for the AI selections. Therefore, an inadequate amount of training data was of interest as a possible significant factor in determining PE's. These concepts were studied, and the results are shown in figure 3.

Figure 3 presents the relationship of PE (with no threshold) to the total number of training fields. The results indicated that the primary source of variability came from classifications using less than 25 training fields.

With a 1-percent threshold (table 16), the variability of the PE's was significantly larger with the AI-selected training fields than with the randomly selected fields. As expected, the PE's were negatively biased for both types of selection.

VII. FRACTION OF TRAINING FIELDS LABELED AS WHEAT

From the nature of the classification procedures, it was expected that with a sufficiently large training sample (i.e., one in which all classes and subclasses were well represented) the wheat proportion estimate would depend only on class and subclass signatures. Specifically, the proportion estimate would be independent of the relative number of training fields representing each class or subclass. If a dependence were found, then it could be concluded that either the size of the training sample was not sufficiently large or the AI intentionally or unintentionally selected his training fields for each class in proportion to the relative abundance of that class.

Additionally, in the case of insufficient training samples, the dependence of the proportion estimate on the fraction of training fields representing a class or subclass would be expected to increase with the number of channels used for classification.

Two heuristic supporting arguments are presented. First, as the number of channels increases, a greater increase occurs in the number of parameters to be estimated. Statistically, for a given training sample size, the more parameters to be estimated, the less precise their estimation becomes. Second, for a given pass, suppose n subclasses are present. When an additional pass is added, each of the original n subclasses either remains the same or is split into one or more new subclasses. Therefore, the number of subclasses for both passes must be greater than or equal to the number of subclasses for the first pass. The greater the number of subclasses, the larger the number of training fields that are required to represent those subclasses.

To determine if the estimated wheat proportion depended upon the fraction of training fields labeled wheat, these two quantities were plotted against each other. Classification results from three separate sources were used to generate these plots. A statistical analysis was performed on each data set using simple linear regression. For the sake of brevity the plots are not presented, but the correlation coefficients are in table 17.

It can be seen that, in all of three separate experiments, the estimated wheat proportion is definitely correlated with the fraction of training fields labeled as wheat.

The correlation coefficients were compared to determine if the correlation becomes significantly larger as the number of channels increases. Only two comparisons could be made; namely, classifications 6 and 7 and classifications 3 and 5 of table 17. The first comparison showed that with the multitemporal classifications the correlation coefficient was significantly higher at the 90-percent confidence level. The second showed that the multitemporal correlation coefficient was higher than the single-phase correlation coefficient at the 99-percent confidence level, which is very significant.

Thus, it has been demonstrated in two separate experiments that the correlation coefficient increased with the number of channels used for classification.

The fact that a significant correlation was found indicates that (1) generally an inadequate training sample was available and/or (2) the AI's selected training fields for a class in proportion to the relative abundance of that class.

The fact that the correlation increased in accordance with the number of channels and that recent unpublished research confirms that AI's did not select numbers of training fields for each class on the basis of class abundance supports the contention that inadequate training samples were used in the experiments.

VIII. SUMMARY

The more salient findings during the course of this investigation are:

a. The variability of the proportion estimation from AI-selected training fields was so large that it masked both the labeling error and the correlation with the ground-truth wheat proportion.

b. Evidence showed that the variability in the proportion estimation could be reduced substantially by increasing the training data.

c. A threshold of 1 percent produced a negative bias in proportion estimation.

d. Indications were that classifications using randomly selected fields were not substantially better than those from the AI-selected and corrected training fields, using comparable training data. This suggests that a so-called "AI bias" in field selection may have a much smaller effect than was previously anticipated.

e. It was demonstrated that the estimated proportion was definitely correlated with the fraction of training fields labeled as wheat. This correlation became larger as the number of channels increased.

Table 1. Estimated Wheat Proportion Minus Ground-Truth Wheat Proportion With No Thresholding (Design 1)

Site \ AI	I	II	III	IV	V	Mean
I Morton	0.399 .187	0.168 .015	-0.183 -.109	-0.037 -.159	0.018 -.166	0.0133
II Finney	.329 .401	.009 .164	.224 .214	.076 .084	.067 .160	.1728
III Ellis	-.031	.068	-.301	-.138	.035	-.0734
IV Saline	-.211	-.116	-.063	-.080	.051	-.0838
V Rice	.117	-.184	-.177	-.017	.172	-.0178
Mean	.1701	.0177	-.0564	-.0387	.0481	.0282

Table 2. Estimated Wheat Proportion Minus Ground-Truth Wheat Proportion With 1-Percent Thresholding (Design 1)

Site \ AI	I	II	III	IV	V	Mean
I Morton	0.002 -.192	-0.018 -.229	-0.304 -.332	-0.242 -.414	-0.153 -.282	-0.2164
II Finney	.083 .094	-.180 -.140	-.131 -.080	-.160 -.120	-.062 .008	-.0688
III Ellis	-.162	-.059	-.360	-.234	-.170	-.1970
IV Saline	-.262	-.222	-.199	-.195	-.078	-.1912
V Rice	-.047	-.229	-.231	-.105	-.011	-.1246
Mean	-.0691	-.1539	-.2339	-.2100	-.1069	-.1547

Table 3. Probability of Correct Classification With No Thresholding (Design 1)

Site \ AI	I	II	III	IV	V	Mean
I Morton	0.563 .836	0.681 .813	0.637 .542	0.760 .567	0.417 .598	0.641
II Finney	.506 .536	.718 .547	.715 .728	.903 .883	.454 .285	.628
III Ellis	.831	.713	.497	.748	.549	.668
IV Saline	.587	.640	.602	.686	.602	.623
V Rice	.667	.763	.747	.772	.731	.736
Mean	.647	.969	.638	.760	.519	.652

Table 4. Probability of Correct Classification With 1-Percent Thresholding and Thresholded Pixels Considered as Nonwheat (Design 1)

Site \ AI	I	II	III	IV	V	Mean
I Morton	0.746 .779	0.814 .749	0.647 .583	0.777 .608	0.693 .580	0.698
II Finney	.596 .720	.807 .757	.822 .844	.803 .847	.541 .412	.715
III Ellis	.775	.657	.486	.700	.488	.621
IV Saline	.590	.598	.605	.594	.601	.598
V Rice	.715	.770	.766	.768	.766	.757
Mean	.703	.736	.679	.728	.583	.686

Table 5. Estimated Wheat Proportion Minus Ground-Truth Wheat Proportion With r = 0.3 and No Thresholding

Site \ AI	I	II	III	IV	V	Mean
I Morton	0.355 .135	0.038 -.141	-0.150 -.095	0.037 -.130	.077 -.179	-0.0053
II Finney	.301 .336	-.133 -.071	.258 .260	.076 .046	.064 .163	.1300
III Ellis	-.008	.048	-.254	-.152	-.036	-.0804
IV Saline	-.186	-.197	-.043	-.246	.084	-.1176
V Rice	.037	-.160	-.152	.017	.284	.0052
Mean	.1386	-.0880	-.0251	-.0503	.0653	.0081

Table 6. Estimated Wheat Proportion Minus Ground-Truth Wheat Proportion With r = 0.2 and No Thresholding

Site \ AI	I	II	III	IV	V	Mean
I Morton	0.349 .122	0.044 -.145	-0.143 -.086	0.000 -.176	0.070 -.173	-0.0138
II Finney	.313 .350	-.133 -.071	.251 .243	.074 .035	.069 .193	.1324
III Ellis	-.031	.068	-.211	-.014	-.050	-.0476
IV Saline	-.178	-.197	-.043	-.246	.090	-.1148
V Rice	.055	-.150	-.097	.031	.297	.0272
Mean	.1400	-.0834	-.0123	-.0423	.0708	.0146

Table 7. Estimated Wheat Proportion Minus Ground-Truth Wheat Proportion With r = 0.3 and 1-Percent Thresholding

Site \ AI	I	II	III	IV	V	Mean
I Morton	0.106 -.128	-0.026 -.235	-0.266 -.281	-0.154 -.358	-0.031 -.239	-0.1612
II Finney	.149 .147	-.180 -.141	.019 .049	-.091 -.083	-.020 .071	-.0080
III Ellis	-.093	-.021	-.293	-.199	-.104	-.1420
IV Saline	-.208	-.236	-.109	-.256	.058	-.1502
V Rice	-.027	-.189	-.177	-.028	.178	-.0486
Mean	-.0077	-.1469	-.1511	-.1670	-.0124	-.0970

Table 8. Estimated Wheat Proportion Minus Ground-Truth Wheat Proportion With r = 0.2 and 1-Percent Thresholding

Site \ AI	I	II	III	IV	V	Mean
I Morton	0.193 -.088	-0.004 -.233	-0.261 -.274	-0.130 -.339	-0.026 -.229	-0.1391
II Finney	.205 .205	-.180 -.141	.052 .094	-.023 -.036	.032 .151	.0359
III Ellis	-.088	.013	-.245	-.018	-.386	-.1448
IV Saline	-.208	-.236	-.109	-.256	.037	-.1544
V Rice	.001	-.172	-.121	.003	.225	-.0128
Mean	.0314	-.1361	-.1234	-.1141	-.0280	-.0740

Table 9. Analysis of Variance for the Proportion Error Before Thresholding

Source	Sum of squares	Degrees of freedom	Mean square	F-statistic
Mean	0.00513	1	0.00513	0.52501
Site(s)	.84810	4	.21202	21.68313[a]
Feature extraction (F)	.00375	2	.00188	.19188
F × S	.02326	8	.00291	.29740
AI	.51086	4	.12771	13.06091[a]
AI × S	1.18088	16	.07380	7.54780[a]
AI × F	.04203	8	.00525	.53724
AI × S × F	.06952	32	.00217	.22218
Error	.29335	30	.00978	

[a]Indicates significance at the 1-percent level.

NOTE: As usual, when the response variable is a proportion, the arcsine square root transformation is applied.

Table 10. Analysis of Variance for the Proportion Error After Thresholding

Source	Sum of squares	Degrees of freedom	Mean square	F-statistic
Mean	196.22948	1	196.22948	72.75373[a]
Site(s)	37.39921	4	9.34980	3.46652[b]
Feature extraction (F)	27.12776	2	13.56388	5.02892[b]
F × S	10.38357	8	1.29795	0.48122
AI	44.23023	4	11.05756	4.09968[b]
AI × S	134.21824	16	8.38864	3.11016[a]
AI × F	32.16168	8	4.02021	1.49053
AI × S × F	107.18278	32	3.34946	1.24184
Error	80.91523	30	2.69717	

[a]Indicates significance at the 1-percent level.
[b]Indicates significance at the 5-percent level.

NOTE: As usual, when the response variable is a proportion, the arcsine square root transformation is applied.

Table 11. Effect of Training-Field Labeling and Boundary Correction on the Estimate of Wheat Proportion With No Thresholding (Design 2)

Site	AI	Boundaries and labels uncorrected	Boundaries and labels corrected	Improvement due to correcting
Morton	1	0.270	0.253	0.017
	2	.111	.084	.027
	3	-.147	-.209	-.062
	4	-.054	-.106	-.052
	5	-.104	.031	.073
Finney	1	.389	.453	-.064
	2	.073	.049	.024
	3	.220	.143	.077
	4	.096	.081	.015
	5	.117	.025	.092
Morton mean		.015	.011	.003
Finney mean		.179	.150	.029
Overall mean		.097	.080	.016
MSE		.034852	.036219	–
Variance		.028248	.033061	.003336
Standard deviation		.168072	.181826	.05776

NOTES: The entries in the table refer to the PE (estimated proportion minus true proportion) with no thresholding.

$t = \frac{.016}{.05776}\sqrt{10} = .876$, not significant.

Mean square error is increased by 3.9 percent.

F = 100.5. The variance of the PE's from stratified, randomly selected fields is significantly lower than from AI-selected and ground-truth-corrected fields.

Table 12. Effect of Training-Field Labeling and Boundary Correction on the Estimate of Wheat Proportion With 1-Percent Thresholding (Design 2)

Site	AI	Boundaries and labels uncorrected, $\hat{p}$ - Pw	Boundaries and labels corrected, $\hat{p}$ - Pw	Improvement due to correcting
Morton	1	-0.075	0.163	-0.088
	2	-.097	-.128	-.031
	3	-.219	-.349	-.130
	4	-.290	-.309	-.019
	5	-.202	-.156	.046
Finney	1	.077	.026	.051
	2	-.160	-.109	.051
	3	-.098	-.168	-.070
	4	-.144	-.153	-.009
	5	-.012	-.149	-.137
Morton mean		-.177	-.156	-.044
Finney mean		-.067	-.111	-.023
Overall mean		-.122	-.133	-.034
MSE		.024991	.037096	–
Variance		.011230	.021504	.005083
Standard deviation		.105973	.146644	.071293

NOTE: The entries in the table refer to the PE (estimated proportion minus true proportion) with 1-percent thresholding. There is no significant improvement and no significant difference in variance.

Table 13. Improvement in Classification Accuracy Using No Thresholding (Design 2, see Section V)

Site	AI	Labels uncorrected	Labels corrected	Improvement
Morton	II	0.111	0.084	0.027
Morton	V	-.104	.031	.073
Finney	II	.073	.049	.024
Finney	IV	.096	.081	.015
Finney	V	.117	.025	.092

NOTES: Correcting the labeling and boundary errors on the average improves the accuracy of proportion estimation by approximately 5 percentage points when adequate training data are available. This represents a relative improvement of 46 percent.

Uncorrected mean square error = 0.010274.

Corrected mean square error = 0.003521.

Mean square error is reduced by 65.7 percent.

The significant average improvement is 0.0462.

Table 14. Improvement in Classification Accuracy Using 1-Percent Thresholding (Design 2, see Section V)

Site	AI	Boundaries and labels uncorrected	Boundaries and labels corrected	Improvement
Morton	II	-0.097	-0.128	-0.031
Morton	V	-.202	-.156	.046
Finney	II	-.160	-.109	.051
Finney	IV	-.144	-.153	-.009
Finney	V	-.012	-.149	-.137

NOTE: Average improvement = -0.016. Even when adequate training data are available, there is no improvement in proportion estimation if 1-percent thresholding is applied.

Table 15. Comparison of AI-Selected Fields With Ground-Truth-Selected Fields With No Thresholding (Design 3)

Site	AI selected and GT corrected training fields	Stratified random selected training fields
Morton	0.253	0.007
	.084	.013
	-.209	.002
	-.106	.002
	.031	.027
Finney	.453	.015
	.049	.048
	.143	.033
	.081	.046
	.025	.000
Morton mean	.0106	.0102
Finney mean	.1502	.0284
Overall mean	.0804	.0193
MSE	.0362188	.000669
Variance	.0330607	.000329
Standard deviation	.181826	.018144

NOTE: The entries are the estimated wheat proportion minus the ground-truth wheat proportion with no thresholding.

Table 16. Comparison of AI-Selected Fields With Ground-Truth-Selected Fields With 1-Percent Thresholding (Design 3)

Site	AI selected and GT corrected training fields	Stratified random selected training fields
Morton	0.163	-0.035
	-.128	-.039
	-.349	-.070
	-.309	-.103
	-.156	-.064
Finney	.026	-.049
	-.109	-.052
	-.168	-.052
	-.153	-.013
	-.149	-.072
Morton mean	-.1558	-.0622
Finney mean	-.1106	-.0476
Overall mean	-.1332	-.0549
MSE	.0370962	.0035513
Variance	.0215044	.0005969
Standard deviation	.1466437	.0244315

NOTE: The entries are the estimated wheat proportion minus the ground-truth wheat proportion with 1-percent thresholding.

Table 17. Correlation of the Proportion Estimate With Fraction of Training Fields Labeled as Wheat for Three Separate Experiments

Description of the classification and training set	Correlation coefficient	Level of confidence, %
1. Twenty-five training-field sets obtained for the five ITS's by the five AI's (design 1)	0.577	99
2. Ten sets of training fields randomly selected from stratified samples from Finney and Morton Counties (design 3)	.782	99
3. Single-phase classifications using cluster statistics for the five ITS's in a separate experiment	.738	99
4. Same as above using class statistics	.655	99
5. Same as above for multitemporal classifications (8 channels)	.964	99
6. Classifications of 54 segments using 4-channel data in a second separate experiment	.464	99
7. Classifications of 21 segments using 8-channel data which were a subset of the 54 segments where 2 passes were available	.752	99

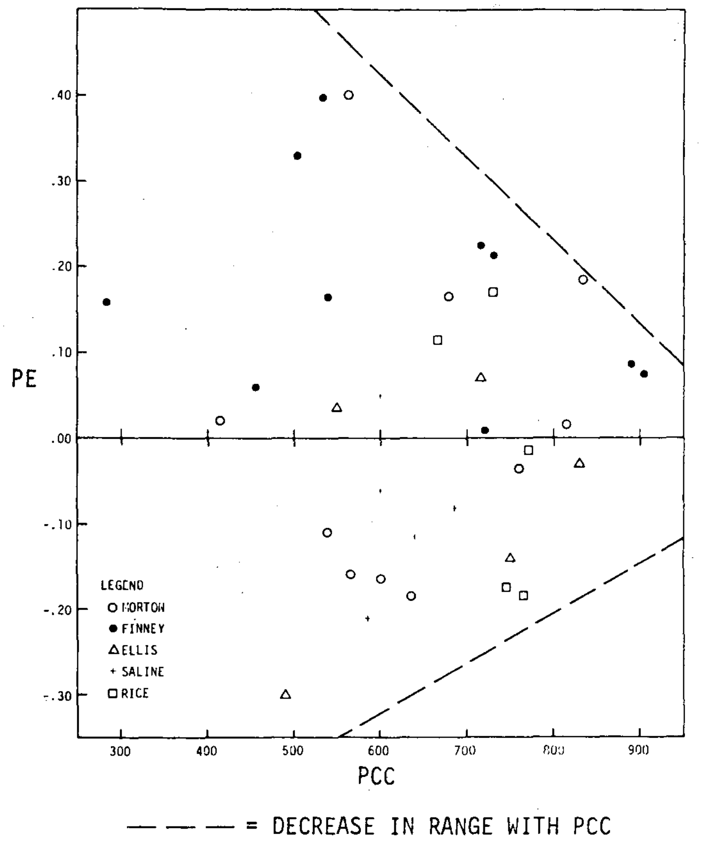

Figure 1. Proportion Error Versus PCC (Design 1).

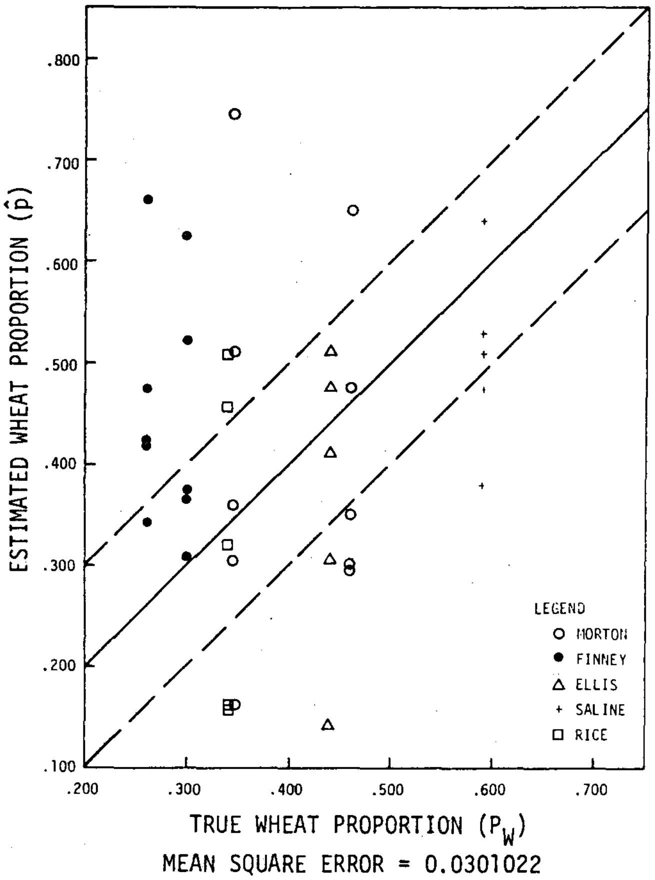

Figure 2. Estimated Wheat Proportion Versus True Wheat Proportion With No Thresholding (Design 1).

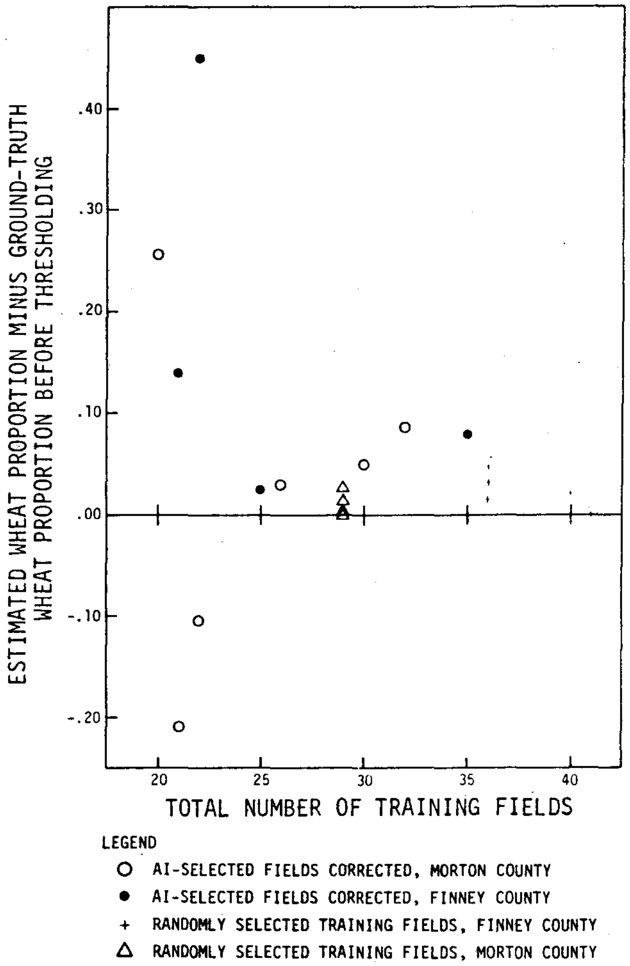

Figure 3. Proportion Error Versus Size of Training Data for the Morton and Finney County Test Sites (Design 3).