Area Estimates by LANDSAT: Kansas 1976 Winter Wheat

Crop
Reporting
Board
New Techniques Section
Research and Development Branch

Economics, Statistics, and
Cooperatives Service
U.S. Department of
Agriculture

Area Estimates by LANDSAT: Kansas 1976 Winter Wheat

by

Michael E. Craig

Richard S. Sigman

Manuel Cárdenas

New Techniques Section

Research and Development Branch

Statistical Research Division

Economics, Statistics, and Cooperatives Service

U.S. Department of Agriculture

Washington, D.C.

August 1978

Table of Contents

I. INTRODUCTION

This paper describes the results of research done in Kansas related to the estimation of winter wheat by the Economics, Statistics, and Cooperatives Service (ESCS) in 1976. The goal of the project was to utilize data gathered by the LANDSAT satellite to improve existing winter wheat estimation procedures at the state, multi-county, and individual county levels. Existing ground surveys, especially the June Enumerative Survey (JES), provide crop hectarage estimates with measureable precision at national and state levels. The ESCS approach to utilizing LANDSAT data is to use it as an auxiliary variable with the JES ground data being the primary variable [1].

The following phases of the 1976 Kansas Winter Wheat project are described in this report:

1). Agriculture in Kansas,

2). Ground data collection and editing,

3). LANDSAT data acquisition and management,

4). Analysis procedures, and

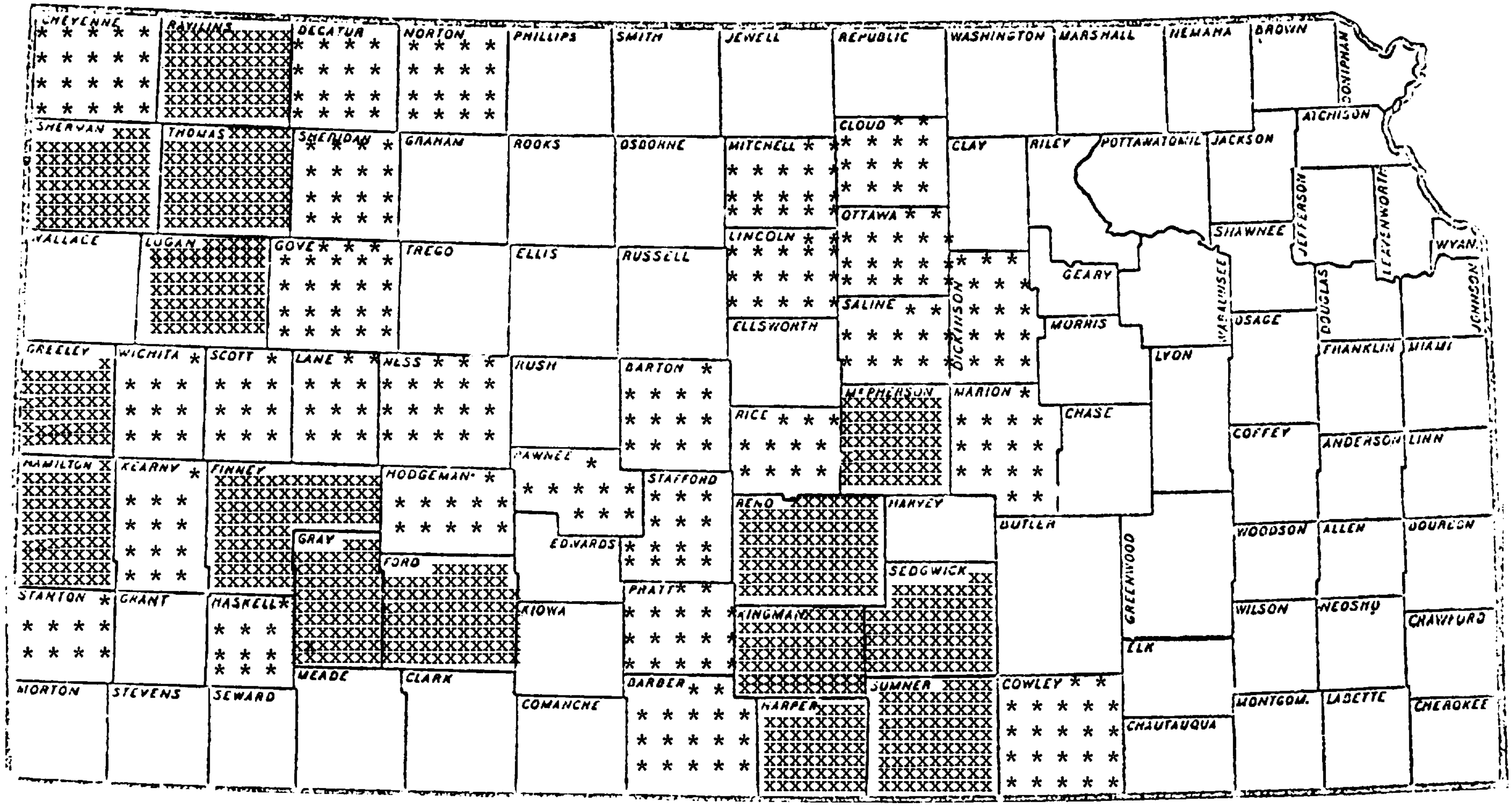
5). Analysis results and comparisons.

II. AGRICULTURE IN KANSAS

Kansas is the number one ranking state nationally in winter (and all) wheat planted, harvested, and produced. In 1976, Kansas ranked fourth in area planted to principal crops with 8.8 million hectares. Over half of this hectarage, some 5.2 million hectares (see Figure 1 for distribution), was planted to winter wheat. Final harvested hectarage from the 1976 crop was 4.5 million hectares and produced 9.2 million metric tons, third largest production ever for winter wheat in Kansas.

Other major crops for Kansas include sorghum, corn, rye, barley, soybeans. oats, alfalfa, and other hay. For this study, crops or land uses of interest were only those that can be spectrally confused with winter wheat. This restriction excluded spring planted crops (soil still bare in spring imagery). Thus, the possible confusion crops were alfalfa, hay, barley, rye, and pasture (see Figure 2).

Rye and barley hectarages in Kansas (32 and 35 thousand hectares, respectively) were small compared to winter wheat hectarage. Hence, rye and barley were not significant as confusion crops. This left alfalfa and other hay as the major confusion crops during the period of study.



xx 13 million and over kilograms

** 95.2 to 136 million kilograms

Figure 1: Distribution of Wheat Production in Kansas by Counties

Figure 2: Kansas Crop Calendar

Crops	Usual Planting Dates	Usual Harvesting Dates		
		Begin	Most Active[a]	End
Barley:				
Fall	Sept. 10 - Oct. 25	June 10	June 15 - July 1	July 5
Spring	Mar. 5 - Apr. 30	June 20	June 25 - July 1	July 10
Corn:				
Grain	Apr. 15 - June 10	Sept. 15	Oct. 10 - Nov. 5	Dec. 5
Silage	Apr. 20 - June 20	Aug. 25	Sept. 1 - Oct. 1	Oct. 10
Hay:				
Alfalfa		May 10		Oct. 30
All		May 25		Sept. 10
Oats	Feb. 25 - May 1	June 25	June 30 - July 10	July 20
Sorghum:				
Grain	May 10 - July 1	Sept. 20	Oct. 10 - Nov. 10	Dec. 1
Silage	May 10 - July 1	Sept. 5	Sept. 10 - Oct. 10	Oct. 15
Soybeans	May 10 - July 5	Sept. 20	Oct. 1 - Nov. 5	Nov. 20
Winter Wheat	Sept. 10 - Oct. 25	June 15	June 20 - July 5	July 15

*Excerpt from Agricultural Handbook No. 283, USDA, Economics, Statistics, and Cooperatives Service.

III. GROUND TRUTH MANAGEMENT

A. Data Collection

Published wheat hectarage estimates at the state and national level are based in part on the JES, a sample survey that utilizes area-frame sampling. The design of the JES is that of a stratified cluster sample (see Appendix D and reference [2]). The clusters (referred to as segments by ESCS) are land areas consisting of several farms or parts of farms. These segments were used as ground-truth information for applying LANDSAT data estimation procedures.

From a total sample of 435 Kansas JES segments, 87 were subsampled for the LANDSAT project. Another 87 were available from segments rotated out of the JES after the 1975 survey making a total of 174 segments to be used for LANDSAT analysis. This number of segments subsampled was determined to reduce the impact of the LANDSAT research on the 1976 JES data collection effort.

Each set of 87 segments contained two replications each from strata 11, 12, 20 and one replication each from strata 31, 32, and 40 (land use strata definitions are given in Appendix A). For the LANDSAT analysis it was decided to study only the major agricultural strata (11, 12, and 20) because the subsample contained very few segments in the urban and rangeland strata (31, 32, and 40). Thus, the size of the subsample was reduced from 174 to 156 (see Table 1).

Table 1: Kansas Segment Allocation

Strata	Population No. Segments	Seg. Size (Sq. Miles)	JES Sample	LANDSAT Sample
11	25058	1.0	170	68
12	21732	1.0	120	48
20	21284	1.0	100	40

The enumerators collected segment data on forms designed by the New Techniques Section with assistance from the Kansas State Statistical Office (SSO). Field boundaries were drawn on the black and white ASCS photos of the segments. Training schools were held on the use of these forms and photos. Enumerated data were collected on two visits to each segment. The first visit, called the April visit, was made during the period from April 12 to May 3, 1976. The second visit, called the June visit, was made during the period from May 21 to June 21, 1976. For fields in subsampled segments, enumerators collected such items as total field and crop area, crop or land-use cover, intended uses of crop fields, field appearance, and data of harvest (see Appendix B).

To assist with the interpretation of ground-truth information low level color infrared (IR) aerial photography of the subsampled segments was taken and prepared by the Remote Sensing Institute of the South Dakota State University. These photos were developed at a scale of 5.25 inches to a mile. The photo acquisition flights over the segments occurred during the period from May 1 to May 8, 1976.

B. Data Edit

As soon as both the ASCS and the color IR photos were received, field, tract, and segment boundaries were transferred to the color IR. These boundaries and field numbers were transferred as reported by the enumerators with no attempt made to interpret them. There were 11 segments with unusable or missing color IR photos.

After editing some of the data, the large amount of time required to correct all field boundaries necessitated restructuring the edit to label only the wheat fields, with all other fields called "other."

Field cover type and boundaries were photo-interpreted on the color IR and compared to enumerator data. Inconsistencies between the IR and reported data were rectified. Area and appearance data were then coded and keypunched. Using county maps with JES segments located on them, the segments were located and drawn on USGS quadrangle maps.

A field determination using the color IR photography was made with a computer process called digitization. This process related field boundary coordinates to a map base (the USGS maps), from which very precise area measurements were available for individual fields. A discussion of the software package used is given in [4].

The coded ground data was then merged with the digitized area determinations to make field level records containing (both April and June) reported and digitized area, field appearance codes, strata, segment number, and dates of visits. Checking the ratio of reported to digitized area at the field level was done along with comparing the total digitized segment area to the planimetered area given by the JES master record. Any discrepancies were checked and updates made as needed to get a final data set. This data set was used to create ground-data files for analysis.

IV. LANDSAT DATA ACQUISITION AND MANAGEMENT

A. Characteristics of LANDSAT Data

The basic element of LANDSAT data is the set of measurements by the satellite's multispectral scanner (MSS) of a .4 hectare area of the earth's surface. The MSS measures the amount of radiant energy reflected and/or emitted from the earth's surface in various regions (bands) of the electromagnetic spectrum. The LANDSAT II satellite used by this study has four bands; one green, one red, and two near-infrared bands.

The individual .4 hectare MSS resolution areas, referred to as pixels, are arrayed along east-west running rows within the 185 kilometer wide north-to-south pass of the LANDSAT satellite. A given point on the earth's surface is imaged once every eighteen days by the LANDSAT II satellite. Satellite passes which are adjacent on the earth's surface are at least one day apart with respect to their dates of imagery.

Satellite passes are cut into scenes, which are just strips of LANDSAT data covering a length of 185 kilometers (same as the width). Adjacent scenes in the same pass overlap several hundred scan lines. Adjacent scenes east-to-west overlap approximately one one-third of the columns.

B. Scene Selection

In order to cover the state of Kansas with LANDSAT imagery, six satellite passes were required. Coverage is composed of five passes of three scenes each and one pass consisting of only one scene (to cover the southeast tip of the state).

It was felt that the separation of other land uses from wheat would be best in early spring imagery. Hence, the first criterion for selection of LANDSAT imagery was the optimum period for images which was beleived to be April or May. The second criterion considered the machine quality of digital data over all four bands. Third, the presence or absence of clouds was considered in the selection.

Cloud cover presented a definite problem [8]. Four passes were available which were nearly cloud free. For another pass, two counties were lost due to a small cloud covered area. The remaining pass (over central Kansas) had no cloud free scenes for the period required (either LANDSAT I or LANDSAT II). Two partially cloud covered scenes on one date were used to cover a seven county area found to be cloud free in this pass. See Figure 3 and Table 2 for final scenes used for coverage. Individual scenes were labeled by pass number and position (north, middle, south) in a pass.

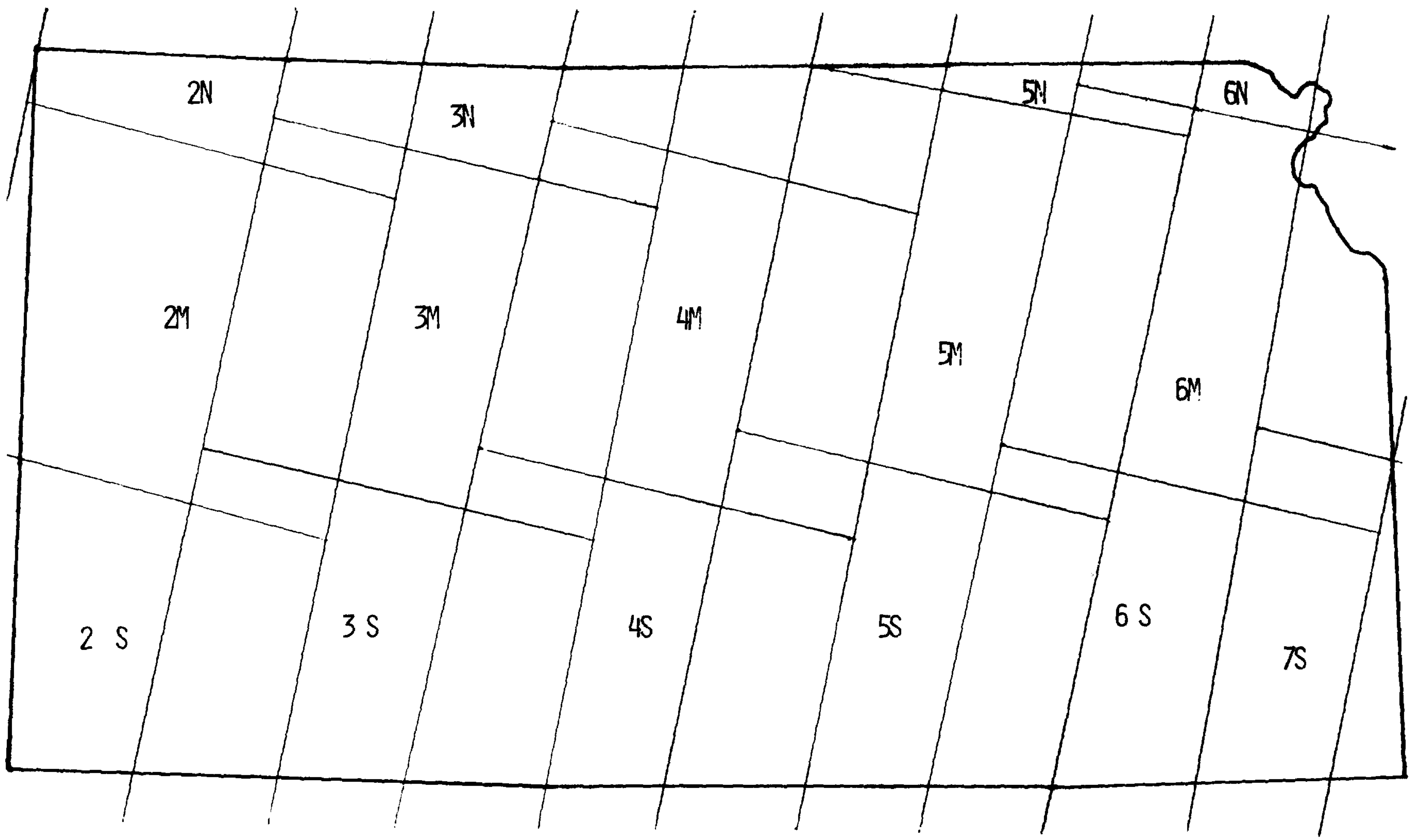


Figure 3: Kansas LANDSAT Coverage

Table 2: LANDSAT II Data, Kansas 1976

Scene	Date	LANDSAT ID-Number	Comments
2N	4/1/76	2435-16404	Clear
2M	4/1/76	2435-16410	Clear
2S	4/1/76	2435-16413	Clear
3N	5/6/76	2470-16335	Clear
3M	5/6/76	2470-16342	Clear
3S	5/6/76	2470-16344	Clear
4M	4/17/76	2451-16291	Heavy Clouds
4S	4/17/76	2451-16293	Heavy Clouds
5N	4/16/76	2450-16230	Clear
5M	4/16/76	2450-16232	Clear
5S	4/16/76	2450-16235	Some Clouds
6N	5/3/76	2467-16165	Clear
6M	5/3/76	2467-16171	Clear
6S	5/3/76	2467-16174	Clear
7S	5/20/76	2484-16113	Clear

An inspection of one pass showed a visible edge separating light pixels in the middle (3M) and dark pixels in the southernmost scene (3S). This edge (or front) ran in a diagonal fashion across the two scenes and was believed to be caused by wet versus dry soil. A possible explanation for this difference was a large rain front over the wet-looking area the day before the imagery. This difference tended to confuse the classification of wheat and other crops between the two areas within the same pass. Healthy wheat fields in the "dry" area looked similar to abandoned wheat or waste fields in the "wet" area.

C. Registration and Segment Calibration

Registration relates LANDSAT row/column coordinates to map based latitude/longitude by means of mathematical equations called affine transformations. These equations allow prediction of specific points on maps to corresponding pixels and vice-versa. Registration of the 15 scenes picked for Kansas analysis to a map base was done using corresponding points found on 1:500,000 scale LANDSAT paper products and on USGS quadrangle maps.

Segment calibration is a local movement of the predicted segment area to a more exact location as determined by field patterns in the segment. These field patterns were found in the LANDSAT data by use of computer generated grayscale printouts. A grayscale is a picture-type line printer product where each printed character represents a pixel. For raw data grayscales, the printed character represents the amount of emergy reflected in the specific light band. A categorized grayscale has numbers representing the category into which the pixel was classified.

Segment calibration was done in two ways. First, grayscales from raw data bands were made of the predicted segment area plus a boundary of 20 pixels in width for each segment. Using the digitized segment files, plots of the segment were made at the same scale as the grayscale prints (see Figures 4 and 5 for an example of segment grayscale plots). Starting from the predicted position of the segment, the plot is overlayed on the grayscale and moved until the field boundaries on the plot best fit the field patterns of light and dark pixels on the grayscale. The new coordinates of the segment, if it needs to be moved, are entered into a local calibration file which will supplement the precision registration when required.

A second way of checking segment location was to use the same procedure as above, with a categorized grayscale. Each pixel in a categorized grayscale had been given a category number based on a preliminary clustering or classification of the segment area and using all 4 LANDSAT bands (not just a single band as above). Thus, field patterns were found that were sometimes hidden in the other type of grayscale. This method did take more resources and if distinction was good in the original grayscales may not have been worth the extra time and effort.

Bands 5 and 6 were most helpful in finding wheat fields and their boundaries. The registration accuracies prior to calibration were very good with many segments needing no movement at all and the rest were generally moved less than 2 pixels in either direction.

Large field sizes (see Table 3) and rectangular field shapes are common in Kansas. These characteristics greatly facilitated the location of "ground-truth" pixels in the LANDSAT data.

Table 3: Kansas Average Wheat Hectares [3]

Agriculture Strata	Per Segment	Per Field
11	97.1	22.7
12	85.0	16.3
20	46.5	13.7
(11,12,20)	80.5	18.0

Figure 4: Segment Grayscale

Segment = 5048 Channel 3, Scene 4M, Band 6;

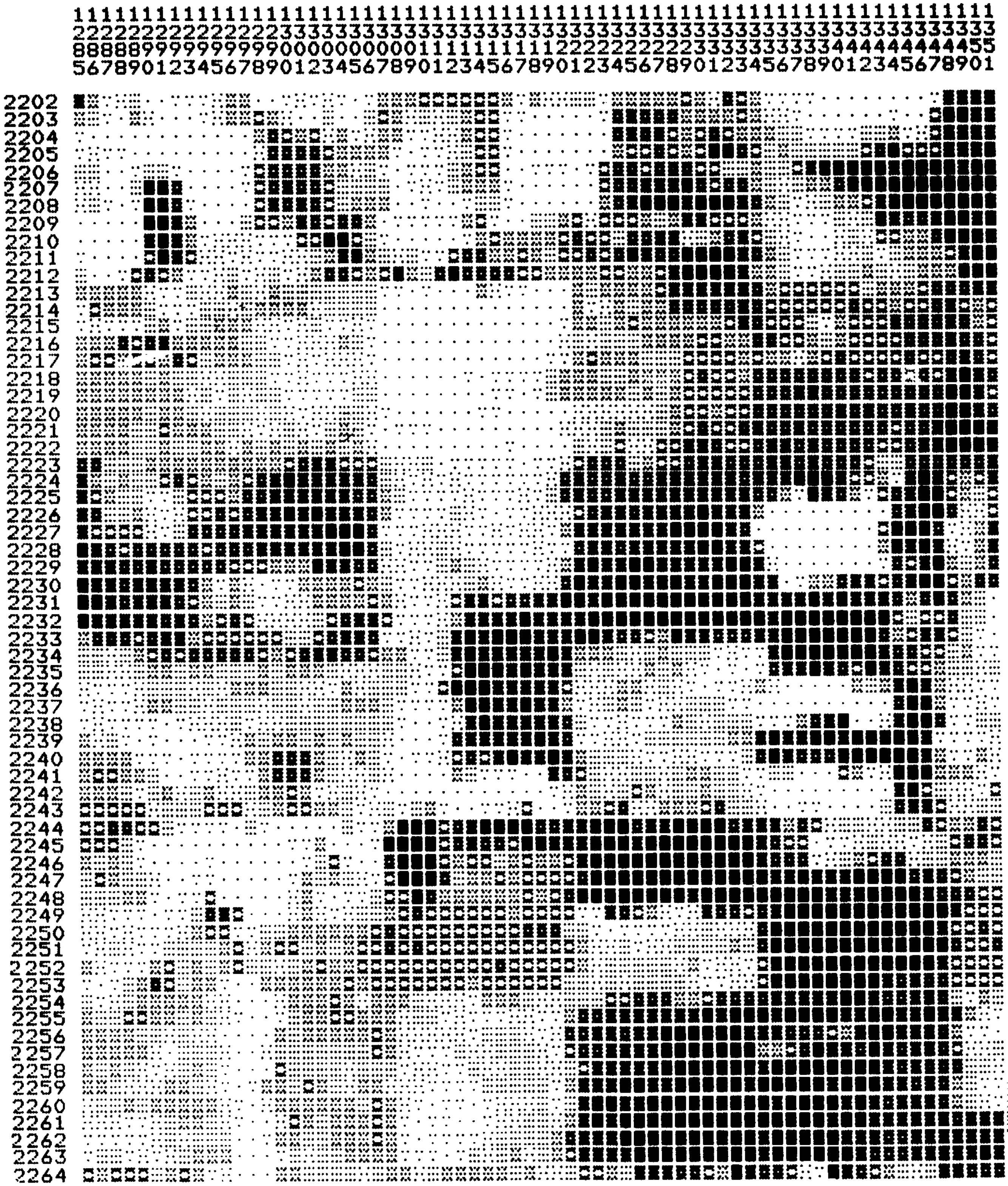
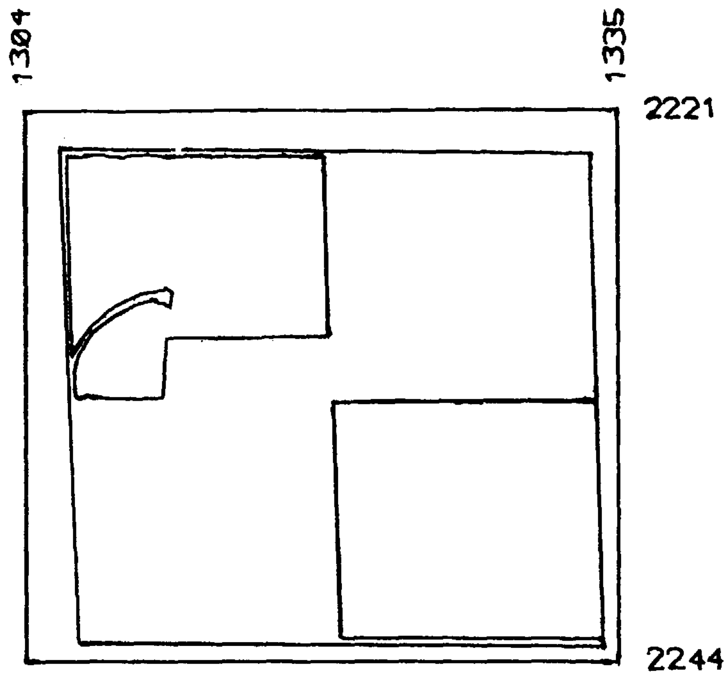


Figure 5: Segment Plot



Segment 5048, Part 1 of 1. 3 Fields 34 Edges 33 Vertices

Total Area 643.4 Acres Seg. Scale 1: 11885

V. ANALYSIS PROCEDURES AND RESULTS

A. Definition of Analysis Districts

One characteristic of LANDSAT data is that it does not consider political boundaries when taking imagery. Thus, the state was divided into analysis districts which were determined by LANDSAT boundaries and not comparable to ESCS's Crop Reporting Districts. An analysis district is a group of counties or parts of counties that is wholly contained in a LANDSAT pass. Estimates for these multi-county areas were made and then individual county estimates were derived from them.

County maps with land-use strata marked on them were ditigized to a latitude-longitude coordinate system (as discussed earlier with segments and fields). The vertices of the outer boundaries of counties were then transformed to the row-column coordinate system of LANDSAT data using each individual scene's registration. These coordinates were then viewed and counties were assigned to LANDSAT scenes (see Figure 6 for final analysis districts for Kansas). Counties were assigned to one and only one pass (and its corresponding analysis district). There were 14 counties with no LANDSAT data because of clouds and 4 counties lost due to lack of training data. Counties without LANDSAT data or with other problems that made them unusable were lumped into one analysis district called Pass-4C.

B. Split Counties

In addition to the 18 counties lost due to clouds or no training data, 13 counties were found to be split across scene boundaries within 4 of the passes. In earlier experiments (see [1]) when this situation was encountered, psuedo-frames were constructed by putting together the bottom part of the northern scene and the top section from the southern scene thus creating a new frame. This method was only valid when the scenes were from the same date and when counties were cut north or south by LANDSAT lines and not east or west by LANDSAT column boundaries. Another drawback of the psuedo-frame approach was that it requires registration of the new scenes.

A new, quicker method was created to handle these split counties whether they were divided by lines or columns. The new approach, called the psuedo-county approach, was to digitize a figure that divided a county into two (or more) parts, or sub-counties, that were each completely within one LANDSAT scene, utilizing the fact that scenes partially overlap. This figure was then used to cut

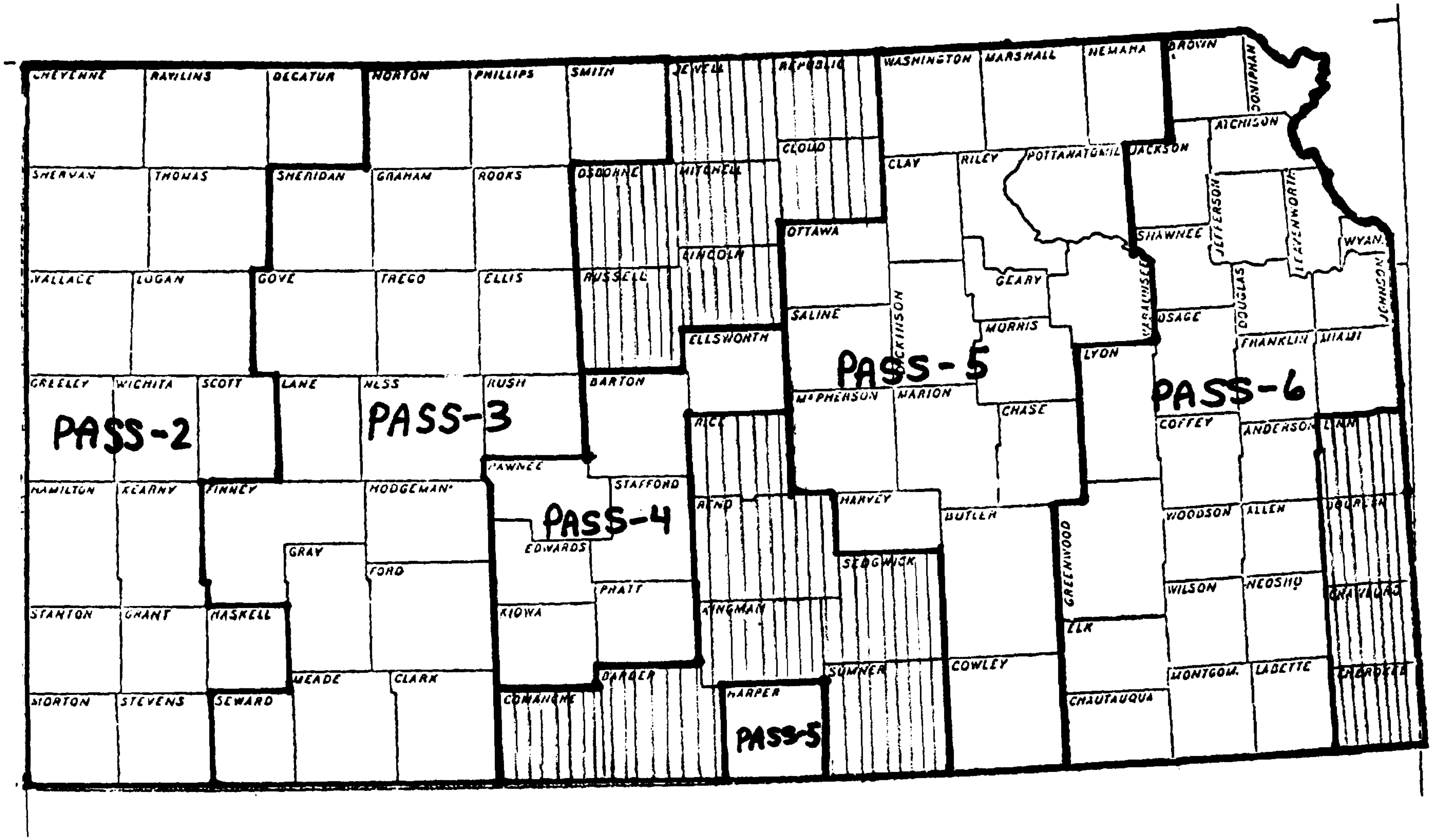


Figure 6: Final Analysis Districts - Kansas 1976

up the original digitized county file into parts called psuedo-counties. Each psuedo-county was distinct from all others and thus was estimated as were the non-split counties. In Kansas, only one county was split across analysis districts and it happened to lie partly in the one-scene pass (Pass-7) that did not have enough training data and so it was not used. The other counties were split across scenes within the same pass and thus the only adjustment to the estimation process was to sum the wheat pixels by strata for each county's parts. For estimation when the psuedo-counties for a given county are in different analysis districts, each psuedo-county would be considered a separate county all the way through the actual estimation and would require adjustment in the number of area frame units for each analysis district. This did not occur in the Kansas study.

After the analysis districts and split counties in Kansas were all decided upon, the segments were labeled by analysis district also. See Table 4 for segment and county allocation to the various passes (districts).

Table 4: The Number of Segments and Counties by Analysis District

Analysis Dist.	Sample Segments By Strata			Sample Segments 11-12-20	Frame Units By Strata			Number Counties
	11	12	20	TOTAL	11	12	20	
Pass-2	21	5	3	29	8067	2681	2189	17
Pass-3	11	12	12	35	4678	5295	5264	19
Pass-4	9	2	0	11	2750	1047	739	7
Pass-5	9	13	9	31	3579	3497	3961	19
Pass-6	2	5	9	16	1518	4520	5478	25
Pass-4C	28	29	15	72*	4470	4696	3652	18

* Total JES Sample Used for Direct Expansion

B. Pixel Clustering and Classification

Separate analyses were conducted for each analysis district using various clustering and classification procedures. For general information on classification and clustering see Appendices C and D. Various factors affecting classifier performance are discussed in the Illinois report [1]. For this Kansas project, the pixel classifier for each pass was based on training data from that specific pass only. Initially each cover type was clustered into distinct groups or categories and calculations made of the signature* means and covariance matrix for the training set of labeled pixels defined by the digitized segments. Signature statistics for several categories or cover types were grouped together in statistics files used with classification software. Different clustering attempts for each cover type were made.

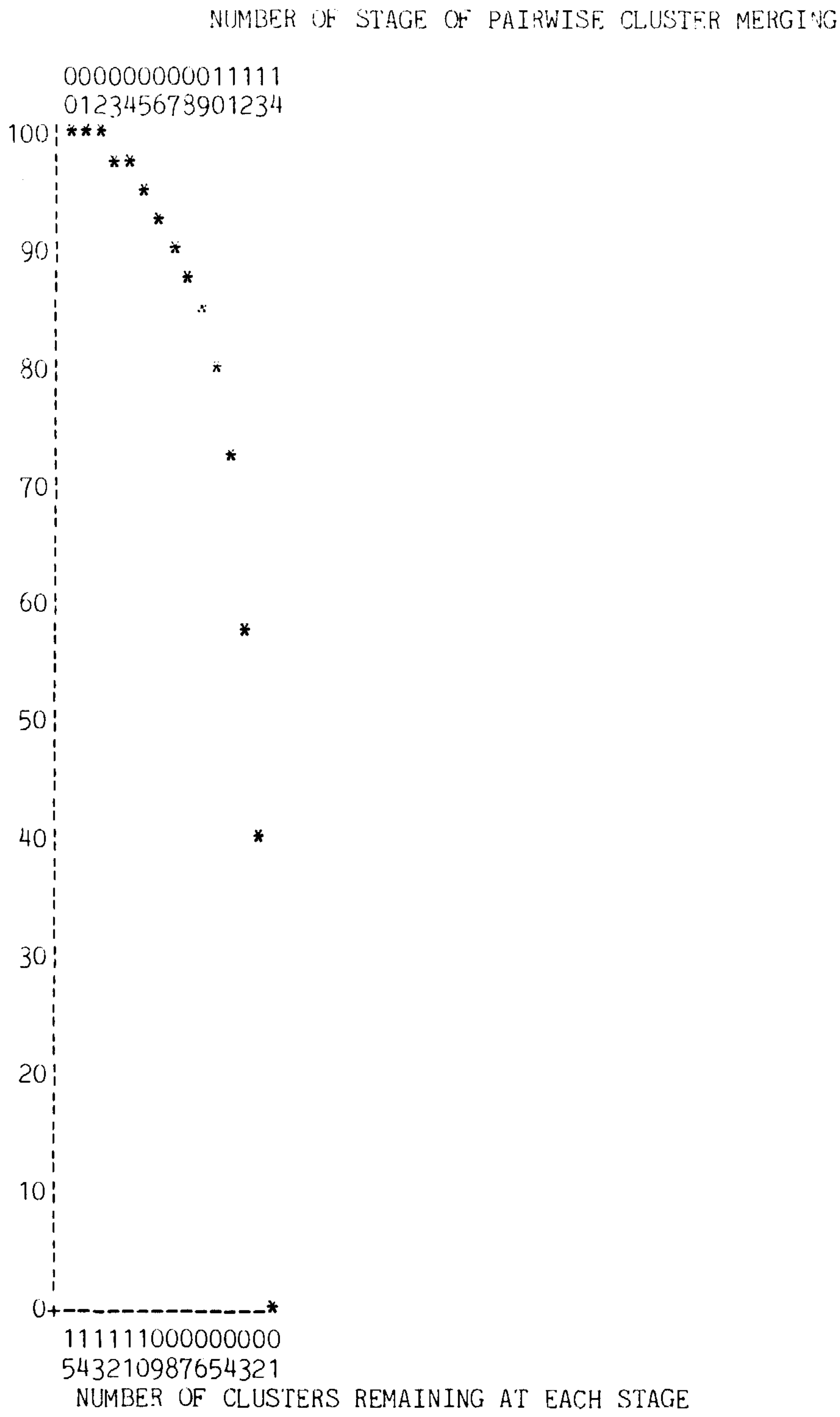
One approach tried was to set the minimum number of categories to a large number (say 12-15 per cover type) and let the clustering algorithm find the best set of that number of categories. This set of categories (now signatures rather than groups of pixels) was then used in a grouping algorithm [6] which merged signature means and variances one at a time from the total number of categories down to one overall cover type category. The merging criterion in the algorithm was a minimum pattern-transmission loss function (an example is shown in Figure 7). Using the loss function graph, the final number of categories was determined considering the natural breaks and an 80-90 percent pattern retention. In the example shown, 5 or 6 final categories could be a likely choice.

Another approach used was to set a small minimum number of categories and let the clustering algorithm find the best set of these, deleting any that seemed to be outliers. The two approaches came up with much the same categories although clustering only without grouping seemed to give slightly tighter clusters. Grouping helped to give an idea of the number of categories desired whereas clustering by itself required predetermination of numbers of categories.

---

* Signature refers to the mean vector and covariance matrix for a specific cover type or category and ideally is separable in the four dimensional LANDSAT Scanner space from other categories.

Figure 7: Example of Pairwise Cluster Merging Tree



PATTERN TRANSMISSION LOSS FUNCTION FOR CLUSTER ANALYSIS

(PERCENT PATTERN TRANSMISSION RETAINED)

Several methods were used to compare the different clustering and grouping approaches. One method, called a scattergram, shows pixels values labeled by category in a two-dimensional graph (two of the four bands must be chosen). Scattergrams show each pixel's value and may become messy when categories overlap. Another method used for signature comparisons was to plot concentration ellipses (again picking two bands). These ellipses were actually just two-dimensional 90 percent, confidence intervals computed from statistics files containing the signature means and variances. These two methods allowed visual comparisons of the amount of confusion (seen as overlap) between categories. Appendix E contains ellipse plots for the final signatures (statistics files) used for classification. Two other methods of comparing the signatures involve classification and estimation are discussed later in this report.

Table 5 gives the final number of categories per cover type used in this study for large scale classification. On three passes (2,3, and 5) the number of wheat categories needed came out five and so in later passes (4 and 5) the wheat training data was clustered directly into five categories. The number of categories needed for "other" was more variable by pass ranging from 4 to 7.

Table 5: Number of Categories Per Cover Type

Analysis District	Wheat	Other	Total
Pass-2	5	4	9
Pass-3	5	7	12
Pass-4	5	7	12
Pass-5	5	5	10
Pass-6	5	5	10

The final statistics file was used to create a set of discriminant functions for classification of LANDSAT pixels. Usually, each analysis district had one and only one statistics file no matter how heterogeneous the data may have been across a county, scene, or pass. This classification was done at two levels, one was to classify only those pixels interior to the sampled segments, the other was to classify entire LANDSAT scenes using the same statistics file. The segment level classifications were used to test the performance of the classifier and for estimation of regression parameters. The large scale (entire scene) classifications were used for the actual acreage regression estimates at district and county levels.

After examining the visible differences in Pass-3 (discussed earlier), it was decided that another level of classification was needed to allow more than one statistics file per scene. This classifier would take into account completely different signatures for the various covers as a function of location in the scene. New software was written [1/] to apply this classifier and also allow its results to be used in estimation. Two statistics files were then created (although both have the same number of categories per cover as in Table 5, the application of the new classification method to estimation did not require this) for Pass-3 by splitting up the set of segments into two parts, then following the normal clustering procedure. Table 6 shows the actual pixels classified by the analysis districts.

Another factor involved in the classification of pixels is the use of "different prior probabilities" (different weighting factors for the likelihood functions in the discriminant functions). The priors used were unequal priors proportional to expanded digitized acres (called PED) and equal priors (called EP). For a given analysis district the PED prior probability for a specific cover was defined as the ratio of the current year direct expansion estimate to the total land area in the region.

Some idea of the performance of the classifier may be obtained from the percent correct, that is, the percentage of the digitized segment information that was classified correctly. Since the classifier was trained and tested on the same data (called resubstitution) the numbers may be somewhat optimistic. One drawback of the percent correct as a measure of classifier performance is that it takes into account only the errors of omitting pixels of a specific cover and does not measure the error of calling a pixel that specific cover when it is actually another cover type. Table 7 contains percents correct for the final statistics files used in this project, with both unequal priors (PED) and equal priors (EP). Notice that the unequal priors statistics file was used for large scale classification in Pass-6 only, even though PED tenas to increase the percent correct. The final criterion for picking a statistics file will be discussed in the Estimation section.

1/ This software was supplied by Martin Ozga and Walter Donovan of the Center for Advanced Computation at the University of Illinois.

Table 6: Pixel Classification Table by Strata and Analysis District

Analysis District	Cover Type	11	12	20	31-62	Total Over All str.
Pass-2	Wheat	1542556	479453	294008	90795	2406812
	Other	3015429	1028932	951584	803402	5799347
Pass-3	Wheat	832104	823850	574361	169791	2400106
	Other	1828384	2177453	2378133	1110488	7494458
Pass-4	Wheat	691804	214542	97788	29275	1033409
	Other	864889	377724	324671	356944	1924328
Pass-5	Wheat	883916	676351	457624	171630	2189521
	Other	1156347	1297593	1799603	2310827	6564420
Pass-6	Wheat	121051	269503	217076	91486	699116
	Other	745203	2303272	2905911	1645465	7599851

Table 7: Percent Correct Pixel Classification for Segments

Analysis District	ED Priors			PED Priors		
	WHEAT	OTHER	OVERALL	WHEAT	OTHER	OVERALL
Pass-2	86.83	89.38	88.73	80.19	94.61	90.97
Pass-3	71.73	87.07	82.81	*	*	*
Pass-4	77.28	73.43	75.10	84.74	78.99	81.49
Pass-5	70.07	84.70	80.10	64.93	88.78	81.29
Pass-6	62.45	79.12	77.76	43.48	97.17	92.79

* Not calculated

C. Estimation Procedures

1. The Regression Approach

As mentioned earlier, ESCS uses LANDSAT data as an auxiliary variable in a regression procedure. In past estimation projects where more segments per analysis district were available, a separate regression equation was estimated for each land-use stratum. The statistical methods involved with the separate regression have been described in the paper by Sigman et al [5], and are excerpted in Appendix D. The technique of pooling strata, used in the Illinois project [1] to alleviate the problem of small sample sizes within stratum was rejected because pooling strata tends to overestimate the variance of the estimate.

A 'combined' regression estimator was developed (see Appendix F) for use with this project. This method assumes that the regression coefficient of the estimator is the same for all strata but the intercepts are obtained from the stratum means. Using the small-scale classifications of the sampled segments the regression coefficients for each stratum in each analysis district (shown in Table 8) were apparently estimates of a common value within each analysis district.

Table 8: Estimated Regression Coefficients for Each Stratum

Stratum	ANALYSIS DISTRICT				
	PASS-2	PASS-3	PASS-4	PASS-5	PASS-6
11	1.1738	1.1785	1.0435	0.9155	1.2140
12	1.1973	1.1132	-	0.6962	1.6004
20	0.9618	1.0929	-	0.3788	1.6604
Combined	1.0648	1.1206	1.0117*	0.7909	1.6206

* Only two data values existed in Stratum 12 and none in Stratum 20.

2. Selection of a Classifier for Estimation

Besides estimating the regression coefficients the small scale classification and estimation provides a measure of the performance of the classifier with respect to the variances of the estimates. The various types of regressions are "separate," "pooling", and the combined strata.

In the "separate" regression, the sample coefficients of determination (r-squared) between digitized wheat acres and classified wheat pixels are determined for each stratum. As shown in Appendix D, maximizing the r-square values minimizes the variance of the regression estimates resulting from a classification. Thus, one criterion used to compare classifier performances on the same strata was the respective r-squares. These values were calculated in all Kansas analysis except for strata 12 and 20 in Pass-4 and stratum 11 in Pass-6. In some analyses districts, however, the small sample sizes per stratum make this figure somewhat unreliable. The pooling of data to derive the classifier relationships in this case assumed only a single stratum was sampled. All segment data from strata 11, 12, and 20 were pooled together and regression was calculated as for an unstratified population. The various r-squared values for the Kansas analysis districts are shown in Table 9 for both EP and PED prior probabilities.

Table 9: R-Square Values by Analysis District and Priors

TYPE & STRATA	Pass 2		Pass4**		Pass 5		Pass 6		Pass 3
	EP	PED	EP	PED	EP	PED	EP	PED	EP
Separate-11	.8516	.7762	.6161	.6398	.8522	.8361	. *	. *	.6719
Separate-12	.9953	.9920	*	*	.4785	.3883	.1454	.9836	.9430
Separate-20	.9965	.9950	*	*	.3962	.5098	.0832	.7429	.7100
Pooling	.8818	.8215	.5975	.5614	.7450	.7181	.1911	.7659	.8073

* Not calculated due to lack of data

** Pass-4 pooling includes strata 11 and 12 only.

Since the major objective of this project is estimation of winter wheat acreages with reduced variances, maximization of percent correct or reduction of the classification error was not considered in the choice of classifiers. Maximization of the r-squared values was the final criterion used for selection of a statistics file to use for large scale classification in a given analysis district. The equal priors (EP) file was selected in all analysis districts except Pass-6. In this analysis district, the classifier tends to classify a large portion of "other" pixels into the wheat categories. Wheat in this area was not a very large crop percentage wise and thus the application of PED priors with small probabilities for wheat tended to give a more reasonable classification and thus better r-squares.

Although in most analysis districts the unequal priors classifier was not chosen for full frame classification, the r-squares found using the PED priors are very close to the corresponding equal priors (EP) values (except in Pass-6). Thus, if the objective of the study was yield computation or some type of stratification based on classified pixels, and not estimation of acreage, the better classifier would be the unequal priors classifier.

3. Large Scale Estimation

Multi-county regression estimates for winter wheat area planted were calculated for the various analysis districts. The regression estimates were compared to estimates calculated by direct expansion of the subsample segments, direct expansion of the total 435 JES segments, and to estimates obtained from the summation of final 1976 county estimates published by the Kansas SSO. The final SSO estimates in Kansas are predominately based on the Kansas State Farm Census. Note that the SSO estimates do not have a calculable variance associated with them because they are based on several non-probability indications, not just the JES direct expansion.

For the multi-strata and multi-county analysis districts, performance of the combined regression estimator was compared to the direct expansion estimator in terms of the relative efficiencies (denoted RE) of the resulting estimates. RE measures the gain, in terms of increased precision, of the combined regression estimate over the respective JES or subsample direct expansion estimate. The equation for calculating the RE follows:

$$RE = \frac{\text{Var (direct expansion)}}{\text{Var (combined regression)}}$$

Table 10 gives the estimated wheat area, coefficients of variation (CV's), and relative efficiencies for all passes with LANDSAT classifications available. Note that the direct expansion estimates shown are based on the subsample chosen for the LANDSAT project.

Table 10: Planted Area Estimates of Winter Wheat for Strata 11, 12, and 20.

Analysis District	Number of Segments	Number of Counties	Estimator	Estimate (hectares)	CV	RE
Pass-2	29	17	Regression	886500	4.9	13.1
			Direct Expansion	876300	18.1	
Pass-3	35	19	Regression	946900	6.7	4.8
			Direct Expansion	1114400	12.5	
Pass-4*	11	7	Regression	382800	7.8	1.3
			Direct Expansion	459300	7.3	
Pass-5	31	19	Regression	876700	5.5	3.2
			Direct Expansion	889800	9.8	
Pass-6	16	25	Regression	358900	4.8	10.6
			Direct Expansion	258500	21.7	
Overall**	122	87	Regression	3488600	2.8	

*Strata 11 and 12 only

** Stratum 20 estimate was prorated from state estimate in Pass-4.

When the relative efficiency was computed for the subsampled segments regression with respect to the whole JES sample, smaller RE were found. Even with this restriction, the regression estimates showed a significant reduction in variance as measured by the relative efficiency, ranging from 1.6 to 4.8 with a RE of 2.7 computed over the 87 county area.

Since some strata were deleted from the classification analysis, "swiss cheese" estimates were computed in order to compare regression estimates with the summations of SSO published county estimates. A swiss cheese estimate consists of regression estimates on the strata included in the classification analysis and prorating the direct expansion estimates of the whole state with respect to area frame units on the strata excluded from the classification analysis. Table 11 gives the pass-level 'swiss-cheesed' estimates for both regression and direct expansion along with the summation of SSO county estimates. The prorated estimate for strata 31, 32, 33, 40, and 50 range from 2.9 percent of the total for Pass-2 to 11.3 percent for Pass-6. The state level estimate uses a direct expansion for Pass-4C.

Table 11: Planted Area Estimates of Winter Wheat for All Strata

Analysis District	Estimator	Estimate (Hectares)	CV	RE
Pass-2	Regression	912900	4.8	13.0
	Direct Expansion	902700	17.6	
	SSO Sum	1035600	-	
Pass-3	Regression	984200	6.5	4.8
	Direct Expansion	1151700	12.1	
	SSO Sum	1106400	-	
Pass-4	Regression	431300	6.9	1.3
	Direct Expansion	507700	6.7	
	SSO Sum	494500	-	
Pass-5	Regression	947500	5.3	3.1
	Direct Expansion	960600	9.1	
	SSO Sum	945800	-	
Pass-6	Regression	404700	4.7	9.0
	Direct Expansion	304300	18.6	
	SSO Sum	382400	-	
State	Regression	5141900	2.7	
	SSO Sum	5220400	-	

* Regression and direct expansion estimators are based on the 'swiss-cheese' technique and use only the subsample segment data for strata 11, 12 and 20.

4. County Estimates

Single county estimates were made on 87 of the 105 Kansas counties. As mentioned before, the unavailability of LANDSAT data (due mostly to cloud cover) prevented the estimation of the remaining 18 counties. Estimates were computed with the "swiss cheese" technique as discussed earlier in the Large Scale Estimation Section. The combined regression estimator as discussed in Appendix F was employed in calculating the estimates for strata 11, 12, and 20. The estimates, their standard errors, and coefficients of variation are listed in Appendix G.

Recently a family of county estimators was developed [7]. County estimates of Kansas winter wheat were computed for strata 11, 12, and 20 using the so-called "ratio" estimator from this family. An empirical comparison of the "ratio" and regression estimates was made. Since the actual county wheat acreage totals were not available, the SSO estimates were used as the "true" totals in the comparison. The county estimates computed by the "ratio" estimator for the total wheat acreage in all strata are also shown in Appendix G together with their standard errors.

Each county in Figure 8 has two numbers associated with it; the first (top) was computed by the formula:

$$\frac{\text{Regression estimate - SSO estimate}}{\text{SSO estimate}} \times 100$$

and the second (bottom) was computed in a like manner with the "ratio" estimate replacing the regression estimate in the formula. The figure actually sheds little, if any, light on which estimator might be superior, however, it does show that the bigger relative differences between both the regression and "ratio" estimates from the SSO estimates occur in counties which are on the border of the state. Also of interest was the fact that both the "ratio" and regression estimates were generally larger than the SSO estimates in the northern counties. The coverse was true for the southern counties. These occurrences were due to the fact that both the "ratio" and regression estimates were highly correlated with the number of pixels classified as wheat which also exhibited these properties. But the reason for such a pattern to be present in the number of pixels classified as wheat was not known and might well serve as a topic for future research.

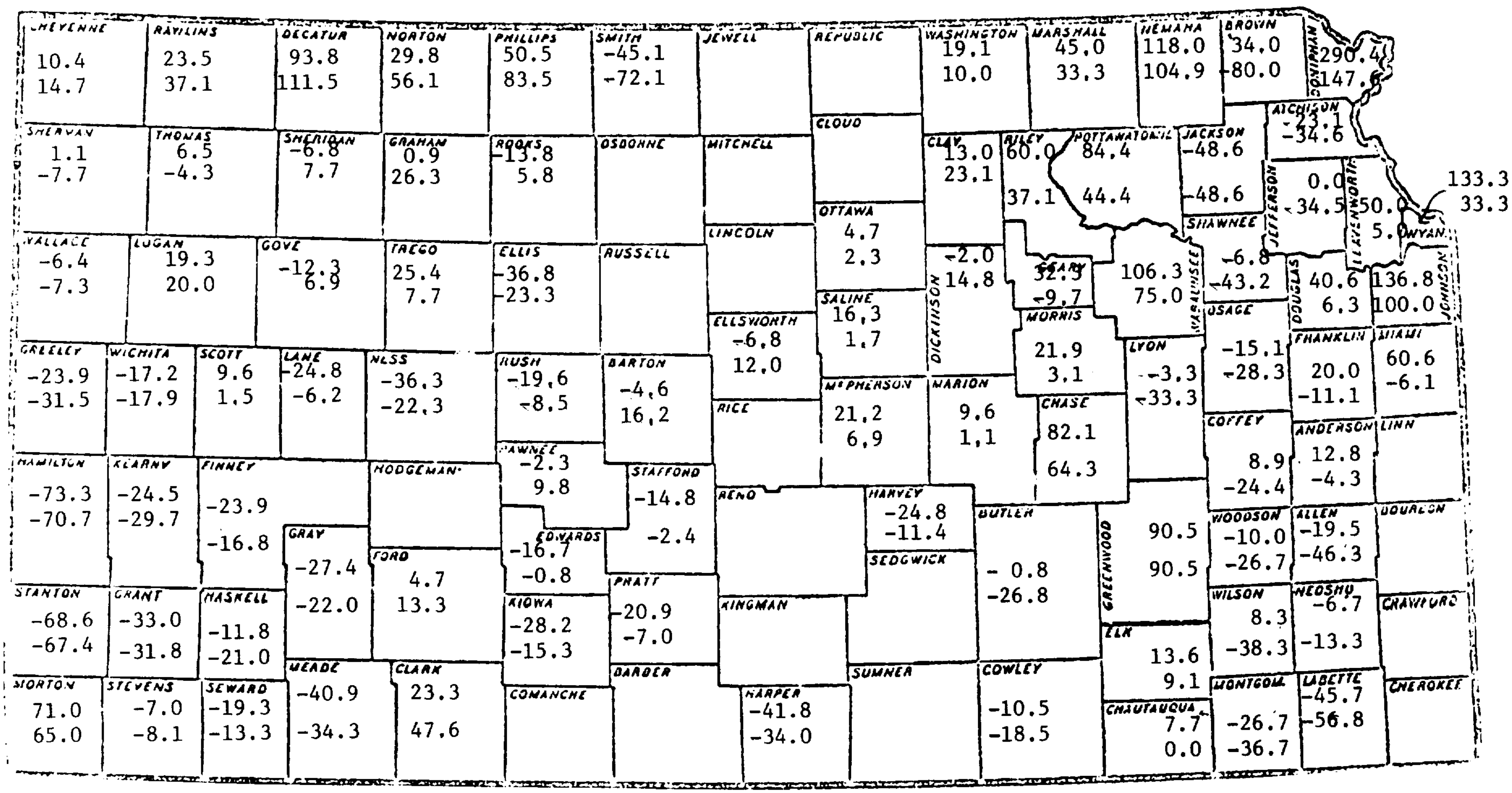


Figure 8: Percent Changes from SSO Estimates; Top Number is from Combined Regression, the Bottom Number derived from "Ratio" estimates.

Since the "ratio" and regression estimates did not exhibit a pronounced difference, we shifted our attention to their coefficients of variation. Figures 9 and 10 show respectively the distribution of the coefficients of variation for the regression and "ratio" estimates of strata 11, 12, and 20. Except for Pass-6, i.e. the eastern part of Kansas, the ratio estimates generally had smaller coefficients of variation. In Pass-6, however, the converse was true. In fact, the smallest coefficient of variation achieved by the "ratio" estimator in Pass-6 was larger than all but two of the C.V.'s attained by this estimator outside Pass-6. This was due, at least in part, to the fact that the pooled within-county variance for the whole state was used to estimate the within-county variance in Pass-6 since there were not enough data values available in Pass-6 alone.

Actually, 67% of the "ratio" estimtes had coefficients of variation less than 20% whereas 48.3% of the regression estimates had C.V.'s less than 20%.

Intervals of one and two standard errors were computed for both the "ratio" and the regression swiss cheese estimates in each county. We then checked to see if the SSO estimates fell within these intervals. Table 13 has the results.

Table 13: Number of counties for which the SSO estimate falls within a one or two standard error tolerence interval for the "ratio" and regression estimates by pass.

Pass	No. of Counties in pass	No. of counties within one-standard error tolerence interval		No. of counties within two-standard error tolerence interval	
		"ratio"(%)	reg.(%)	"ratio"(%)	reg(%)
2	17	8 (47.1)	8 (47.1)	11 (64.7)	13 (76.5)
3	19	8 (42.1)	9 (47.4)	12 (63.2)	16 (84.2)
4	7	4 (57.1)	5 (71.4)	5 (71.4)	6 (85.7)
5	19	7 (36.8)	8 (42.1)	10 (52.6)	14 (73.7)
6	25	9 (36.8)	9 (36.0)	17 (68.0)	10 (40.0)
TOTAL	87	36 (41.4)	39 (44.8)	55 (63.2)	59 (67.8)

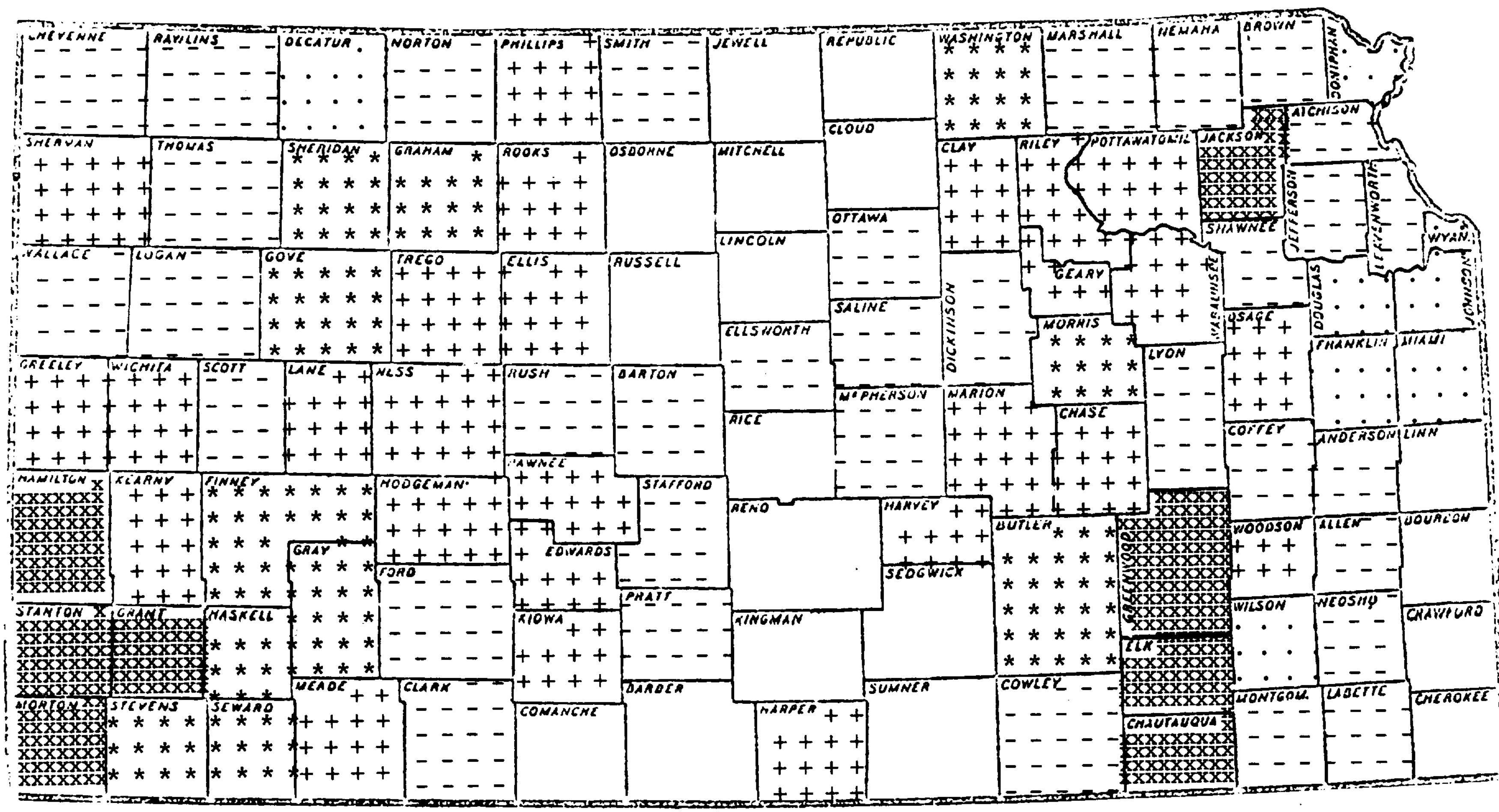
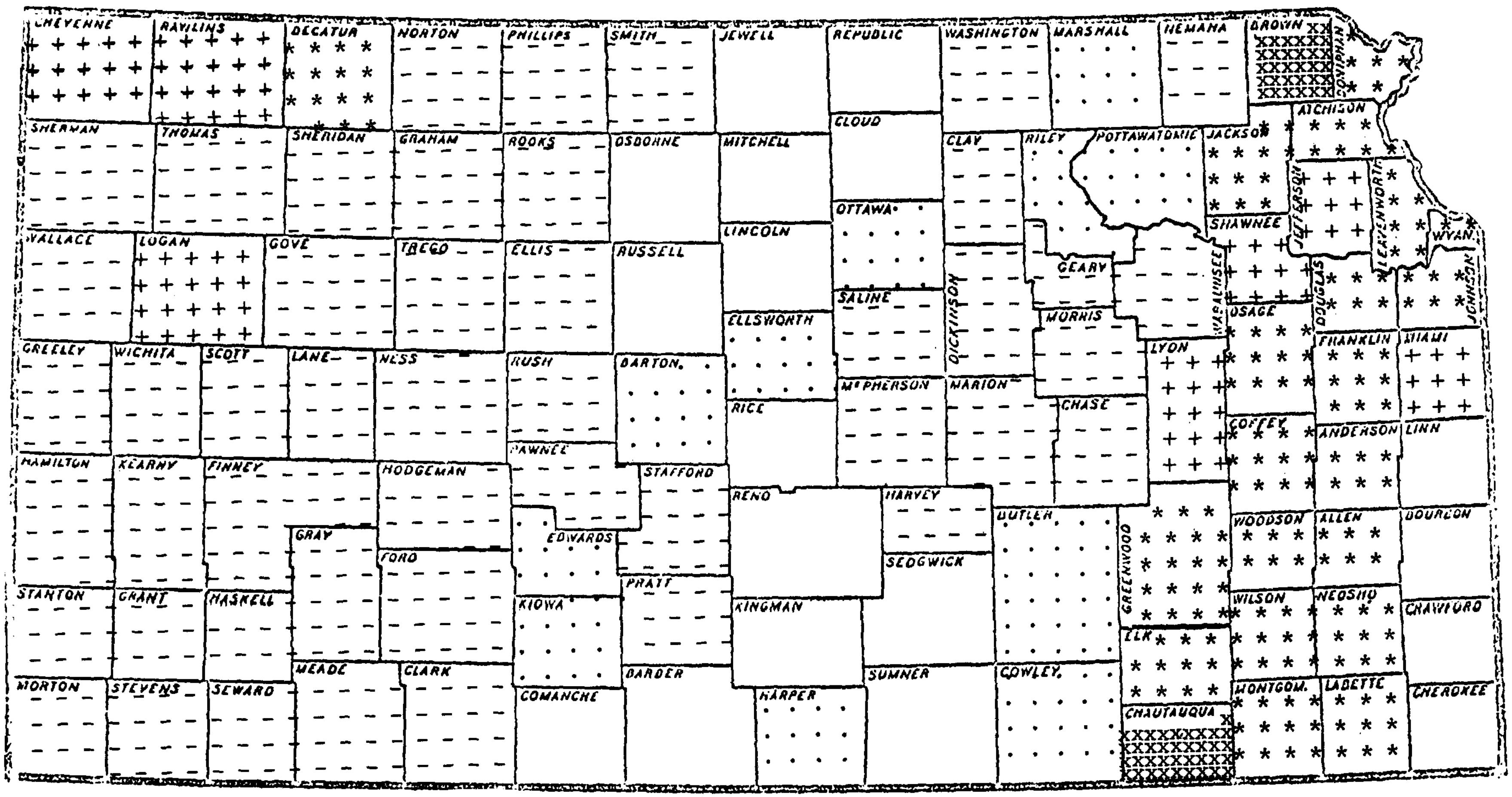


Figure 9: County Level CV's - Regression

. . 0-9.9% CV
- - 10.0-19.9% CV
+ + 19.9-29.9% CV
* * 29.9-39.9% CV
xxx 40.0+% CV
otherwise no estimates were made due to cloud cover



. . 0-9.9% CV

- - 10.0-19.9% CV

+ + 19.9-29.9% CV

* * 29.9-39.9% CV

xxx 40.0+% CV

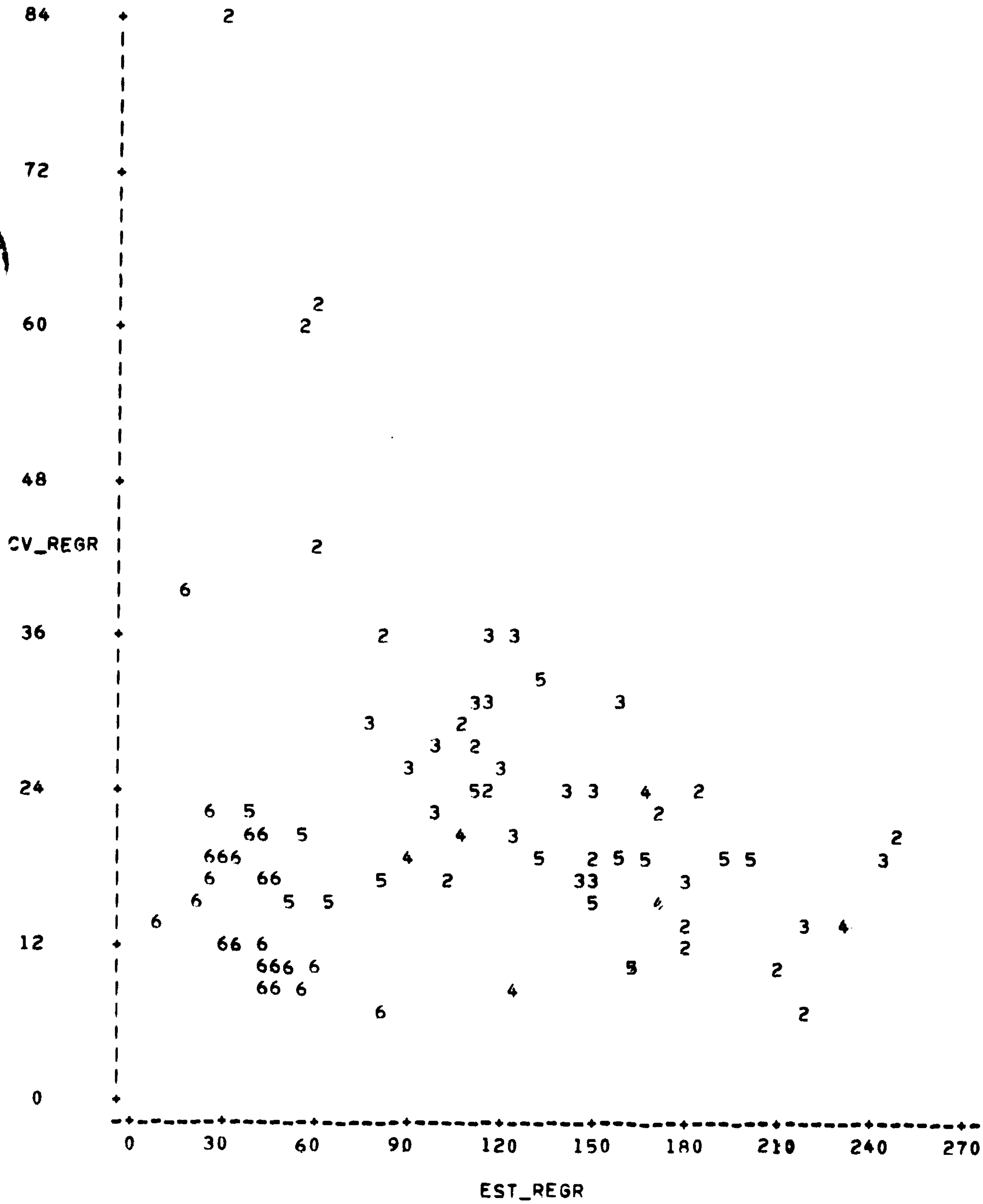
otherwise no estimates were made due to cloud cover

Figure 10: County Level CV's - "Ratio"

The percentage of county SSO estimates falling within a given standard deviation interval was lower for the regression estimates in Pass-6 than in the other passes. Otherwise the percentages were pretty consistent by passes for each estimator (excluding Pass-4 which consists of only 7 counties and a change of one county SSO estimate falling with a tolerence interval brings about a change of over 14 percentage points). The percentages of SSO estimates falling within a one-standard error tolerence interval for both estimators were 41.4 to 44.8%. For a two-standard error tolerence interval the percentage was a little higher for the regression estimator (67.8%) than for the "ratio" estimator (63.2%).

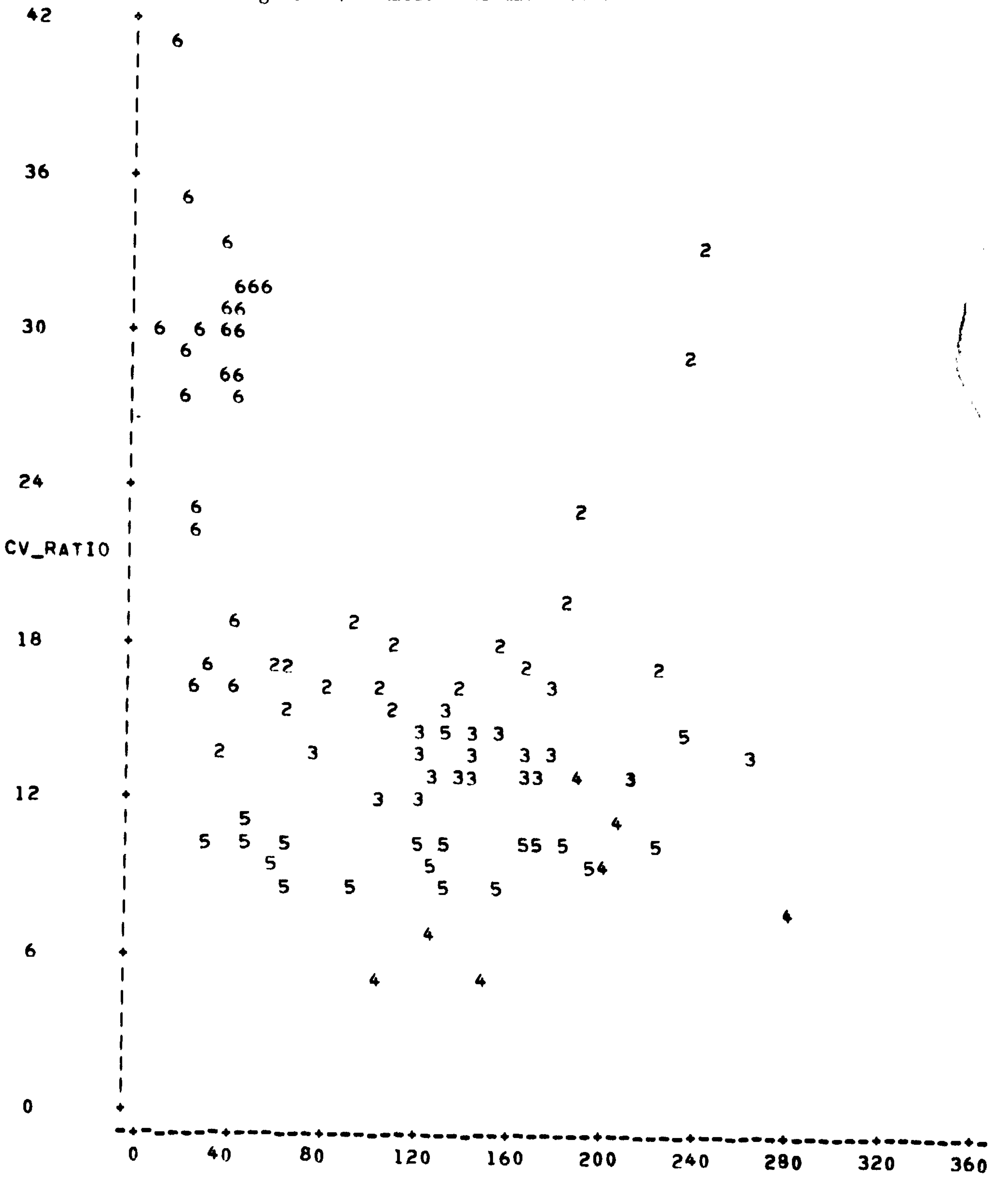
The coefficients of variation were plotted versus the estimate produced by the regression (Figure 11) and "ratio" (Figure 12) estimates. At first glance it appears that for the ratio estimates small C.V.'s are associated with the large estimates and vice-versa. However, upon closer inspection, we see that the size of the estimates and the satellite pass are confounded. The small estimates with the large C.V.'s all belonged to Pass-2. In all the other passes the C.V.'s seemed to be independent of the size of the estimate with the possible exception of Pass-2 which apparently exhibited the opposite trend. On the other hand, for the regression estimator the small C.V.'s generally corresponded to large estimates and conversely. The degree to which this relationship is true varied from pass to pass. Further research is needed to determine all relationships.

Figure 11: Regression Estimates Versus C.V.'s



NOTE: 9 OBS HIDDEN

Figure 12: "Ratio" Estimates Versus C.V.'s



NOTE: 2 OBS HIDDEN

VI. CONCLUSION

The goal of this project was to utilize LANDSAT data to improve existing winter wheat estimation procedures. Winter wheat planted area estimates were made at multi-county and county levels for 87 of the 105 Kansas counties. Eighteen counties were not estimated due to cloud cover or lack of training data.

Attainment of the project goals was measured by the reduction in variance of the area estimate computed using LANDSAT data as compared to the direct expansion estimate over the same area. The use of LANDSAT data as an auxiliary variable was seen to reduce the variation of the multi-county (17 to 25 counties each) areas from 68 to 92 percent. One analysis area which contained only 7 counties and 11 segments showed a 23 percent reduction in variance due to use of LANDSAT data. For the 87 county area as a whole, a reduction of 64 percent was seen for the variation of the planted area estimate.

Several new procedures for analysis of LANDSAT data in general were also explored in support of this project. The split county approach will be especially useful in the future. Combined regression techniques were seen to be applicable to LANDSAT based estimation. More study is needed to compare the suggested county estimators.

VII. REFERENCES

1. Gleason, Chapman; Starbuck, Robert R.; Sigman, Richard S; Hanuschak, George A; Craig, Michael E.; Cook, Paul W.; and Allen, Richard D.; "The Auxiliary Use of LANDSAT Data in Estimating Crop Acreages: Results of the 1975 Illinois Crop-Acreage Experiment," Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C., October 1977.

2. Pratt, William L.; "The Use of Interpenetrating Sampling in Area Frames, "Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C., May 1974.

3. Craig, Michael E.; and Cardenas, Manuel; "Kansas Wheat Non-Sampling Error Analysis," Economics, Statistics, and Cooperatives Service, U.S. Department of Agriculture, Washington, D.C., April 1978.

4. Ozga, Martin; Donovan, Walter E.; and Gleason, Chapman P.; "An Interactive System for Agricultural Acreage Estimates Using LANDSAT Data," Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

5. Sigman, Richard R.; Gleason, Chapman P.; Hanuschak, George A.; and Starbuck, Robert A.; "Stratified Acreage Estimtes in the Illinois Crop-Acreage Experiment," Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

6. Ray, Robert M. III; "An Entrophy Maximization Approach to the Description of Urban Special Organization," CAC Document 237, University of Illinois, September 1977.

7. Cardenas, Manuel; Craig, Michael E.; and Blanchard, Mark M.; "Small Area Estimators: County Crop Acreage Estimates Using LANDSAT Data," contributed paper, 1978 annual ASA meeting, San Diego, California.

8. Hanuschak, George A.; "The Effect of the LANDSAT Cloud Cover Domain on Winter Wheat Acreage Estimation in Kansas During 1976," Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

Appendix A: Kansas Strata Definitions

Stratum	Description	Population Size	Average Segment Size
11	80 + % cultivated	25028	1.00 sq. mile
12	50 - 80% cultivated	21704	1.00 sq. mile
20	15 - 49% cultivated	21286	1.00 sq. mile
31	Agric. Urban	2774	.25 sq. mile
32	Urban	2941	.10 sq. mile
33	Resort	247	.25 sq. mile
40	Range	3147	4.00 sq. mile
50	Non Agric.	294	1.00 sq. mile
61	Water	29	.50 sq. mile

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

Form Approved
O.M.B. Number 40-R2766
Approval Expires 4-30-77

Appendix B: Example Form

1976 KANSAS SATELLITE CROP INFORMATION SUPPLEMENT

NAME OF OPERATOR: ______________________________
(Last) (First) (Middle Initial)

ADDRESS: ______________________________
(Route or Street) (City) (State) (Zip)

NAME OF OPERATION: ______________________________
(Record if different than operator's name above)

ENUMERATOR ____________

DATE OF VISIT: Month/Day (numeric) __ __/__ __ SEGMENT NO. ________ TRACT LETTER ______ COUNTY ____________

Tract Acres 199

	RECORD FIELD NUMBER				
	TOTAL ACRES IN FIELD	828 .	828 .	828 .	828 .
	CROP OR LAND USE *(Specify)*				
	ACRES IRRIGATED AND TO BE IRRIGATED	101 .	101 .	101 .	101 .
LAND USE COVER	LAND USE COVER WATER COVER *(Lakes, Ponds, Rivers Etc.)*	102 .	102 .	102 .	102 .
	DENSELY WOODED COVER	103 .	103 .	103 .	103 .
	GENERAL WASTELAND *(Farmstead, Roads, Ditches, Etc.)*	104 .	104 .	104 .	104 .
	SUMMER FALLOWED DURING 1976	847 .	847 .	847 .	847 .
	PERMANENT PASTURE *(Not in crop rotation)*	842 .	842 .	842 .	842 .
	BARE SOIL OR PREPARED LAND NOT YET PLANTED	105 .	105 .	105 .	105 .
	STALK FIELD *(Stalks from last years spring planted crops)*	106 .	106 .	106 .	106 .
	FIELD APPEARANCE CODE *(See Table 1)*	107	107	107	107
CROP TYPE COVER	CROP TYPE COVER WINTER WHEAT — Planted	540 .	540 .	540 .	540 .
	RYE — Planted	547 .	547 .	547 .	547 .
	OATS — Planted	533 .	533 .	533 .	533 .
	BARLEY — Planted	535 .	535 .	535 .	535 .
	CORN *(No Intentions)* — Planted	530 .	530 .	530 .	530 .
	SORGHUM *(No Intentions)* — Planted	570 .	570 .	570 .	570 .
	SOYBEANS *(No Intentions)* — Planted	600 .	600 .	600 .	600 .
	ALFALFA AND ALFALFA MIXTURES — Seeded	653 .	653 .	653 .	653 .
	HAY – OTHER THAN ALFALFA — Kind				
	Acres	65_ .	65_ .	65_ .	65_ .
	OTHER CROPS — Name				
	Acres Planted	--- .	--- .	--- .	--- .
	INTENDED USES OF CROP TYPE OTHER THAN GRAIN — Use *(See Table 2)*	150	150	150	150
	Acres	151 .	151 .	151 .	151 .
	FIELD APPEARANCE CODE *(See Table 3)*	829	829	829	829
	DATE OF HARVEST: If Field Has Been Harvested in 1976	154 MO/DAY __ __/__ __	154 MO/DAY __ __/__ __	154 MO/DAY __ __/__ __	154 MO/DAY __ __/__ __
	NOTES ON FIELD CONDITION(S) — OFFICE USE	155	155	155	155

Field Number ________ Notes ______________________________
________ ______________________________
________ ______________________________
________ ______________________________

Table 1

FIELD APPEARANCE CODE FOR LAND USE	
10	GREEN COVER (not in planted crop)
20	BARE SOIL (Prepared land not yet planted)
30	DRIED GRASS (brown pasture or fallow)
40	OTHER (Water, F.S. Feed lots, etc.)

Table 2

INTENDED USE OF CROP TYPE	
01	SILAGE
02	CROP TO CUT FOR HAY
03	CROP FOR SEED
04	PASTURED OR GRAZED
05	ABANDONED – left standing
06	ABANDONED – Plowed
07	OTHER – (Soil Imp., etc.)

Table 3

FIELD APPEARANCE CODE FOR CROP TYPE	
50	BARE SOIL (planted but not emerged)
60	GREEN (emerged with green cover, even if partial)
70	MATURE (turning color or ready for harvest) – (not to be used for green hay)
80	HARVESTED CROP (but not worked or prepared, stubble, cut hay etc.)
90	HARVESTED CROP (land worked or plowed)

Appendix C

Categorization or Classification Procedures

A. Description of LANDSAT Data*

The satellite data used in this report is LANDSAT Multispectral Scanner (MSS) data and it 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 wavelenths. 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 multivariate statistical techniques. This particular product in system corrected images refers to products that contain the radiometric and initial spatial corrections introduced during the film conversion. Every picture element (pixel) is recorded with 4 variables corresponding to one of the 4 MSS bands.

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

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.

*Excerpted from Wigton, W. "The Technology of LANDSAT Imagery and Its Value in Crop Estimation for the U.S. Department of Agriculture." Statistical Reporting Service, March 1976.

B. Discriminant Analysis*

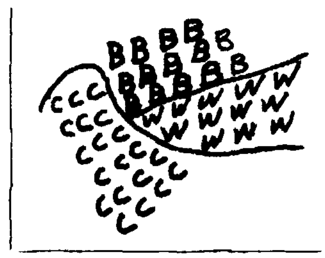
This background is intended to be general and enable the reader to understand the detailed computations and results in this report. Kendall and Stuart formulate Discriminant Analysis and Classification by stating...

> "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."

For example, the land population of interest was a portion of San Joaquin Valley in California. Cotton, wheat, and barley are the major crop populations of interest. From every acre in the San Joaquin Valley 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 cotton, wheat, or barley.

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 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 A: Two-dimensional Measurement Space



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.

---

*Excerpted from Wigton, W. "The Technology of LANDSAT Imagery and Its Value in Crop Estimation for the U.S. Department of Agriculture." Statistical Reporting Service, March 1976.

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 multivarate 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 i^{th} population is:

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

where P_i is the prior probability for the i^{th} crop

Σ_i is the covariance matrix (qxq) for the i^{th} crop

μ_i is the mean vector (q length) for the i^{th} crop

χ is the set of measurements of an individual from the i^{th} population or its equivalent discriminant score the $\log_{(e)}$ of

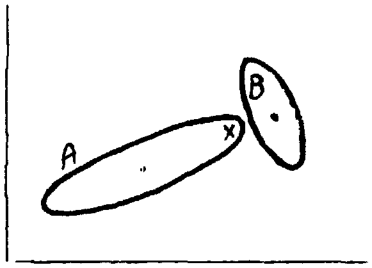
$$D_i = \log_e (P_i) - 1/2 \log_e \left|\Sigma_i\right| - 1/2 (\chi-\mu_i)' \Sigma_i^{-1} (\chi-\mu_i)$$

The boundary between two populations is quadratic (curved) and the point χ that falls in the boundary has 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 B.

Figure B: Measurement Space Showing Two Crop Density Functions and An Unknown Point (χ).



The point 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) $(X - \bar{X}_i)' S_i^{-1} (X - \bar{X}_i)$, which is a sample estimate of

$(X - \mu_i)' \Sigma_i^{-1} (X - \mu_i)$ if linear discriminant functions are used, and

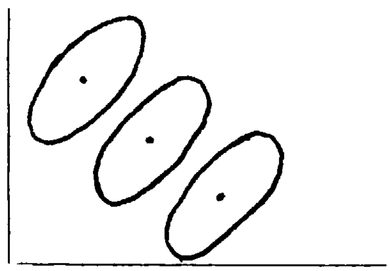
(2) $-1/2 \log_e |S_i| - 1/2 (X - \bar{X}_i)' S_i^{-1} (X - \bar{X}_i)$ if quadratic discriminant functions are used. These functions involve the exponent of the density formula of the multivarate normal distribution

$$C_{exp} = -1/2 (X - \mu_i)' \Sigma_i^{-1} (X - \mu_i)$$

of 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 for each crop has the same slope. Such an event in two-space is shown in Figure C.

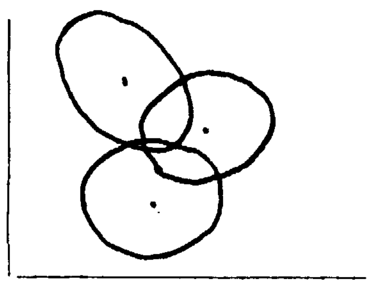
Figure C: Measurement Space Where Crop Types Have Same Covariance Matrix and Slope



This space can be partitioned effectively with straight lines. Thus, we can use linear discriminant functions.

Figure D 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 D: Measurement Space When Crops Have Different Covariance Matrices



In Figure C, even though a common center point is not present, a common covariance (ellipse) matrix would be computed. In Figure D, 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 not 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 the 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.

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 cotton where we have no prior information except sample data on cotton is extremely difficult. If a sample of 1000 points were avilable, each of these 1000 data points would need to be stored in the computer. On the other hand, if we are working with a multidimensional normal distribution, theory tells us that the sufficient statistics are computed (mean vector, and covariance 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).

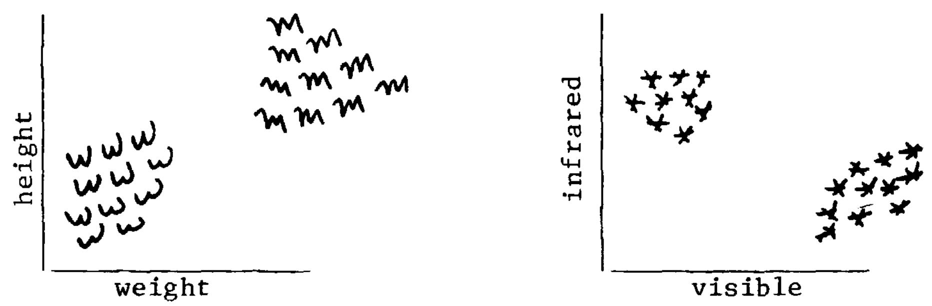
Another consideration is that all the techniques we have described require independent random samples from each crop in order to estimate the population density in the MS (training data). This point is mentioned because most remote sensing analyst do not work with randomly selected points. In this study, 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 or occurrence (prior probabilities) for each of the K populations in the total land population. For example, if 1/3 of all land is cotton, and 1/4 is barley, 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 change of occurrence, then the area in MS would be adjusted downwards.

C. Clustering*

Clustering is a data analysis technique by which one attempts to determine the natural or "inherent" relationships in a set of observations or data points. To get an intuitive idea of what is meant by natural or inherent relationships in a set of data, consider the examples in Figure E. If one were to plot height versus weight for a random sample of students, without regard to sex, on a college campus, it is likely that two relatively distinct clusters of observations would result, one corresponding to the men in the sample (heavier and taller) and another corresponding to the women (lighter and shorter). Similarly, if the spectral reflectance of vegetation in a visible wave band, were plotted against reflectance in an infrared wave band, dry vegetation and green vegetation could be expected to form discernible clusters.

Figure E: Clustering Patterns



If the data of interest never involved more than two attributes (measurements or dimensions), cluster analysis might always be performed by visual evaluation of two-dimensional plots such as those in Figure E. But beyond two or possibly three dimensions, visual analysis is impossible. For such cases it is desirable to have a computer perform the cluster analysis and report the results in a useful fashion.

In regards to the application of clustering to remote sensing research, the greatest use of cluster analysis has been for the purpose of assuring that the data used to characterize the crop or land use classes do not seriously violate the assumption of Gaussian statistics. In general, it may be expected that each distinct cluster center will correspond to a mode in the distribution of the data. Therefore, with the objective of defining a crop or land use subclass for each cluster center, the possibility of multimodal (and hence definitely non-Gaussian) crop or land use distributions is essentially eliminated.

A more detailed report on the technical development of several clustering algorithms, is provided by Swain.

---

* Excerpted from Swain, P. H., Pattern Recognition: A Basis for Remote Sensing Data Analysis. LARS information Note 111572.

Appendix D

Crop Acreage Estimation Procedures and Classifier Design Methods

A. Direct Expansion Estimation (Ground Data Only)*

Aerial photography obtained from the Agricultural Stabilization and Conservation Service is photo-interpreted using the percent of cultivated land to define broad land-use strata. Within each stratum, the total area is divided into N_h area frame units. This collection of area frame units** for all strata is called an area sampling frame. A simple random sample of n_h units is drawn within each stratum. The Statistical Reporting Service then conducts a survey in late May, known as the June Enumerative Survey (JES). In this general purpose survey, acres devoted to each crop or land use are recorded for each field in the sampled area frame units. Intensive training of field statisticians and interviewers is conducted providing rigid controls to minimize non-sampling errors.

The scope of information collected on this survey is much broader than crop acreage alone. Items estimated from this survey include crop acres by intended utilization, grain shortage on farms, livestock inventory by various weight categories, and agricultural labor and farm economic data.

Let $h = 1, 2, \ldots, L$ be the land-use strata. For a specific crop (corn, for example) the estimate of total crop acreage for all purposes and the estimated variance of the total are as follows:

Let Y = Total corn acres for a state (Illinois, for example).

$\hat{Y}$ - Estimated total of corn acres for a state.

y_{hj} = Total acres in the j^{th} sample unit in the h^{th} stratum.

Then,

$$\hat{Y} = \sum_{h=1}^{L} N_h \left(\sum_{j=1}^{n_h} y_{hj} \right) / n_h$$

---

*Excerpted from Sigman, Richard R.; Gleason, Chapman P.; Hanuschak, George A.; and Starbuck, Robert S.' "Stratified Acreage Estimation in the Illinois Crop-Acreage Experiment." Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

**In this context, all area frame units mean all the segments in the population and is not the same concept of area frame unit (count unit) used in the body of this report.

The estimated variance of the total is:

$$v(\hat{Y}) = \sum_{h=1}^{L} \frac{N_h^2}{n_h (n_h - 1)} \quad \frac{N_h - n_h}{N_h} \cdot \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2$$

Note that we have not yet made use of an auxiliary variable such as classified LANDSAT pixels. The estimator is commonly called a direct expansion estimate, and we will denote this by $\hat{Y}_{DE}$.

As an example, for the state of Illinois in 1975, the direct expansion estimates were:

Corn $\hat{Y}_{DE}$ = 11,408,070 Acres

Relative Sampling Error = 2.4% = $\sqrt{v(\hat{Y})} \ / \ \hat{Y}$

Soybeans $\hat{Y}_{DE}$ = 8,569,209

Relative Sampling Error = 2.9% = $\sqrt{v(\hat{Y})} \ / \ \hat{Y}$

B. Regression Estimation (Ground Data and Classified LANDSAT Data)

The regression estimator utilizes both ground data and classified LANDSAT pixels. The estimate of the total Y using this estimator is:

$$\hat{Y}_R = \sum_{h=1}^{L} N_h \cdot \bar{y}_{h(reg)}$$

where

$$\bar{y}_{h(reg)} = \bar{y}_h + \hat{b}_h(\bar{X}_h - \bar{x}_h)$$

and $\bar{y}_h$ = the average corn acres per sample unit from the ground survey for the h^{th} land-use stratum

$$= \sum_{j=1}^{n_h} y_{hj} \ / \ n_h$$

$\hat{b}_h$ = the estimated regression coefficient for the h^{th} land-use stratum when regressing ground-reported acres on classified pixels for the n_h sample units.

$$\hat{b}_h = \frac{\sum_{j=1}^{n_h} (x_{hj} - \bar{x}_h)(y_{hj} - \bar{y}_h)}{\sum_{j=1}^{n_h} (x_{hj} - \bar{x}_h)^2}$$

$\bar{X}_h$ = the average number of pixels of corn per frame unit for <u>all</u> frame units in the h^{th} land-use stratum. Thus <u>whole</u> LANDSAT frames must be classified to calculate $\bar{X}_h$. Note that this is the mean for the population and not the sample.

$$= \sum_{i=1}^{N_h} X_{hi}/N_h$$

X_{hi} = number of pixels classified as corn in the i^{th} area frame unit of the h^{th} stratum.

$\bar{x}_h$ = the average number of pixels of corn per sample unit in the h^{th} land-use stratum.

$$= \sum_{j=1}^{n_h} x_{hj}/n_h$$

x_{hj} = number of pixels classified as corn in the j^{th} sample unit in the h^{th} stratum.

The estimated (large sample) variance for the regression estimator is:

$$V(\hat{Y}_R) = \sum_{h=1}^{L} \frac{N_h^2}{n_h} \; \frac{N_h - n_h}{N_h} \cdot \sum_{j=1}^{n_h} (y_{hj} - y_h)^2 \cdot \frac{1 - r_h^2}{n_h - 2}$$

where

r_h^2 = sample coefficient of determination between reported corn acres and classified corn pixels in the h^{th} land-use stratum.

$$r_h^2 = \frac{\left[\sum_{j=1}^{n_h} (y_{hj} - \bar{y}_j)(x_{hj} - \bar{x}_h)\right]^2}{\sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2 \left] \left[\sum_{j=1}^{n_h} (x_{hj} - \bar{x}_h)^2 \right]}$$

Note that,

$$v(\hat{Y}_R) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h - 2} (1 - r_h^2)\, v(\hat{Y})$$

and so $\lim v(\hat{Y}_R) = 0$ as $r_h^2 \to 1$ for fixed n_h. Thus a gain in lower variance properties is substantial if the coefficient of determination is large for most strata.

The relative efficiency of the regression estimator compared to the direct expansion estimator will be defined as the ratio of the respective variance:

$$\text{R.E.} = v(\hat{Y}_{DE}) \;/\; v(\hat{Y}_R)$$

C. Ratio Estimation*

A ratio estimate of the total Y for a particular cover type is:

$$\hat{Y}_{RATIO} = \sum_{h=1}^{L} (\bar{y}_h / \bar{x}_h) X_h$$

$$= \sum_{h=1}^{L} r_h X_h, \text{ where } r_h = \bar{y}_h / \bar{x}_h$$

*Excerpted from Ozga, Martin; Donova, Walter E.; and Gleason, Chapman, P.; "An Interactive System for Agricultural Acreage Estimates Using LANDSAT Data," Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

The variance of the ratio estimate is:

$$v(\hat{Y}_{RATIO}) = \sum_{h=1}^{L} \frac{(N_h - n_h) N_h}{n_h} (S^2_{h,y} + r^2_h S^2_{h,x} - 2 r_h \rho_h S_{h,y} S_{h,x})$$

where,

ρ_h = sample correlation coefficient between x and y for the h^{th} stratum

$S^2_{h,y}$ = sample variance for the h^{th} stratum for the y variate

$S^2_{h,x}$ is similarly defined.

D. Designing a Classifier

The pixel classifier is a set of discriminant functions corresponding one-to-one with a set of classification categories. Each discriminant function consists of the category's likelihood probability multiplied by the category's prior probability. If the prior probabilities used are correct for the population of pixels being classified, then the resulting Bayes classifier minimizes the posterior probability of misclassifying a pixel for a 0-1 loss function.

In crop-acreage estimation, however, the objective is to minimize the variance of resulting acreage estimates. Since minimizing the posterior probability of misclassification does not necessarily achieve this objective, optimum acreage estimation may require the use of prior probabilities different than the optimum Bayes set.

For the case of multivariate normal signatures, the category likelihood functions are completely specified by the population means and covariances of the category signatures. Thus, the calculations of category discriminant functions involves the estimation of signature means and covariances and category prior probabilities.

Designing the classifier for this experiment consisted of the following steps:

1. Identification of classification categories.
2. Calculation of signature means and covariances and category prior probabilities from a training set of labeled pixels (called "training the classifier").
3. Measurement of classifier performance on a test set of labeled pixels (called "testing the classifier").

4. Heuristic optimization of the classifier by repeating steps 1 through 3 for different numbers of categories and/or different prior probabilities, and then proceeding to step 5 for the "optimized" classifier.

5. Estimation of classifier performance in classifying the entire pixel population.

Because of the availability of ground data, which supplied the location and cover type of agricultural fields, supervised identification of classification categories was possible. A classification category was created for each cover type in which the number of training pixels exceeded a specified threshold, usually 100 pixels. In addition, a classification category for surface water was created using pixels from rivers, lakes, and ponds.

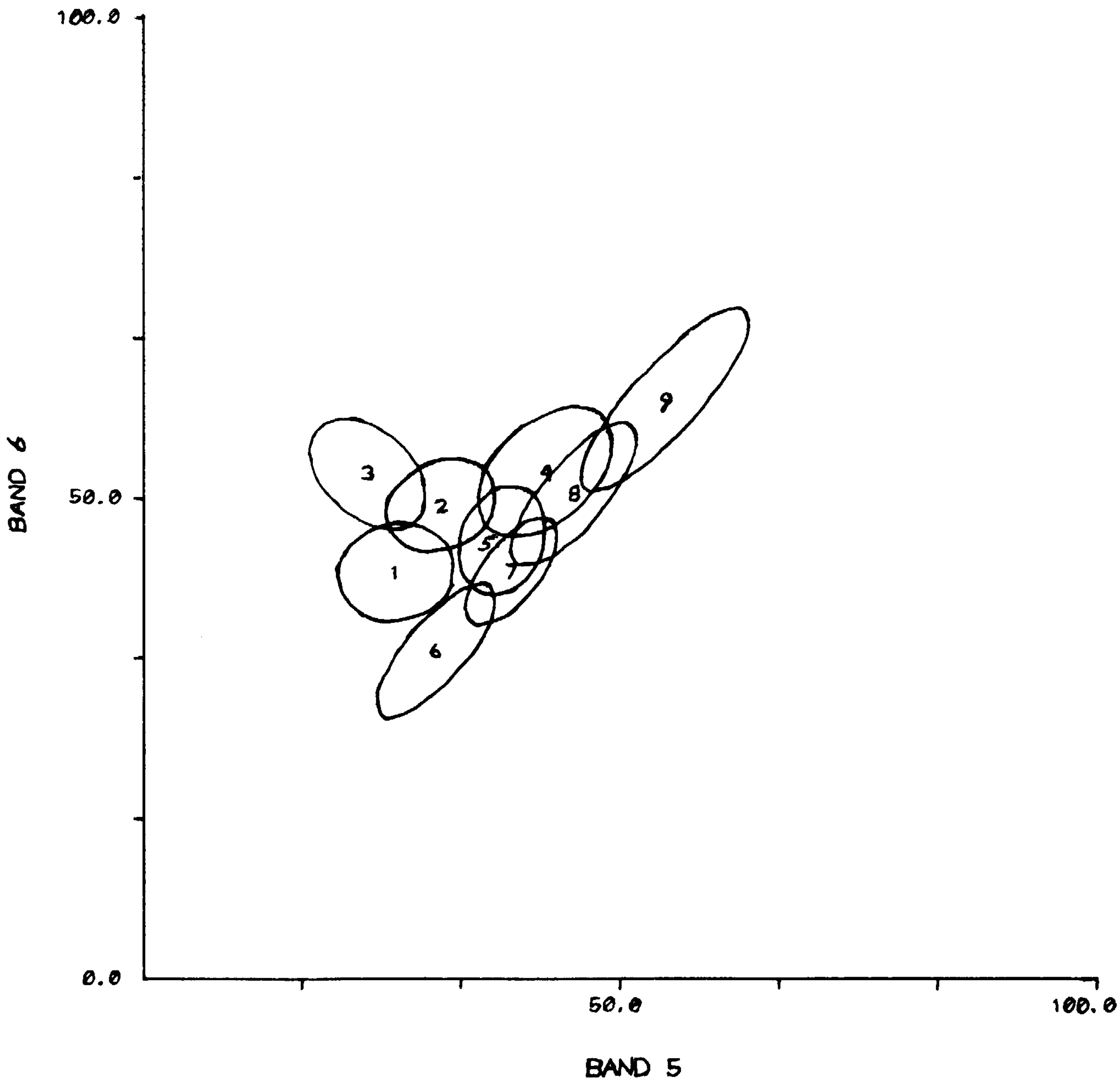
APPENDIX E:

Ellipse Plots for Final Signatures

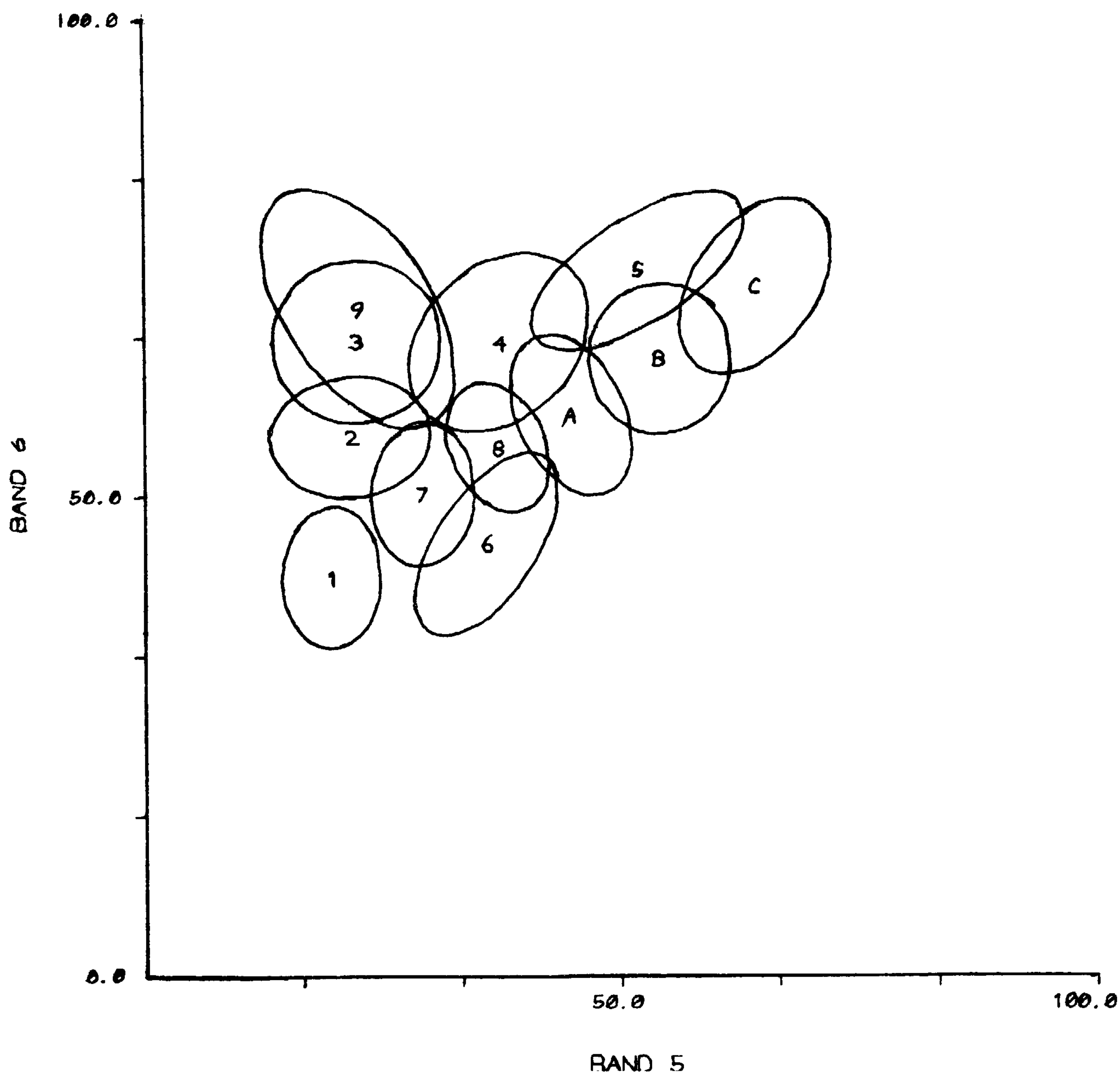
90.0% Concentration Ellipses for Bands 5 and 6

Categories	Cover Type
1, 2, 3, 4, 5	Winter Wheat
6, 7, 8, 9, A, B, C	Other

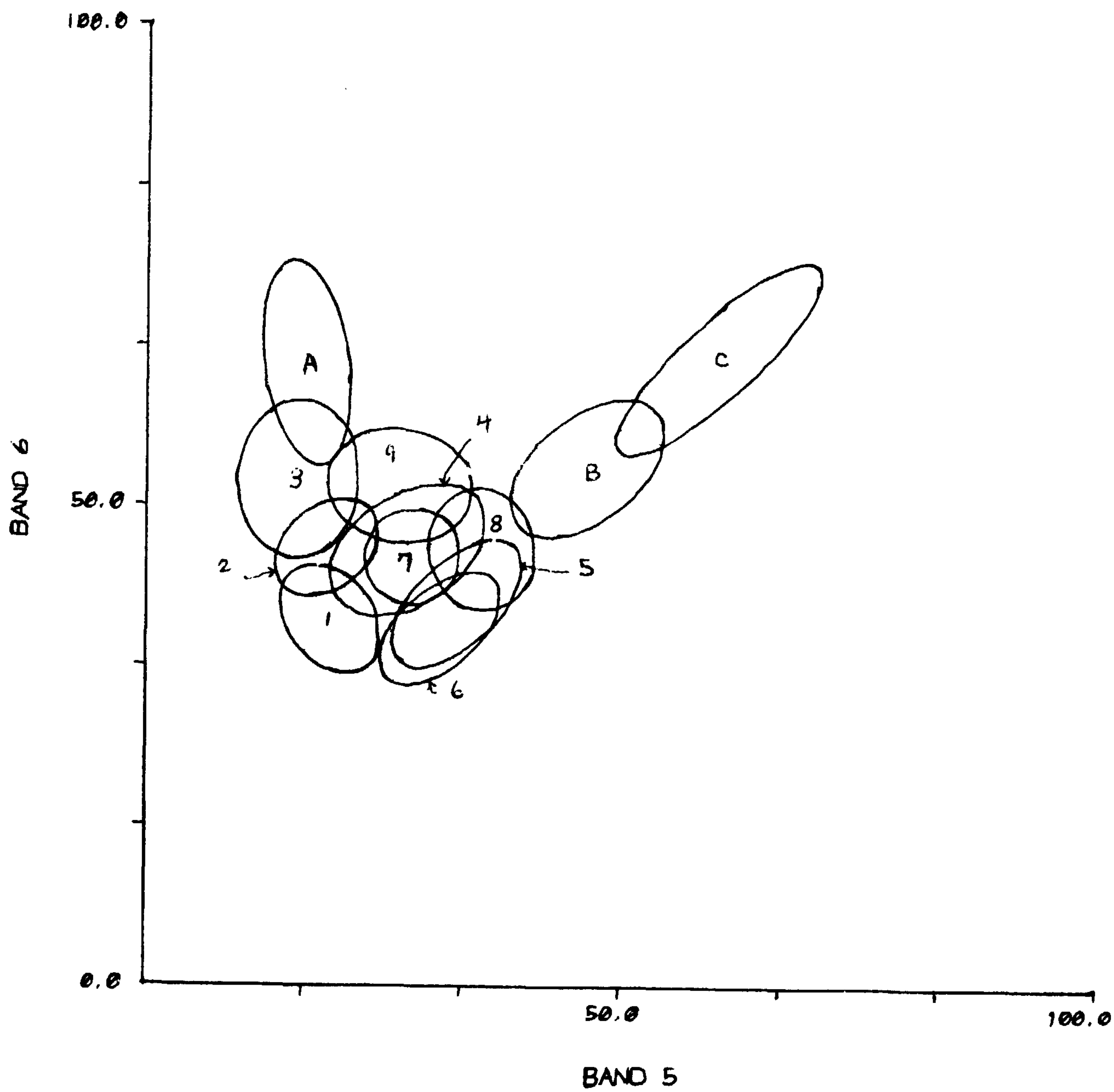
KANSAS 1976 PASS-2 STATISTICS FILE, 1-5 WHEAT



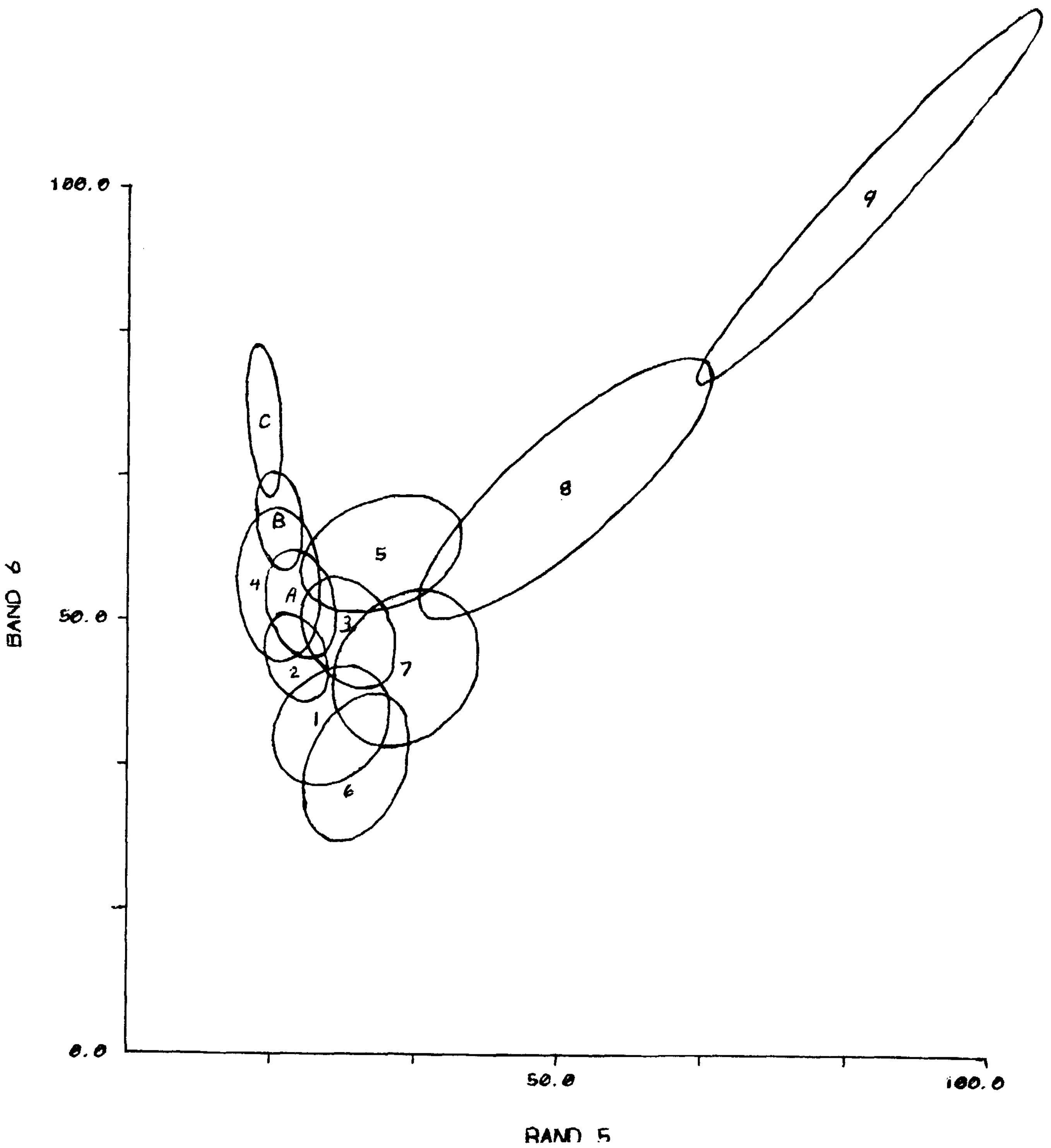
KANSAS 1976 PASS-3A STATISTICS FILE, 1-5 WHEAT



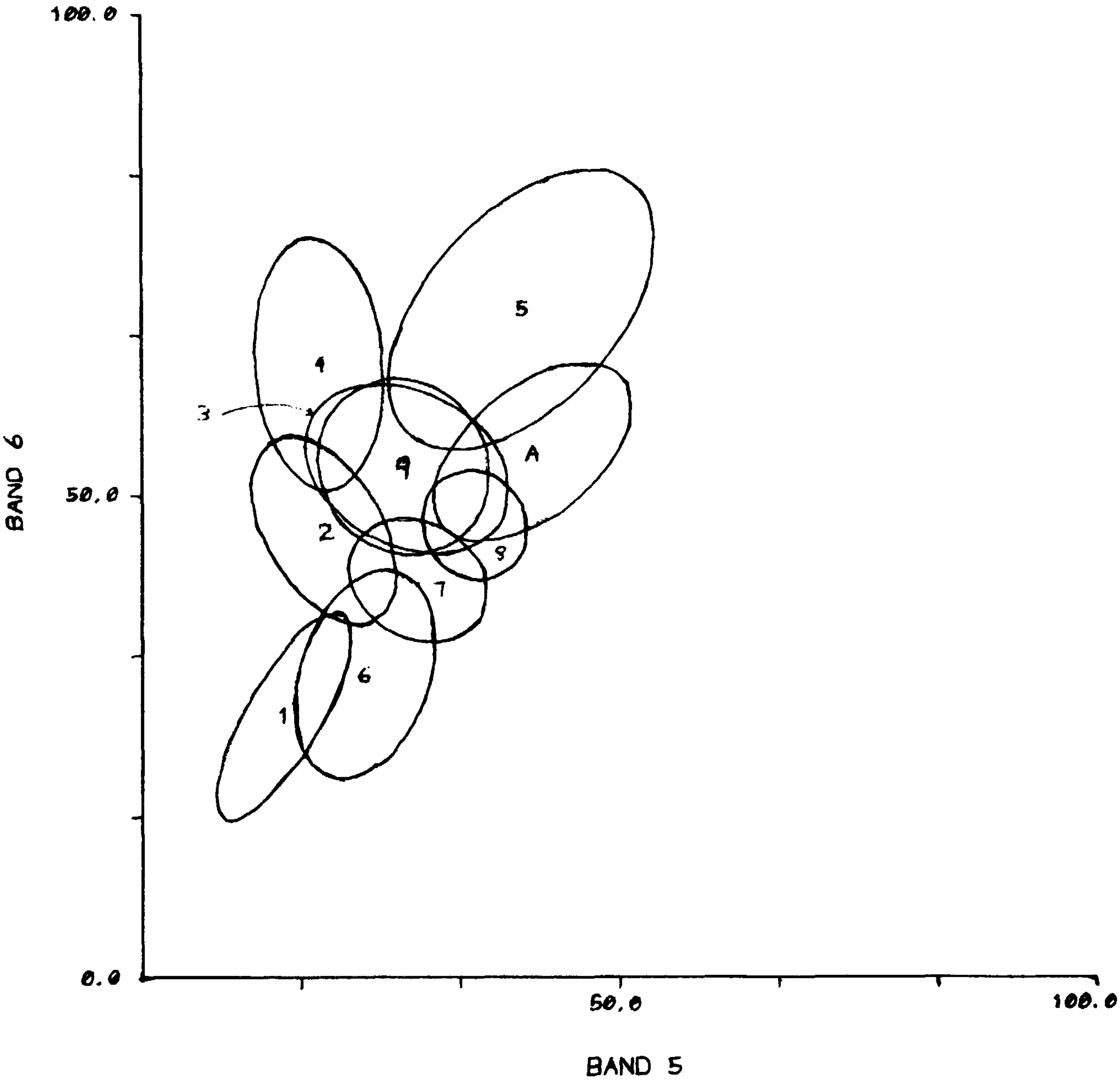
KANSAS 1976 PASS-3C STATISTICS FILE, 1-5 WHEAT



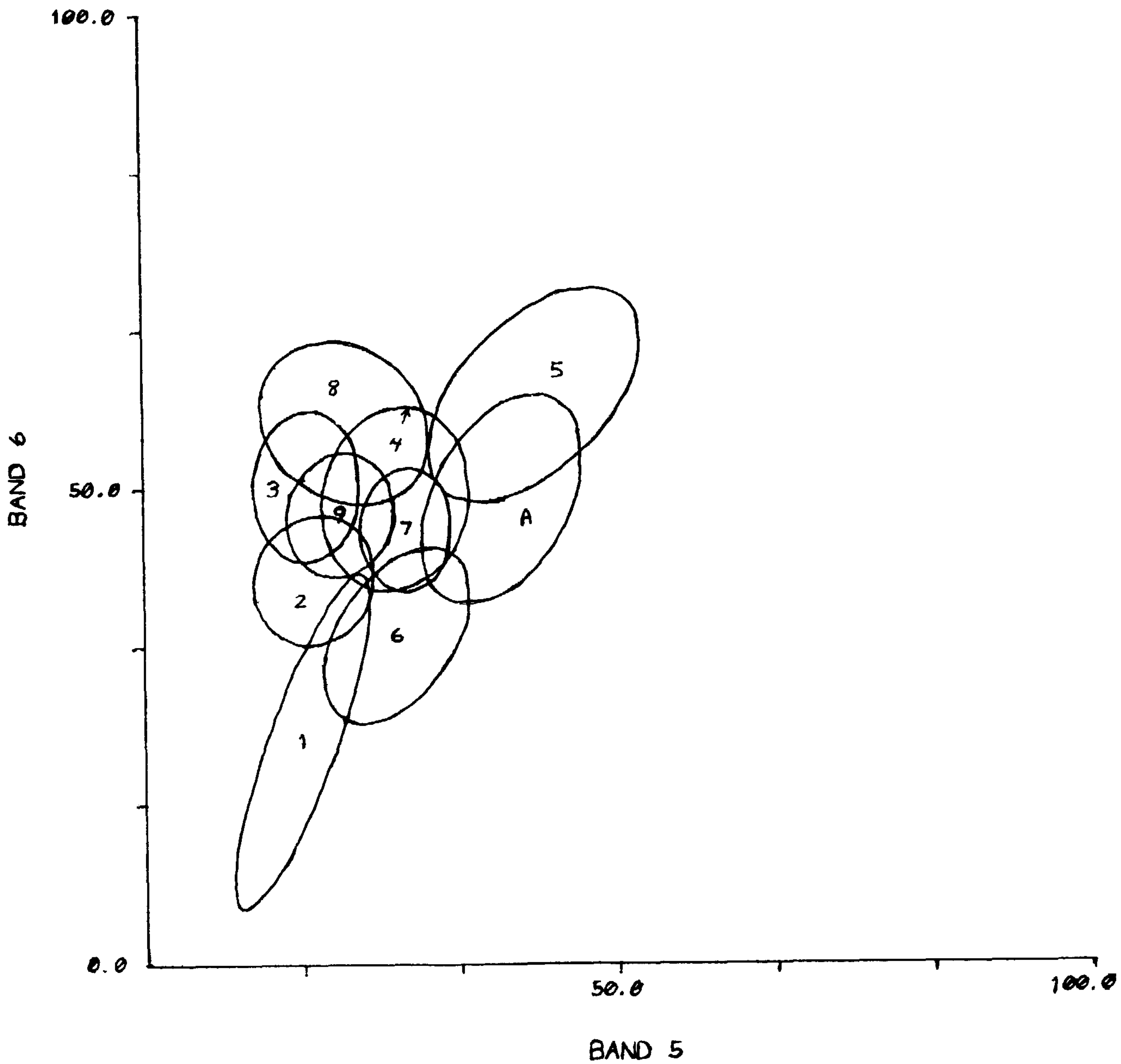
KANSAS 1976 PASS-4 STATISTICS FILE, 1-5 WHEAT



KANSAS 1976 PASS-5 STATISTICS FILE, 1-5 WHEAT



KANSAS 1976 PASS-6 STATISTICS FILE, 1-5 WHEAT



Appendix F: The Combined Regression Estimator

1. Estimation of the Total

The combined regression estimator utilizes both ground data and classified LANDSAT pixels. Since the entire state of Kansas was not covered by LANDSAT passes in one date, it was necessary to work with analysis districts post-strata which were wholly contained within a LANDSAT pass. In this study, the analysis districts were collections of counties wholly contained in a LANDSAT pass. The estimate for the state total, Y, is:

$$\hat{Y}_R = \sum_{i=1}^{P} \hat{Y}_{Ri}$$

where:

P = the number of passes required to cover the state.

$\hat{Y}_{Ri}$ = the combined regression estimate for the i^{th} post-stratum.

$$= \sum_{h=1}^{L_i} N_{ih}\, \bar{y}_{ih}(reg)$$

$$\bar{y}_{ih}(reg) = \bar{y}_{ih} + \hat{b}_i\,(\bar{X}_{ih} - \bar{x}_{ih})$$

$\bar{y}_{ih}$ = the average wheat acres per sample unit from the ground survey for the h^{th} land-use stratum within the i^{th} post-stratum.

$\hat{b}_i$ = the estimated regression coefficient when regressing ground reported acres on classified pixels for the combined data contained in the n_i sample units within the i^{th} post-stratum.

X_{ih} = the average number of pixels classified as wheat per population segment in the h^{th} land-use stratum within the i^{th} post-stratum.

$$= \frac{\sum_{h=1}^{Li} \frac{W_{ih}(1-f_{ih})}{n_{ih}(n_{ih}-1)} \sum_{j=1}^{n_{ih}} (y_{ihj} - \bar{y}_{ih})(x_{ihj} - \bar{x}_{ih})}{\sum_{h=1}^{Li} \frac{W_{ih}^2(1-f_{ih})}{n_{ih}(n_{ih}-1)} \sum_{j=1}^{n_{ih}} (x_{ijh} - \bar{x}_{ih})^2}$$

y_{ihj} = the total wheat acres in the j^{th} sample segment of the h^{th} stratum within the i^{th} post-stratum.

x_{ihj} = number of pixels classified as wheat in the j^{th} sample segment of the h^{th} stratum within the i^{th} post-stratum.

x_{ih} = the average number of pixels of wheat per sample segment in the h^{th} land-use stratum within the i^{th} post-stratum.

L_i = number of strata in the i^{th} post-stratum.

n_{ih} = number of sampled segments in the h^{th} land-use stratum in the i^{th} post-stratum.

$W_{ih}^2 = \frac{N_{ih}}{N_i}$ = number of population segments in the h^{th} land-use stratum in the i^{th} post-stratum divided by the number of population segments in the i^{th} post-stratum.

$f_{ih} = n_{ih}/N_{ih}$

The estimated variance of $\hat{Y}_R$ is given by:

$$V(Y_R) = \sum_{i=1}^{P} V(Y_{Ri})$$

$$\sum_{i=1}^{P} \sum_{h=1}^{Li} N_{ih}^2 V(\bar{y}_{ih}(reg))$$

where

$$V(\bar{y}_{ih}(reg)) = \sum_{h=1}^{Li} \frac{W_{ih}^2(1-f_{ih})}{N_{hi}} (s_{iyh} + 2b_i s_{iyxh} + b_i^2 s_{ixh}^2).$$

with

$$s^2_{ihy} = \frac{\sum_2 (y_{ihj} - y_{ih})^2}{j - 1}$$

$$s_{1yxh} = \frac{\sum_i (y_{ihj} - \bar{y}_{ih})(x_{ihj} - \bar{x}_{ih})}{j - 1}$$

and

$$s^2_{ixh} = \frac{\sum_j (x_{ihj} - \bar{x}_{ih})^2}{j - 1}$$

2. County Estimates

Let $N_{ih,c}$ = total number of segments in the h^{th} stratum of the c^{th} county within the i^{th} post-stratum.

$\bar{X}_{ih,c}$ = total number of pixels in the c^{th} county classified as wheat for the h^{th} stratum in the i^{th} post-stratum divided by N_{ih}.

Then the estimate based on the combined regression estimator is the total wheat acreage for the c^{th} county is,

$$\hat{Y}_{reg,c} = \sum_{h=1}^{Li,c} N_{ih} \left[y_{ih} + \hat{b}_i (\bar{X}_{ih,c} - \bar{x}_{ih}) \right]$$

where

$L_{i,c}$ = the number of strata in county c within the i^{th} post-stratum.

The estimated variance of $\hat{Y}_{reg,c}$ is:

$$V(\hat{Y}_{reg,c}) = \sum_{h=1}^{li,c} N^2_{ih} \left\{ W^2_{ih,c} + \frac{W^2_{ih}}{n_{ih}} + \left[\sum_{k=1}^{Li} \frac{d^2_{ih}\, s^2_{ixh}}{n_{ih} - 1} \Big/ \left(\sum_{k=1}^{Li} d_{ih} s^2_{ixh} \right)^2 \right] \left(W_{ih,c}\bar{X}_{ih,c} - W_{ih}\bar{x}_{ih} \right)^2 \right\}$$

where

$$d_{ih} = \frac{W_{ih}^2}{n_{ih}} (1 - f_{ih})$$

$$w_{ih,c} = \frac{W_{ih,c}}{N_{i,c}}$$

$N_{i,c}$ = number of segments in county c in i^{th} post-stratum.

$x_{ih,c}$ = mean of classified pixels classified as wheat per segment in the h^{th} stratum in county c of the i^{th} post-stratum.

The variance formula given above for the county estimator is derived by treating the part of county c in stratum h as a single segment.

Appendix G: Kansas 1976 Winter Wheat County Estimates
(Estimates in thousands of hectares, CV in percent)

County	Pass	SSO Estimate	Combined Regression			"Ratio"		
			EST.	S.D.	C.V.	EST.	S.D.	C.V.
Cheyenne	2	65.96	72.72	8.91	12.25	75.76	17.36	22.92
Decatur	2	45.73	88.75	6.03	6.79	96.92	32.05	33.07
Grant	2	35.61	23.88	10.12	42.37	24.32	4.13	16.97
Greeley	2	89.84	48.51	15.74	22.98	61.39	11.09	18.06
Hamilton	2	91.05	24.40	15.15	62.07	26.59	4.13	15.52
Haskell	2	48.16	42.53	12.63	29.69	38.04	7.29	19.15
Kearny	2	62.73	47.35	11.39	24.07	44.03	6.84	15.53
Logan	2	60.70	72.32	10.15	14.03	73.01	14.23	19.49
Morton	2	40.47	11.53	10.01	86.79	14.12	1.93	13.70
Rawlins	2	68.80	85.11	9.11	10.70	94.17	27.20	28.88
Scott	2	54.63	60.06	11.74	19.54	55.36	9.19	16.60
Sherman	2	73.65	74.58	17.56	23.55	68.11	11.45	16.82
Stanton	2	70.82	22.38	13.35	59.67	23.07	4.05	17.54
Stevens	2	34.80	32.42	11.66	35.96	32.01	5.34	16.69
Thomas	2	93.89	99.88	19.83	19.85	90.00	15.22	16.91
Wallace	2	44.52	41.52	6.83	16.44	41.36	6.91	16.70
Wichita	2	54.23	44.96	12.67	28.17	44.39	8.13	18.32
Clark	3	41.68	58.84	10.16	17.27	61.59	9.08	14.74
Ellis	3	53.82	37.27	9.61	25.78	41.16	5.01	12.17
Finney	3	79.72	64.51	20.37	31.58	66.25	8.97	13.55
Ford	3	94.29	99.43	19.34	19.46	106.92	14.37	13.44
Gove	3	52.61	46.30	16.67	36.01	56.25	7.20	12.81
Graham	3	46.13	47.63	14.93	31.35	58.27	8.62	14.79
Gray	3	67.99	50.18	17.85	35.57	52.85	8.22	15.55
Hodgman	3	57.06	50.26	10.28	20.46	57.63	7.29	12.65
Lane	3	52.20	40.19	8.70	21.66	48.81	7.12	14.60
Meade	3	73.25	47.71	12.61	26.43	48.24	5.88	12.19
Ness	3	87.01	57.06	14.09	24.69	67.70	8.87	13.10
Norton	3	46.13	88.22	12.71	14.40	71.83	9.75	13.58
Phillips	3	39.25	61.16	14.29	23.37	71.87	11.74	16.34
Rooks	3	55.85	50.67	10.49	20.71	58.92	7.94	13.48
Rush	3	76.49	61.39	10.52	17.14	69.97	8.94	12.78
Seward	3	33.59	30.59	8.72	28.51	29.22	4.05	13.87
Sheridan	3	47.35	45.08	13.68	30.34	50.87	6.56	12.89
Smith	3	49.37	72.20	12.51	17.32	84.82	11.21	13.22
Trego	3	52.61	40.71	11.09	27.25	48.44	6.48	13.38
Barton	4	97.53	92.92	12.07	12.99	113.19	9.15	8.09
Edwards	4	50.99	42.53	8.59	20.20	50.79	3.55	6.99
Ellsworth	4	53.82	50.30	4.08	8.12	60.50	2.99	4.95
Kowa	4	50.18	36.14	6.89	19.06	42.37	2.25	5.32
Pawnee	4	70.01	68.23	16.15	23.67	76.93	9.80	12.73
Partt	4	87.01	68.72	10.25	14.91	80.90	7.46	9.22
Stafford	4	84.98	72.44	12.88	17.77	83.04	9.48	11.41
Butler	5	49.78	49.21	12.75	25.91	36.50	3.01	8.25
Chase	5	11.33	20.76	3.20	15.41	18.74	2.10	11.22
Clay	5	43.71	49.49	10.16	20.53	53.86	5.75	10.67
Cowley	5	65.56	58.52	10.28	17.57	53.42	4.44	8.32
Dickinson	5	78.32	77.62	14.85	19.13	91.22	9.39	10.29
Geary	5	12.55	16.39	3.79	23.10	11.37	1.17	10.25
Harper	5	120.19	70.01	15.10	21.56	79.52	7.53	9.47

County	Pass	SSO Estimate	Combined Regression			"Ratio"		
			EST.	S.D.	C.V.	EST.	S.D.	C.V
Harvey	5	60.30	45.16	10.64	23.57	53.58	7.69	14.35
Marion	5	71.63	63.54	12.43	19.59	69.57	7.41	10.66
Marshall	5	44.92	68.15	12.88	18.90	62.77	5.38	8.57
McPherson	5	104.81	80.82	15.30	18.93	95.95	13.80	14.38
Morris	5	25.90	31.65	9.37	29.59	25.09	2.49	9.94
Nemaha	5	24.69	53.94	10.52	19.51	50.59	4.86	9.60
Ottawa	5	69.20	66.09	7.36	11.14	67.70	6.76	9.98
Pottawatomie	5	18.21	33.63	6.01	17.87	26.35	2.23	8.48
Riley	5	14.16	22.54	4.68	20.78	19.59	1.98	10.09
Saline	5	72.03	60.34	9.48	15.72	73.17	7.55	10.31
Wabaunsee	5	12.95	26.79	4.12	15.37	22.62	2.18	9.62
Washington	5	44.52	52.93	17.56	33.18	48.76	4.86	9.96
Allen	6	16.59	13.23	1.58	11.93	9.11	2.75	30.23
Anderson	6	19.02	21.41	2.15	10.03	18.25	5.75	31.49
Atchison	6	10.52	8.22	1.30	15.77	6.88	2.43	35.30
Brown	6	20.23	13.44	2.63	19.58	3.97	1.62	40.83
Chautauqua	6	10.52	11.21	1.97	17.62	10.52	1.79	17.00
Coffey	6	18.21	19.67	3.48	17.70	13.72	4.25	30.98
Doniphon	6	8.50	33.14	2.02	6.11	21.04	6.64	31.54
Douglas	6	12.95	18.21	1.74	9.56	14.00	4.37	31.22
Elk	6	8.90	10.12	2.28	22.57	9.87	1.63	16.52
Franklin	6	16.19	19.55	1.82	9.32	14.41	4.78	33.15
Greenwood	6	8.50	16.11	3.44	21.37	16.19	2.57	15.88
Jackson	6	14.97	7.73	3.00	38.76	7.69	2.27	29.49
Jefferson	6	11.74	11.90	2.19	18.38	7.85	2.19	27.85
Johnson	6	7.69	18.13	1.47	8.11	15.22	4.74	31.15
Labette	6	32.78	17.85	2.27	12.71	14.33	4.29	29.95
Leavenworth	6	8.09	12.10	1.42	11.75	8.62	2.55	29.61
Lyon	6	24.28	23.55	2.43	10.30	16.07	3.02	18.81
Miami	6	13.35	21.41	2.02	9.45	14.20	4.05	28.49
Montgomery	6	24.28	18.01	3.12	17.33	15.58	4.41	28.33
Neosho	6	18.21	19.34	2.15	11.10	15.94	5.06	31.73
Osage	6	21.45	18.05	3.70	20.48	15.38	4.67	30.34
Shawnee	6	17.81	16.75	1.90	11.32	9.96	2.34	23.46
Wilson	6	24.28	22.14	2.00	9.02	15.05	4.13	27.47
Woodson	6	12.14	10.81	2.12	19.62	8.90	1.96	22.05
Wyandotte	6	1.21	2.63	0.36	13.76	1.54	0.46	29.77