75-2

THE TECHNOLOGY OF LANDSAT IMAGERY AND ITS VALUE IN CROP ESTIMATION FOR THE U.S. DEPARTMENT OF AGRICULTURE

Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C.

MARCH 1976

The Technology of LANDSAT Imagery and Its Value in Crop Estimation for the U.S. Department of Agriculture

by

William H. Wigton

Statistical Reporting Service

U.S. Department of Agriculture

Washington, D. C.

December 1975

"Any sufficiently advanced technology is indistinguishable from magic."

Author Unknown

This is a synopsis of a much larger report of the Statistical Reporting Service, of the United States Department of Agriculture (USDA) to NASA. The title is "Crop Identification and Acreage Measurement Utilizing LANDSAT Imagery." and the contract number is 1013A, S-70251AG3. The full report can be obtained from Goddard Space Flight Center, Greenbelt, Maryland, U.S.A. 20771.

This paper was presented at the UN-FAO conference at Jakarta, Indonesia to dispel some of the confusion about the 'myth' of satellites. It is concerned with the data obtained from the satellite, with the method of computer crop identification and the process of integration of this data into the present estimating system used by the United States Department of Agriculture.

Description of LANDSAT Data

The satellite data used in this report is LANDSAT Multi-Spectral Scanner (MSS) data and is described in Section 3 of data User's Handbook. 1/

The MSS is a passive electro-optical system that can record radiant energy from the scene being sensed. All energy coming to earth from the sun is either reflected, scattered, or absorbed, and subsequently, emitted by objects on earth. 2/ The total radiance from an object is composed of two components, reflected radiance and emitted radiance. In general, the reflected radiance forms a dominant portion of the total radiance from an object at shorter wavelengths of the electromagnetic spectrum, while the emissive radiance becomes greater at the longer wavelengths. The combination of these two sources of energy would represent the total spectral response of the object. This, then, is the "spectral signature" of an object and it is the differences between such signatures which allows the classification of objects using the statistical techniques just discussed. The particular product in system corrected images refers to

---

1/ Published by Goddard Space Flight Center.

2/ Baker, J.R. and E.M. Mikhail, Geometric Analysis and Restitution of Digital Multispectral Scanner Data Arrays. LARS information note 052875.

products that contain the radiometric and initial spatial corrections introduced during the film conversion. Every picture element (pixel) is recorded with 4 variables - each variable corresponds to one of the 4 MSS bands. Table 23 shows the relationship between the MSS bands and light wavelengths.

Table 23—Sensor spectral band relationships.

Sensor	Spectral Band Number	Wavelengths (micrometers)	Color	Band Code
MSS	1	.5 - .6	Green	4
MSS	2	.6 - .7	Red	5
MSS	3	.7 - .8	Near Infrared	6
MSS	4	.8 - 1.1	Infrared	7

The numbers are similar to transmission values - zero radiances at Step 15 which is black on positives and maximum radiance at Step 1 which is white on positives. The radiance varies linearly with grey scale stop transmission between these values with the difference between each step corresponding to 1/14th of the maximum radiance. The recording format in the CCT is 8 bits, the sensor range is 7 bits, and the actual dynamic range of usable data is between 5 and 6 bits.

The analysis was started by first locating the test and training data (ground observations with either the Penn State University program (NMAP) or an in-house program (RADMAP) that produces grey scale maps. After the ground enumeration information was located on LANDSAT CCT's, rectangular areas within fields were located and punched using the LARS field description card format. Once these cards were obtained and checked, the statistics function in LARSYS was employed to extract univariate graphs to detect bimodal classes.

In most cases, analysis proceeded from the statistics program to the program for classification of points, but with the introduction of a feature to use prior probabilities. These classifications were stored on tape by file number so the print results function could be run more than once.

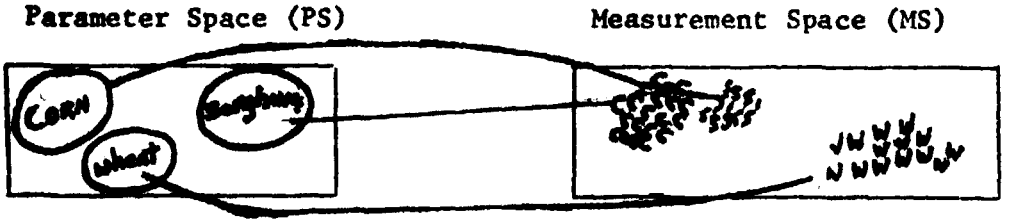
Discriminant Analysis

This background is intended to be general and enable the reader to understand the detailed computations and results that follow. Kendall and Stuart formulate Discriminant Analysis and Classification by stating ....

"We shall be concerned with problems of differentiating between two or more populations on the basis of multivariate measurements... We are given the existence of two or more populations and a sample of individuals from each. The problem is to set up a rule, based on measurements from these individuals, which will enable us to allot some new individual to the correct population when we do not know from which it emanates." 1/

For example, the land population of interest was the Southwest Crop Reporting District (CRD) in Kansas. Wheat, sorghums, corn, oats, rye, and pasture are the major populations of interest. From every acre in the CRD, we have light intensity readings for green light, red light, and two infrared wavelengths. These light intensities are multivariate measurements that will be used to allot or classify each data point into a crop type such as corn, wheat, or sorghums. A graphical representation of the above formulation would be as follows:

Figure 8--Conceptualized mapping from agricultural fields into measurement space.



A sample of fields from each crop type is selected and their respective light intensities obtained. These sample points are plotted on a two-dimensional graph showing relative positions of each crop type in the Measurement Space (MS). The problem is to partition the measurement space in some optimal fashion so that points are allotted as nearly correct as possible. Figure 9 shows the measurement space as it might be partitioned.

Figure 9--Partitioned measurement space.

1/ M.G. Kendall and A. Stuart, The Advanced Theory of Statistics, 2nd Ed., Vol. 3, page 314.

Any point, no matter where it is in MS will be classified as one of the three crops. An unknown point where the number 1 is located in Figure 9 will be classified as wheat because wheat is probably the group to which it belongs. Likewise, a point in position 2 would be classified as sorghum and a point in position 3 would be classified as corn. A point in position 4 would also be classified as wheat, but the probability that it is actually wheat is not as great as that of a point in position 1.

There are many ways to partition a measurement space. We have done a simple non-statistical partition above, merely by drawing lines. Visually partitioning the measurement space may work when it is one or two dimensional, but for more than two dimensional measurement spaces, a visual partition is not possible. For most LANDSAT and aerial photography classification studies a four dimensional measurement space has been used.

The method used in this report was that of constructing contour "surfaces" in the MS. These dividing surfaces were constructed so that points falling on the dividing surface have equal probabilities of being in either group on each side. Those points not on the dividing surface always have a greater probability of being classified into the crop for which the point is interior to the contour surface. If prior knowledge of the population density function indicates that the density is multivariate normal, then a multivariate normal density distribution will be estimated for each crop. It is hoped that the data is approximately multivariate normal since only the mean vector and covariance matrix is required to estimate a discriminant function. Usually small departures from normality will not invalidate the procedure, but certain types of departures (for example, bimodal data) may be very detrimental to the statistical technique. However, the error rate and estimator properties are dependent on the assumptions of the distributions and prior information.

For example, in this study a multivariate normal density was assumed so it becomes quite simple to estimate the density functions and the discriminant scores which in turn determine boundaries.

The discriminant score for ith population is:

$$P_i \, (2\pi)^{-\frac{q}{2}} \, |\Sigma_i|^{-\frac{1}{2}} \, e^{-\frac{1}{2}(\chi-\mu_i)' \Sigma_i (\chi-\mu_i)}$$

where P_i is the prior probability for the ith crop

Σ_i is the covariance matrix (qxq) for the ith crop

μ_i is the mean vector (q length) for the ith crop

χ is a set of measurements of an individual from the ith population.

or its equivalent discriminant score the $\log_{(e)}$ of S_i =

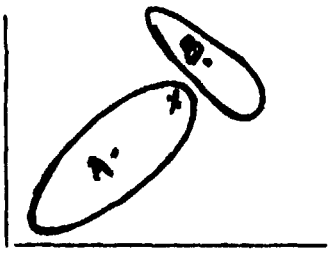
$$\log_e (P_i) - 1/2 \log_e |\Sigma_i| - 1/2 (\chi-\mu_i) \Sigma_i^{-1} (\chi-\mu_i)$$

The boundary between two populations is quadratic (curved) and the point χ that falls in the boundary have an equal probability of being in either population.

When an unknown land point is classified, its measurement vector is compared to the mean vector for each crop represented. The point is assigned to the crop whose mean point is "nearest" from a statistical point.

The procedure used for finding the "nearest" mean uses the Mahalanobis measure of distance, not the Euclidean. This is illustrated in Figure 10.

Figure 10--Measurement Space showing two crop density functions and an unknown point (χ).



The point χ is actually closest (Euclidean distance) to the mean vector (center point) of B. However, when one takes into account the variance and covariances, χ is found to be closest to Group A based on a probability concept and an outlier of Group B. Therefore, the point would be classified into Group A, because the probability that the point (χ) is a member of Group A is much greater than for Group B.

So the partitioning of the MS is done by computing the means for each crop type and using the Mahalanobis distances from this mean. This distance depends on the covariance matrix and is a measure of probability. The discriminant functions without prior probabilities are:

1) $(X - \bar{X}_i)' S^{-1} (X - \bar{X}_i)$, which is a sample estimate of $(X - \mu_i)' \Sigma^{-1} (\chi-\mu_i)$ if linear discriminant functions are used, and

2) $-1/2 \log_e |\Sigma_i| - 1/2 (X - \bar{X}_i)' S_i^{-1} (X - \bar{X}_i)$ if quadratic discriminant functions are used. These functions are the exponents of the density formula of the multivariate normal distribution C_{exp} - $1/2 (X - \mu_i) \Sigma_i^{-1} (X - \mu_i)$ depending on the i'th crop. If $\Sigma_i = \Sigma_j$ for all $i \neq j$ linear discriminant functions are used.

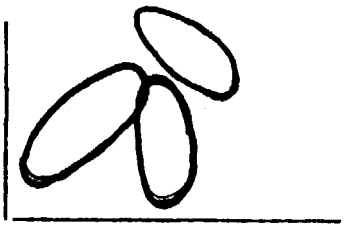
It is worth pointing out that if linear discriminant functions are used, one assumes (1) that $\Sigma_i = \Sigma_j$ and (2) that for all crops in the MS the major and minor axes are equal, and (3) the sample data of each crop has the same slope. Such an event in two-space is shown in Figure 11.

Figure 11--Measurement Space where crop types have same covariance matrix and slope.

This space can be partitioned effectively with straight lines thus we can use <u>linear</u> discriminant functions.

Figure 12 shows a MS where covariance matrices are not equal, and therefore, linear discriminant functions are not appropriate. In either case, the Mahalanobis distance is used.

Figure 12--Measurement Space when crops have different covariance matrices.



In Figure 11, even though a common center point is not present, a common covariance (ellipse) matrix would be computed. In Figure 12 a different covariance matrix will be needed for each crop type. When the off-diagonal elements in the covariance matrix are unequal, the slopes of the data are different and linear discriminant functions are <u>not</u> appropriate.

The above techniques follow from our first assumption that the data is normally distributed in the MS. In practice, however, one does not decide what the distribution of the population density is in MS and program the correct procedure. One uses the available procedures for analyzing data. Most available programs assume multivariate normal data because the program and the calculations are greatly simplified. Thus, it becomes necessary to justify the use of these simplified programs.

In order to explain better how a parametric procedure can reduce the work load, consider that the first step in the discriminant analysis (DA) is to estimate the population density function in the MS, with a sample of points from each crop. Once these population density functions have been estimated, then partitioning the space is extremely simple.

To estimate a multivariate population density in MS for corn where we have no prior information except sample data on corn is extremely difficult. If a sample of 1000 points was available, each of these 1000 data points would need to be stored in the computer. On the other hand, if we are working with a multi-dimensional normal distribution, theory tells us that the sufficient statistics are computed (mean vector variance matrix) and stored in the computer.

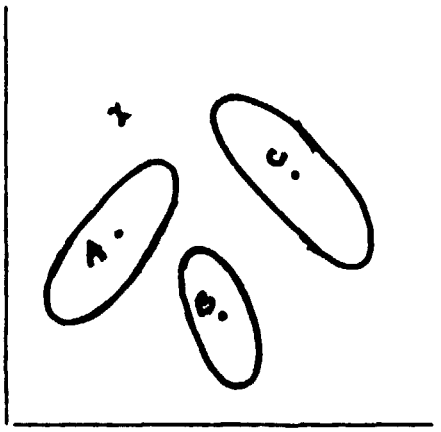
The individual data points could be discarded because no additional information about the population distribution in the MS is available in these points. (There would be information about how well the data fits the normal distribution in these 1000 data points).

Another consideration is that all the techniques we have described require independent random samples from each crop in order to estimate the population density in the MS (training data). This point is mentioned because most remote sensing analysts do not work with randomly selected points. In this study, we have tried to work with randomly selected fields. However, the points within these fields are not a random sample of all possible points in a given crop, but the data are nested within fields. Consequently, the random selection is restricted to the selection of fields within the randomly selected segments.

One type of prior information that can be used in the classification procedure is the relative frequency of occurrence (prior probabilities) for each of the K populations in the total land population. For example, if 1/3 of all land is wheat, and 1/3 is pasture as it might be in parts of Kansas, this information would be used and it would effect the partitioning of the measurement space accordingly. If a crop has a high chance of selection, then the area in the MS would be increased. Conversely, if a certain crop has a very low chance of occurrence, then the area in MS would be adjusted downwards.

One last point to be covered on procedures used would be to define what is meant by thresholding. Suppose some unknown crop for which there is no sample in the original data set is to be classified. With the present system, the point will be classified as Crop A, B, or C, depending on its probability of being in either A, B, or C. For example, in Figure 13, if the probability P(A|X) that the point χ was Crop A is .01 and P(B|X) = .001, and P(C|X) = .02 the point χ would be classified as belonging to Crop C, even though the probability is only .02. It would be an outlier in MS for Crop C, and therefore, we may want to let it remain unclassified.

Figure 13--Measurement Space showing an outlier and three crop areas with 95% confidence limits.



4.1.4 Results

The results will be presented by state since there was a slightly different situation in each state. All LANDSAT analysis is presented first then the aircraft follows.

Missouri LANDSAT:

The Crop Reporting District (CRD) that was the test site was in the southeast corner of the state. This area is outlined in black on the map of Missouri.

Summary of Results

The Missouri test site covers 4,660 square miles. There are 50 segments, each about a mile square. These segments constitute a random sample from all land areas. The ground enumeration was taken from these segments. This information was used for both training and testing.

Analysis of Missouri data was done using a tape that was assembled at LARS. The data for three dates, August 26, September 13, and October 21, 1972, were geometrically corrected then overlayed to create a tape with temporal data. Therefore, data used for analysis from three different times in the growing season was available and covered an area that contained 29 of the JES segments in this CRD. The principle results are summarized below:

1. A test was run on the covariance matrices between crops to see if they were equal. The results of this test were that they very likely were not equal. Thus, linear discriminant functions seemed inappropriate.

2. Best overall correct classification rate was 70%. This included using temporal overlays and using unequal prior probabilities.

3. Unequal prior probabilities for crops improved classification results by 10% over using the assumption of equal probabilities for crops.

4. The temporal data improved the classification by 10% even though the dates were not optimal.

5. One classification was run on data to estimate the effect of independent data. The difference was 6%, and was an over-estimate.

Data Analysis - LANDSAT

In the analysis, the equality of the covariance matrices was checked first because this is essential for the linear discriminant analysis assumptions to be valid. A test presented in Morrison's Multivariate Statistical Methods, page 152, was used to test the within crop covariance of LANDSAT data. This test is not robust with respect to certain departures from normality.

For the following example, August 26, 1972 imagery bands 4, 5, and 7 were used. The covariance matrices for cotton, soybeans, and grass were tested. The test was conducted as follows. The null hypothesis states that the covariance matrices are equal.

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3$$

The alternative hypothesis is:

$$H_1: \Sigma_i \neq \Sigma_j \quad \text{for some } i \neq j$$

S_i is an estimate of Σ_i based on m_i degrees of freedom where i is a crop.

$$S \text{ cotton} = \begin{vmatrix} 6.76 & 7.01298 & .4914 \\ 7.01298 & 11.0889 & -5.6643 \\ .4914 & -5.6643 & 39.69 \end{vmatrix}$$

$$S \text{ soybeans} = \begin{vmatrix} 6.6049 & 8.3623 & .8265 \\ 8.3623 & 13.9876 & -6.3146 \\ .8265 & -6.3398 & 64.6416 \end{vmatrix}$$

$$S \text{ grass} = \begin{vmatrix} 5.6169 & 5.8416 & .7525 \\ 5.8416 & 9.7344 & -6.3398 \\ .7525 & -6.3398 & 40.3225 \end{vmatrix}$$

Now we form the pooled estimate of Σ.

$$S = \sum_{i=1}^{k} \frac{m_i s_i}{m_i} = \begin{vmatrix} 6.5567 & 7.4436 & .6638 \\ 7.4436 & 12.1519 & -6.0189 \\ .6638 & -6.0189 & 50.2976 \end{vmatrix}$$

The statistic for the modified likelihood - ratio test is:

$$M = \sum_{m_i} \ln |S| - \sum_{i=1}^{k} m_i \ln |S_i|$$

= 149.25

Next, we form the scale factor:

$$C^{-1} = 1 - \frac{2P^2 + 3P - 1}{6(p+1)(k-1)} \sum_{i=1}^{k} \frac{1}{m_i} \frac{1}{\Sigma m_i} = .00678$$

and MC^{-1} is distributed approximately chi-squared with degrees of freedom $1/2\ (K-1)p(p+1)$ as m_i tends to infinity if H_0 is true.

$$MC^{-1} = 48.77 \quad d.f. = 12 \quad \alpha = .05 \quad \chi^2(12, \alpha = .05) = 22.36$$

Thus, we must reject the null hypothesis i.e. the data does not support the assumption that the covariance matrices are equal.

Therefore, the necessary assumptions for valid linear discriminant analysis are not met and better results might be attained by using quadratic discriminant functions. Generally, we used the quadratic approach on our analysis. However, it should be pointed out that upon close examination, the covariance matrices are very similar in many respects. Corresponding elements in the three covariance matrices are of at least the same order of magnitude and have the same sign. Under such conditions, it is possible to get acceptable results from a linear approach.

Conclusions of similar tests for the September 14, 1972 data were the same, the covariance matrices were unequal.

Results of the discriminant analysis (DA) are presented in a classification matrix (CM). Table 24 is an example of a CM using quadratic discriminant functions with unequal prior probabilities. The prior probabilities came from the June Survey early in the season. That is, it was not assumed that corn, cotton, soybeans, grass, and others all have the same probability of occurrence. The classification parameters were obtained from the same data that was used in the testing phase.

Although 12 bands were available, since three dates were involved, only nine were used in this study because three were of poor quality. There were two consecutive LANDSAT images that contained 29 segments. All data were used both to partition the measurement space (MS) and test the partition. The CM will be biased upward because data was used for both purposes, however, this bias should be small if ample data are available.

Table 24--Classification matrix of quadratic discriminant functions with unequal prior probabilities using data from three overflights1/, Missouri Study Area.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton....	927	79.7	739	2	137	26	23
Corn.....	58	44.8	9	26	7	14	1
Soybean...	852	71.8	99	12	612	96	23
Grass.....	240	53.3	42	1	66	128	4
Misc.....	140	89.3	17	2	44	13	64
Totals....	2217		906	43	866	277	125
Overall performance 70.8 percent							

1/
August 26, 1972, MSS bands 4, 5, 7
September 14, 1972, MSS bands 5, 7
October 2, 1972, MSS bands 4, 5, 6, 7

The leftmost column in Table 24 identifies the crop - cotton, corn, soybeans, grass, and miscellaneous. The next column gives the number of sample values in each of the crop classes. For example, there are 927 pixels to be classified. The next column tells the percent of these that were classified correctly as cotton (79.7%). The rest of the columns give the number of these pixels that were classified into each crop class, i.e. 739 were classified correctly as cotton, while the remainder were misclassified as follows: 2 of the 927 as corn, 137 as soybeans, 26 as grass, and 23 as miscellaneous. The overall performance in this table was 70.8 percent. To compute this figure, the correctly classified pixels were divided (the diagonal elements - 1569) by the total pixels 2217.

The prior probabilities used in this study were based on a statistical sampling of the entire land area. Data that is collected in this way enables the user to estimate the prior probability and take advantage of this procedure. Historic data could be used, but they are more difficult to justify when important changes between years are occurring.

The next table is the same as the last, except that equal prior probabilities were used.

Table 25--Classification matrix of quadratic discriminant functions with equal prior probabilities using data from three overflights 1/, Missouri Study Area.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton....	927	74.3	689	21	83	36	98
Corn.....	58	58.6	4	34	3	10	7
Soybean...	852	39.7	101	49	338	137	227
Grass.....	240	57.1	34	22	22	138	25
Misc.....	140	75.0	14	5	7	9	105
Total.....	2217		842	131	453	329	462
Overall performance 58.8 percent							

1/
August 26, 1972, MSS bands 4, 5, 7
September 14, 1972, MSS bands 5, 7
October 2, 1972, MSS bands 4, 5, 6, 7

Most classifications done so far by other remote sensing analysts have used this assumption that the crop classes are all equally likely to occur. Most people feel this assumption is not detrimental, however, this example illustrates that it can make a difference. Especially, if acreage for the crop classes does vary vastly or when crops are hard to distinguish. Two properties are worth noting, classification results, and the statistical properties are much better in Table 24 than in Table 25. For example, in Table 24 the total number of pixels classified as cotton is 906, compared to the actual number of 927. In Table 25, the number of cotton pixels is 842.

A similar comparison is even more drastic with soybeans. In Table 24, 866 pixels were classified as soybeans while 842 actual points were soybeans. In Table 25, there were 453 points classified as soybeans. Further, the statistical properties of the estimates are better since if the data is normal, and the prior probabilities are correct, we obtain unbiased estimates of crop categories and we can estimate the Bayes error rates (minimum error rates) using the classification.

A chi-square test for discriminatory power was run on the CM of Table 24 and 25. 1/ The null hypothesis is that the classification was done strictly at random. If the null hypothesis is correct, then the spectral information was useless as far as giving information that would help assign the data to a crop class. If the above hypothesis is correct, then the statistic $\frac{(n-e)^2}{e} + \frac{(\bar{n}-\bar{e})^2}{\bar{e}}$ has a chi-square distribution with 1 degree of freedom. Where n and $\bar{n}$ are the number of correctly classified and misclassified points respectively and e and $\bar{e}$ are the expected number of correctly classified and misclassified points under the null hypothesis.

The chi-square for Table 24 is 4626 and for Table 25 is 2782. These chi-square values with one degree of freedom are highly significant, and therefore, we conclude that the classification was not done at random. Another chi-square test based on the difference between the marginal sums and the correct number of data points in each class for Table 25 is as follows:

$$X^2_{(5)} = \frac{(906-927)^2}{927} + \frac{(43-58)^2}{58} + \frac{(866-852)^2}{852} + \frac{(277-240)^2}{240} + \frac{(140-125)^2}{125} =$$

$$.47 + 3.87 + .23 + 5.70 + 1.61 + = 11.89$$

This chi-square statistic is similar to the one before, except that there are 4 degrees of freedom. $\sum_{i=1}^{k} \frac{(n-e)^2}{e_i}$ where n and e have the same meaning as before.

This chi-square value of 11.89 is significant, and therefore, the hypothesis that the marginal totals in Table 24 are estimating the actual row totals is rejected. Note that the components for grass and corn are the major contributors to the significant chi-square.

The authors know of no statistical test that compare one C.M. with another C.M., but there are two criteria that can be used to help evaluate a certain C.M. The first criterion simply assigns each misclassified point a loss of 1 and each correctly classified point as loss of 0. Under this criterion, Table 24 has a loss value of 648 and Table 25 has a loss value of 914. This criterion is crude, but it seems reasonable for our purposes to give a misclassified corn pixel the same weight as the misclassified cotton pixel.

1/ S. James Press, Applied Multivariate Analysis, pages 381-383.

The next criterion is a bit more subtle. It uses the marginal totals in the C.M. For example, in Table 24 the column sum for cotton is 906. This means that 906 pixels were classified as cotton. Actually, there were 927 cotton pixels. In Table 25, there were 842 pixels classified into the cotton group. This is not close to the correct number of 927. The marginal estimate (906) from Table 24 is within 2 percent of the actual. In Table 25, the marginal estimate of 842 or within 9 percent. Table 26 presents these estimates along with the percentages of the true value.

Table 26--Marginal estimate and difference from actual values.

Group	Actual	Unequal Prior Probabilities			Equal Prior Probabilities		
		Estimate	Difference	Percent	Estimate	Difference	Percent
Cotton..	927	906	21	2.2	842	85	9.2
Corn....	58	43	15	25.9	131	73	125.9
Soybean.	852	866	14	1.6	453	399	46.8
Grass...	240	277	37	15.4	329	89	37.1
Winter Wheat...	85	27	27	68.2	346	261	307.1
Odd.....	55	98	43	78.2	116	61	110.9

In every case, unequal prior probabilities were superior to the equal prior probabilities model and in some cases, substantially so. For example, the number of corn pixels for Table 25 was 131 or 125.9 percent of the difference from the actual 58. The number of corn pixels for Table 24 is 43 or 25.9 percent of the difference from the actual 58 pixels. Soybeans, a very important item, also shows a significant improvement over the equal probability model. Actually, the soybean estimate for the equal prior probability model was 46.8 percent while the estimate for the unequal prior probability model was 1.6 percent.

Next, the point classification systems were compared to the per-field classification scheme. Table 27 presents the C.M. for the per-field classifier system. With a point classification system, each point in a field can be assigned to any of the crop categories. With the sample classifier, all points in the field are assigned to the same crop class. One drawback to the procedure is that there were a large number of fields that were not assigned to a crop because the data set was not large enough. The technique requires the covariance matrix to be inverted and therefore, p+1 data points are required (where p is the number of variables). However, if enough points are present, classification performance has generally been found to be excellent.

In the work done in Missouri using the sample classifier, about 40 percent of the fields were not classified because the required number of points for the classifier (10 is this particular case) exceeded the number of points present within the defined fields. Of the total number of fields, 32.9 percent were correctly identified. Considering only those fields which were classified, 54 percent were classified correctly.

Table 27--Per-field classification matrix based on data from 3 overflights.1/

Group	No. of fields	Percent fields correct	No. of samples	COTTON	CORN	SOY-BEANS	GRASS	MISC.	NOT CLASSIFIED
Cotton:	38	63.2	927	24	0	2	0	1	11
Corn..:	7	14.3	58	0	1	0	1	1	4
Soy-beans.:	58	25.9	852	9	3	15	3	8	20
Grass.:	31	9.7	240	3	1	1	3	2	21
Misc..:	9	44.4	140	1	0	1	1	4	2
Totals:	143	32.9	2217	37	5	19	8	16	58

1/
August 26, 1972, MSS bands 4, 5, 7
September 14, 1972, MSS bands 5, 7
October 2, 1972, MSS bands 4, 5, 6, 7

Temporal Overlay

The next analysis investigated the value of a temporal overlay of the three LANDSAT passes. This particular data set was a temporal overlay of three LANDSAT passes. Each pass could also be compared with the three passes. However, there were 3 bad bands in the total of 12. Two poor quality bands were in the September 14 imagery and one poor quality band was in the August 26 imagery. This makes it difficult to compare the three dates since the number of bands were confounded with dates. Nevertheless, the C.M.'s for each date are presented in Tables 28, 29, and 30. These tables can be compared to the 9 band-overlay of Table 24 since they are all unequal prior probability models.

Table 28--Classification matrix using August 26, 1972, MSS bands 4, 5, and 7 with unequal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	60.6	562	1	311	22	31
Corn.....	58	10.3	12	6	30	2	8
Soybean..	852	86.0	70	2	733	29	18
Grass....	240	8.3	42	7	167	20	3
Misc.....	140	31.4	9	3	76	8	44
Totals...	2217		696	19	1317	81	104

Overall performance 61.5 percent

Table 29--Classification matrix using September 13, 1972, MSS bands 5 and 7 with unequal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	69.7	646	0	246	14	21
Corn.....	58	0.0	12	0	16	20	10
Soybean..	852	67.6	175	1	576	74	26
Grass....	240	42.1	40	0	97	101	2
Misc.....	140	22.8	14	2	82	10	32
Totals...	2217		887	3	1017	219	91

Overall performance 61.0 percent

Table 30--Classification matrix using October 2, 1972, MSS bands 4, 5, 6, and 7 with unequal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	73.2	679	6	161	59	22
Corn.....	58	12.1	30	7	14	1	6
Soybean..	852	62.4	200	7	532	76	37
Grass....	240	27.9	83	0	89	67	1
Misc.....	140	17.9	30	1	73	11	25
Totals...	2217		1022	21	869	214	91

Overall performance 59.1 percent

Table 31 summarizes these three classification matrices in 1 table.

Table 31--Comparison of multitemporal classification performance to classification of single dates. 1/ Missouri Study Area.

Group	Multitemporal	Aug. 26	Sept. 14	Oct. 2
Cotton	29.7	60.6	69.7	73.2
Corn	44.8	10.3	0.0	12.1
Soybeans	71.8	86.0	67.6	62.4
Grass	53.3	8.3	42.1	27.9
Misc.	89.3	31.4	22.8	17.9
Overall	70.8	61.6	61.1	59.2

1/ Unequal prior probabilities were used for all classification.

The same classifications were run for all dates individually except that equal prior probabilities were used.

Table 32--Classification matrix for August 26, 1972, based on MSS bands 4, 5, and 7 using equal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classifed into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	60.7	563	92	108	63	101
Corn.....	58	56.9	2	33	0	7	16
Soybean..	852	15.3	57	72	130	245	348
Grass....	240	45.4	32	41	26	109	32
Misc.....	140	62.9	11	10	13	18	88
Totals...	2217		665	248	277	442	585
Overall performance 41.6 percent							

Table 33--Classification matrix for September 13, 1972 based on MSS bands 5 and 7 using equal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	60.7	563	92	108	63	101
Corn.....	58	56.9	2	33	0	7	16
Soybean..	952	15.3	57	72	130	245	348
Grass....	240	45.4	32	41	26	109	32
Misc.....	140	62.9	11	10	13	18	88
Totals...	2217		665	248	277	422	585
Overall pefformance 50.8 percent							

Table 34--Classification matrix for October 2, 1972 based on MSS bands 4, 5, 6, and 7 using equal prior probabilities.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	66.7	618	35	30	149	95
Corn.....	58	37.9	21	22	4	4	7
Soybean..	952	20.8	142	46	177	141	346
Grass....	240	42.5	58	9	23	102	48
Misc.....	140	60.7	20	8	8	18	85
Totals...	2217		860	120	242	414	581
Overall performance 45.3 percent							

Table 35 summarizes these tables.

Table 35--Comparison of multitemporal classification performance to classifications of single dates using equal prior probabilities. 1/ Missouri Study Area.

Group	Multitemporal	Aug. 26	Sept. 13	Oct. 2
Cotton	74.3	60.7	71.4	66.2
Corn	58.6	56.9	34.5	37.9
Soybeans	39.7	15.3	28.9	20.8
Grass	57.1	45.4	44.6	42.5
Misc.	75.0	62.9	65.7	60.7
Overall	58.8	41.6	50.8	45.3

The temporal overlay classification of Table 25 shows an overall performance of 58.8 percent as compared to 41.6 percent, 50.8 percent, and 45.3 percent, respectively, for Tables 32, 33, 34. Based on these comparisons, the temporal overlay does improve the classification, However, the evaluation can become more difficult to interpret in the temporal overlay tapes because of changes in land use from one date to the next. Thus, the time of year becomes very important in areas where double-cropping is common or preparation of land follows each crop. It should be pointed out that these dates were not optimal. Other dates would have given different results.

Independent Test Data

The last exercise was completed to estimate the C.M. in Missouri on independent data. Since the number of fields and points within are small and the area covered is large, we need more training data to represent the total area. It did not seem possible to divide the set into halves and still have enough training data. It was decided to use a jacknife procedure. This procedure has the advantage of giving unbiased estimates that are simple to calculate. The data were divided into three equal subgroups, two groups were used to train with and the third group was used as a test group. This was repeated three times, each time with a different group used as test data. These three tables are presented separately, then the three are combined and presented to give an unbiased estimate of the classification matrix where independent test data is used. By using independent data, it is hoped that the bias caused by using the same data for both training and testing would be eliminated, but the variance of each item in the latter tables may be somewhat higher than those in the previous tables since a smaller data set was used.

One cotton field of 27 points was not included in any of the three groups. So the total in Table 39 is 27 pixels smaller than the total of earlier tables. Table 39 is the matrix sum of Tables 36, 37, and 38.

Table 36--Classification matrix using August 26, 1972, MSS bands 4, 5, and 7 with subgroups 2 and 3 as training data and subgroup 1 as test data.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton	479	56.2	269	11	129	36	34
Soybean..	138	45.7	35	6	63	17	17
Grass	66	34.8	15	7	15	23	6
Misc.	68	16.2	1	4	39	13	11
Totals	751		320	28	246	89	68
Overall performance 48.7 percent							

Table 37--Classification matrix using August 26, 1972 MSS bands 4, 5, and 7 with subgroups 1 and 3 as training data and subgroup 2 as test data.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	290	57.6	167	36	11	19	57
Corn.....	29	13.8	1	4	0	8	16
Soybean..	308	13.0	48	53	40	20	147
Grass....	42	28.6	1	11	4	12	14
Misc.....	57	78.9	0	2	8	2	45
Totals...	726		217	106	64	63	279
Overall performance 36.9 percent							

Table 38--Classification matrix using August 26, 1972 MSS bands 4, 5, and 7 with subgroups 1 and 2 as training data and subgroup 3 as test data.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	131	47.3	62	22	1	22	24
Corn.....	29	41.4	3	12	2	5	7
Soybean..	406	200	6	29	8	137	226
Grass....	132	43.2	20	27	0	57	28
Misc.....	15	0.0	5	2	0	8	0
Totals...	713		96	92	11	229	285
Overall performance 19.5 percent							

Table 39--Classification matrix combining Tables 36, 37, and 38.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	900	55.3	498	69	141	77	115
Corn.....	58	27.6	4	16	2	13	23
Soybean..	852	13.0	89	88	111	174	390
Grass....	240	28.3	36	45	19	92	48
Misc.....	140	40.0	6	8	47	23	56
Totals...	2190		633	226	320	379	632
Overall performance 34.6 percent							

The comparable classification where non-independent data was used is shown in Table 40.

Table 40--Classification matrix using August 26, 1972, MSS bands 4, 5, and 7.

Group	No. of sample points	Percent Correct	Number of samples classified into				
			Cotton	Corn	Soybean	Grass	Miscellaneous
Cotton...	927	60.7	563	92	108	63	101
Corn.....	58	56.9	2	33	0	7	16
Soybean..	852	15.3	57	72	130	245	348
Grass....	240	45.4	32	41	26	109	32
Misc.....	140	93.6	11	10	13	18	131
Totals...	2217		665	248	277	442	585
Overall performance 43.6 percent							

Anytime the results differ this much between data sets, we know the data set is either too small or the bias is large. Obviously, we have not reached the point where we have covergence of parameters based on independent and non-independent data sets. The sample sizes necessary depend on the variation in the data set and the variation in the data set is generally a function of how dispersed the data really is. One thing is certain with a small data set, either procedure may lead to erroneous conclusions.

Kansas:

The LANDSAT analysis was done on the CRD in the southwest corner of the State.

Analysis of Kansas LANDSAT Data

The objectives of the analysis of Kansas LANDSAT data were the following:

1. Test of the covariance matrices of the most important crops to see if they were equal.

2. Computations of the classification rates for the Kansas test site.

3. Computations of the correlation coefficients between ground observation acreage and classified pixels.

4. A study of the effect of classification in one LANDSAT frame using training parameters from an adjoining pass taken one day apart.

5. A study of the classification of a Kansas county.

Approach:

1. LANDSAT imagery for the study area was too cloudy to be useful, prior to September 21, 1972. The study was based on September 21 and 22 imagery. The area of interest in Kansas was divided by two LANDSAT passes, thus the training data was also divided. Twenty-two segments were in the September 21 imagery. Seven of these segments were hidden by clouds. Therefore, 15 segments were used as training and test data.

   Since the time of year was not conducive to optimal results, a visual inspection of the grey-scale printout of MSS band 5 and ground truth was used to select particular fields to use as training fields; i.e. those fields which were partially harvested and those with a confusion of symbols were discarded. Another reason for selecting fields was to compare parameters from one pass with those from another as described in this report.

As a first step, the covariance matrices of the most important crops were compared and tested within frames and between frames. Tables 41 and 42 show the pertinent data.

The test criterion was computed and indicates that the within-crop covariances are statistically different. Also, the covariances between frames for the same crops were tested and are significantly different.

This would indicate that quadratic discriminant analysis could produce better results. In addition, a method of signature extension would be complicated if one wished to go from one frame to another.

2. The next step was to employ the quadratic classifier for the training data. The classification based on these select fields is presented in Table 43.

   The overall performance was 91.2%. The classification used the standard pointwise quadratic discriminant functions found in LARSYS with the added feature of allowing unequal prior probabilities for the different crops. The unequal prior probabilities use information that is available about the likelihood of certain crops. If, for example, corn is more likely to be encountered than grain sorghum, corn is given a higher chance of occurrence. In most classifications using unequal prior probabilities done in Kansas, the prior probabilities were:

   1) Alfalfa - .03
   2) Pasture - .72
   3) Corn - .09
   4) Grain Sorghum - .16

Prior probabilities in this report were computed from a probability survey conducted by the Statistical Reporting Service in June 1972, (June Enumerative Survey).

In Table 43, the number of pixels to be classified are not proportional to the prior probabilities. The prior probabilities are based on acreage of all segments in the Crop Reporting District, and not the segments in frame 1060-16512. Development of proper prior probabilities for areas divided by LANDSAT passes presents additional problems. A better correspondence would have resulted in higher overall classification; however, 91.2% is very good.

Table 41--Covariance matrices and mean vectors for frame 1060-16512. (September 21, 1972).

	Mean	Covariance			
Alfalfa n = 43					
	26.63	3.430			
	19.58	4.531	8.535		
	50.81	-2.357	-8.199	27.346	
	30.28	-2.751	-7.357	16.363	12.301
Pasture n = 6378					
	29.70	10.926			
	26.36	12.975	21.821		
	56.88	10.351	12.698	22.487	
	20.07	4.405	4.332	11.388	7.339
Corn n = 332					
	31.63	46.883			
	29.71	77.701	133.003		
	43.03	26.525	42.905	33.798	
	24.84	2.728	-6.399	11.275	10.978
Grain Sorghum n = 508					
	32.21	115.096			
	27.32	130.402	154.965		
	43.78	78.251	85.757	76.431	
	25.65	18.089	16.152	29.548	18.198

Table 42--Covariance matrices and mean vectors for frame 1061-16570. (September 22, 1972).

	Mean	Covariance			
Alfalfa n = 78					
	24.23	8.180			
	15.96	12.793	24.701		
	55.61	-18.345	036.494	71.234	
	34.51	-15.063	-29.604	50.802	39.313
Pasture n = 320					
	28.62	5.290			
	25.53	6.109	11.002		
	35.98	3.534	3.061	19.272	
	19.81	1.056	0	11.213	8.237
Corn n = 337					
	24.52	1.877			
	19.91	2.183	9.120		
	36.88	0.339	-5.114	17.056	
	22.82	-0.081	-5.291	11.039	8.820
Grain Sorghum n = 177					
	27.16	32.718			
	22.76	49.217	77.088		
	43.69	2.100	2.865	16.646	
	27.09	-15.639	-24.393	10.975	19.448

Table 43--Classification matrix for September 21, 1972 MSS bands 4, 5, and 7, using quadratic discriminant functions with unequal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Sorghum	Threshold
Alfalfa..	43	100.0	43	0	0	0	0
Pasture..	172	98.3	0	169	2	1	0
Corn.....	51	90.2	0	1	46	4	0
Grain Sorghum..	78	69.2	0	10	14	54	0
Totals...	344		43	180	62	59	0

Overall performance 91.2%

A classification was then done using all identifiable fields in the 15 segments. The results of this classification are presented in Table 44. The overall performance was 90.2%.

There was a small decrease in overall performance between Table 43 and Table 44. However, a random sample of ground truth yields a better representation of all land and allows statistical inferences about the pixels.

The second pass required to cover the Kansas test site was analyzed in the same way as described above. The second scene contained 23 segments, but one of these segments fell in a non-agricultural area. In addition, to the random segments, two additional segments were selected which contained sugar beets.

Table 45 presents the classification of select fields for the second pass. The fields were selected from the grey-scale printout as described above. The overall performance was 75.5%.

Table 44--Classification matrix for September 21, 1972 imagery (MSS bands 4, 5, 6, and 7), using quadratic discriminant functions with unequal prior probabilities in Kansas test site.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	43	93.0	40	2	0	1	0
Pasture..	6378	95.0	23	6061	123	142	29
Corn.....	332	37.7	38	110	125	59	00
Grain Sorghum..	508	64.8	38	77	60	329	44
Totals...	7261		139	6250	308	531	33
Overall performance 90.2%							

Table 45--Classification matrix for September 22, 1972 imagery (MSS bands 4, 5, 6, and 7), using quadratic discriminant functions with unequal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	78	84.6	66	12	0	0	0
Pasture..	230	93.0	0	214	11	5	0
Corn.....	337	65.0	0	93	219	25	0
Grain Sorghum..	177	63.9	3	34	18	122	0
Totals...	822		69	353	248	152	0
Overall performance 75.5%							

Table 46 represents a classification of the second scene, using all identifiable fields. The overall performance was 65.8%. This decrease in performance could be attributed to several things. The number of crops being classified was increased from four to seven. Increasing the number of crops will reduce the performance. Secondly, there was a confusion between most crops and pasture. This could have resulted from using late September imagery; all crops are spectrally similar. Thirdly, the frequency of the data pixels presented for classification differed drastically from the prior probabilities used.

Table 47 is a classification study using the same select training fields that were used in Table 45. However, in Table 47 equal prior probabilities were applied. In Table 47, the overall performance at 79.2% is actually better than the 75.5% in Table 45. Applying prior probabilities based on all fields to a non-random selection of fields in a particular area is the cause for the lower classification in Table 45.

Table 48 presents a classification of all identifiable fields in scene 1061-16570, using equal prior probabilities. This table is comparable with the weighted classification presented in Table 46. The overall performance was increased 4.4% by using prior probabilities. When all fields are used in the classification, the total acres per crop more closely estimate the true prior probabilities of the model.

The increase caused by using unequal prior probabilities in Kansas was not as great as it had been in other areas. The smaller gain from prior probabilities is perhaps caused by the fact that the LANDSAT data contained more information; i.e., the classes were more separable. Ihus, the expected gain from prior probabilities is greater in areas where classification is poorer.

3. The correlations between acres and pixels were calculated. Coordinates of ground truth segments were carefully defined. The training data from each scene were used to classify the segments in that scene. The classified pixels in the two scenes were then combined (i.e., Tables 44 and 46 were combined) and correlations with known ground truth acreage were computed.

Correlations between acreage and pixels were calculated as follows:

Total Acreage vs Total Pixel	$r^2 = .88$	$r = .94$
Pasture Acreage vs Pasture Pixel	$r^2 = .84$	$r = .92$
Corn Acreage vs Corn Pixel	$r^2 = .62$	$r = .79$
Grain Sorghum vs Grain Sorghum Pixel	$r^2 = .58$	$r = .76$

Table 46--Classification matrix for September 22, 1972 imagery (MSS bands 4, 5, 6, and 7), using unequal prior probabilities, Kansas, all fields.

Class	No. of sample points	Percent Correct	Number of samples classified into						
			Alfalfa	Pasture	Corn	Grain Sorghum	Winter Wheat	Sugar Beets	Threshold
Alfalfa	287	56.4	162	57	12	23	16	6	0
Pasture	4975	90.6	19	4508	45	44	156	0	23
Corn	1698	40.8	1	684	693	174	99	0	0
Grain Sorghum	2869	55.3	89	300	357	1586	265	0	4
Winter Wheat	863	13.3	14	431	16	41	115	0	4
Fallow	1508	64.6	10	285	44	56	134	2	3
Sugar Beets	25	0.0	16	2	1	1	5	0	0
Totals	12225		311	6267	1168	1925	790	8	34

Overall performance 65.8 percent

Table 47--Classification matrix for September 22, 1972 imagery, MSS bands 4, 5, 6, and 7, using quadratic discriminant functions with equal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	78	84.6	66	11	0	1	0
Pasture..	230	75.2	3	173	38	16	0
Corn.....	337	87.5	0	29	295	13	0
Grain Sorghum..	177	66.1	14	16	30	117	0
Totals...	822		83	299	363	147	0
Overall performance 79.2%							

When pixels and acreage are this highly correlated, remotely sensed data is beneficial.

4. In this study, the statistics compiled on one LANDSAT frame were used to classify points in the adjacent frame. As described earlier, two adjacent passes were used to obtain necessary coverage of Kansas. The select fields from both scenes (as described in Section A), had four classes (alfalfa, pasture, corn, grain sorghum. These four classes were also the classes for the "all fields" in frame 1060-16512. One requirement is that the same classes be used for training as those classified. The classification used the quadratic discriminant function with unequal prior probabilities.

   Table 49 presents the results of classifying the select fields in frame 1060-16512, using training statistics generated from select fields in frames 1061-16570. The overall performance was 54.4%; however, the average performance by classes 1/ was 33.3% correct classification. The 100% correct classification of the pasture class greatly influenced the overall classification.

1/ The average performance by classes is computed by averaging the percent identified for each class.

Table 48--Classification matrix for September 22, 1972 imagery, 4 bands using equal prior probabilities, Kansas.

Class	No. of sample points	Percent Correct	Number of samples classified into							
			Alfalfa	Pasture	Corn	Grain Sorghum	Winter Wheat	Fallow	Sugar Beets	Threshold
Alfalfa.....	287	50.5	145	18	30	9	24	4	57	0
Pasture.....	4975	80.1	61	3986	371	66	340	106	22	23
Corn,.....	1698	70.3	80	267	1193	69	39	32	18	0
Grain Sorghum..	2869	42.1	496	115	620	1209	149	103	174	3
Winter Wheat...	863	23.4	20	350	50	44	202	149	44	4
Fallow.....	1508	50.5	18	208	79	120	256	762	62	3
Sugar Beets....	25	56.0	6	2	2	0	1	0	14	0
Totals.....	12225		826	4946	2345	1517	1011	1156	391	33

Overall performance 61.4%

Table 49--Classification matrix of select fields in frame 1060-16512 classification, using statistics from select fields in frame 1061-16570.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	43	0.0	0	41	0	1	1
Pasture..	172	100.0	0	172	0	0	0
Corn.....	51	0.0	3	7	0	41	0
Grain Sorghum..	78	33.3	7	28	15	26	2
Totals...	344		10	248	15	68	3
Overall performance 54.4%							

Table 50 is a classification of all identifiable fields in the segments in frame 1060-16512, using the statistics generated from the select fields in frame 1061-16570. The classifications with an overall performance of 65.5% and an average class performance of 48.5% are very good. Here again, it was the correctly classified pasture points which kept the averages high. In Table 50, more fields were classified and the influence of prior probabilities was more beneficial than in the cases where select fields were classified.

Table 51 shows a classification of select fields in frame 1061-16570, using statistics generated from all fields in frame 1060-16512. In this study the overall performance slipped to 49.0% but the average class performance was 59.1%. Classification was very good in all classes except corn, which was confused with pasture and grain sorghum. The time of year may have caused this confusion.

5. The border of Stevens County, Kansas was drawn on a grey-scale map of MSS band 5. The area was then defined on punch cards and classified. Training data for the classification were obtained from segments in the Crop Reporting District which contains Stevens County. Three of these segments were actually in Stevens County. A total of 410,505 pixels were classified which correspond to a calculated 466,560 acres in the county.

Table 50--Classification matrix of all fields in frame 1060-16512 classification, using statistics generated from "select fields" in frame 1061-16570.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	43	65.1	28	3	0	12	0
Pasture..	6378	93.2	7	5943	11	277	140
Corn.....	332	7.5	8	79	25	204	16
Grain Sorghum..	508	28.3	16	105	75	144	168
Totals...	7261		59	6130	111	637	324

Overall performance 85.5%

Table 51--Classification matrix of select fields in frame 1061-16570 classification, using statistics generated from "all fields" in frame 1060-16512.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa..	78	80.8	63	12	0	0	3
Pasture..	230	94.3	0	217	4	8	1
Corn.....	337	9.2	5	140	31	161	0
Grain Sorghum..	177	52.0	12	30	43	92	0
Totals...	822		80	399	78	261	4

Overall performance 49.0%

Alfalfa, pasture, corn, and grain sorghum were the crops classified. The following classification was obtained:

Number of Pixels	Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
410,505	5,362	172,021	30,448	165,107	37,567
	1.3%	41.9%	7.4%	40.2%	9.2%

The prior probabilities as a percentage which were applied were the following:

Alfalfa	3%
Pasture	72%
Corn	9%
Grain Sorghum	16%

There is confusion between pasture and grain sorghum. Ways to use this data to produce a final estimate will be discussed in the section on estimation.

South Dakota

The test site in South Dakota is in the eastern part of the State.

Analysis of LANDSAT Data in South Dakota

Objectives:

The objective of this section was to determine the classification accuracy in the South Dakota test site.

Approach:

Imagery for three dates was available. However, the August and early September imagery was too cloudy to be useful. Thus, later September imagery was used. All 34 segments were contained in one LANDSAT frame (1060-16491). The segments and fields within segments were located and defined on punch cards. These segments were used for both training and classifying.

The LARS classifier with unequal prior probabilities was used. The classifier is a standard discriminant analysis.

Table 52 presents a classification of pixels in all segments in South Dakota. The overall performance was 30%, but the average class performance was 15%. Almost all classes in Table 52 were classified as either pasture or oats.

There were two reasons for this. First, prior probabilities used were large for pasture and oats, and second, the spectral data is quite similar at this period of time for all crops.

An attempt to improve the classification results was made by selecting fields that looked homogeneous.

These selected fields were used as training data and then classified. The results of this classification are presented in Table 53. The overall performance was 26% and the average class performance was 44%. There appears to be very little information in the data which would aid in the separation of crops. The influence of the prior probabilities again was the reason pasture and oats had high correct classification rates.

There must be reasons for the very poor classification rates. As an attempt to determine the reasons for the poor results, we have studied the means and covariances. They are in Table 54. It appears to be impossible to separate these classes with this data. Simply looking at the data does not necessarily show the true multivariate situation is four dimensional - but it does give an indication.

Summary

In South Dakota, late September imagery was used because of cloud cover in earlier imagery. Classification results were poor. Examination of Table 54 showed very little information in the data for the separation of the classes of interest. This late in the season, crops were classified as either pasture or oats.

The use of homogeneous fields selected from gray scale printouts and ground truth did not improve classification, and actually reduced the overall performance rates.

Table 52--Classification matrix for September 21, 1972 imagery (MSS bands 4, 5, 6, and 7), using unequal 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.

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

Idle	Means	Number 19		Covariance Matrix		
	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 Fallow	Means	Number 82		Covariance Matrix		
	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.

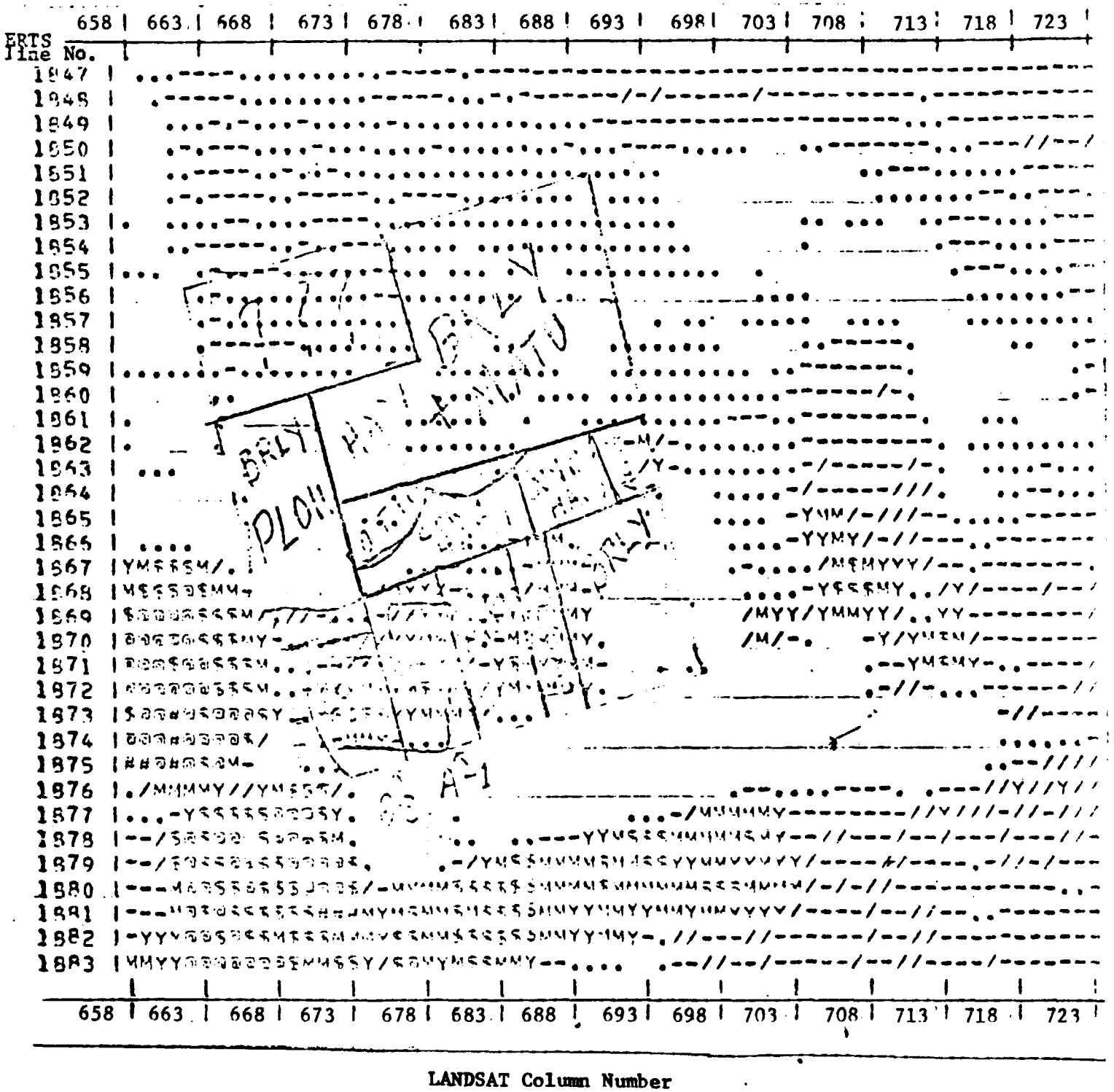
Obviously, the classification is not as good as we expected; however, by chance, one would expect only 8% correct classification for 12 crop categories. Another possible problem with the classification is that some field boundaries, sometimes, fall on adjacent points and since the pixels are partially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

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

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



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 has 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 imagewas 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.

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	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.

4.2 Crop Acreage Estimation

The objective of this section is to present a procedure that will use classifcation 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 57.

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 59--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:

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

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:

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

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

$\bar{x}_{\text{sample } i}$ 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:

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

where $\text{Var}(\hat{Y}_i)$ 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 $\text{Var}(\hat{y})$.

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:

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

$\hat{Y} = (280)(332) = 92,960$ 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:

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

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

Table 60 shows acreage estimates with variance and coefficients of variation for various crops with the aid of LANDSAT data.

Table 61 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 com-

Table 60--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 61--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%

puter 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:

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

The variance would be similar to the previous formula:

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

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

Area Sampling Frame Stratification

An additional advantage in satellite data, that has been run through the computer and given a crop "tag," is its function for the improvement of land use stratification. Under the present procedure for construction of area sampling frames, land use stratification is used. Trained frame construction personnel inspect aerial photography and estimate the area that is cultivated for each block of land. Those blocks of land, which have similar percentages of cultivated land are grouped and assumed to be homogeneous. In most instances, they are homogeneous. However, there are exceptions - instances when the cultivated acreage is similar, the crops are different and the variances must be computed for the individual crops. The objective in homogeneous grouping is the reduction of variance for individual crops.

Strata are divided - crops in areas of high concentration are placed in one category and those in areas of low concentration in another. This method assumes that year to year variation does not change from area to area but merely from field to field within an area. Usually, this assumption is true. However, the results presented here do not deal with the year to year variation since only one year has been studied. 1/

The study was conducted as follows:

1. Milan County, Texas was divided into 105 primary sampling units. Each primary sampling unit (PSU) has unambiguous boundaries such as rivers and roads and dimensions which are between eight and twelve square miles wherever these boundaries are available. In addition, agriculture within each primary sampling unit should be relatively homogeneous. However, the agriculture in one PSU may be quite different from the agriculture in the next PSU.

   Satellite data was located, computer classified, and tabulated for each PSU. For example, PSU number 43 had 7,076 pixels approximately and it was 10 square miles in area. Nine hundred and sixty-one pixels were computer classified as cotton, 136 pixels classified as sorghum, 76 pixels classified as hay, 2,673 pixels as pasture, 2,990 pixels as other uses and 1,240 pixels which were not recognizable. PSU 43 had, on an average, 90 acres of cotton and 265 acres of pasture per square

---

1/ Huddleston, H. F. and William H. Wigton, Use of Remote Sensing in Sampling for Agricultural Data, ISI/IASS, Paper #47.

mile. Similar data is available for all other PSU's. Hence, other PSU's with similar agriculture can be grouped in homogeneous strata and PSU's that are different can be readily separated from them. Although total ground data is not available, the computer data is correlated with what is on the ground; thus it can be used to make sampling more efficient.

To obtain the maximum reduction possible through stratification, the primary sampling units were assigned to four (4) strata based on the square root of the pixel count for cropland and the two principal individual crops. 1/

For the stratification variable total cropland pixels, the reductions in variance were 27 percent for cotton and 35 percent for sorghum. When the stratification was based on individual crops, the reduction was 60 percent for cotton and 58 percent for sorghum. Since cropland pixels are likely to be constant over years for the PSU's stratification based on this variable should have lasting benefits. However, stratification on an individual crop may not be as effective the second year, since individual crops may change from one year to the next. If current crop year data is available before harvest, then it is possible to use satellite data for post-stratification or in the estimation more directly with regression estimators (Refer to section on Acreage Estimation).

Finally, supplementary information can be used as a size variable. One could assign sample units to PSU's based on total pixels. It is possible to obtain substantial gains following this strategy also. 2/

1/ Cochran, W.G., Sampling Techniques. Second Edition, P. 129-130.

2/ Huddleston, H. F., and William H. Wigton, Use of Remote Sensing in Sampling for Agricultural Data." U.S.D.A.-SRS 1975.