LACIE-00467
JSC-13731

LARGE AREA CROP INVENTORY EXPERIMENT (LACIE)

7.9-10.123
TM-79987

NASA NOAA USDA

LACIE THIRD INTERIM PHASE III
ACCURACY ASSESSMENT REPORT

(E79-10123) LARGE AREA CROP INVENTORY EXPERIMENT (LACIE). LACIE THIRD INTERIM PHASE 3 ACCURACY ASSESSMENT REPORT Interim Report (NASA) 141 p HC A07/MF A01 CSCL 02C

N79-18390
Unclas
G3/43 00123

National Aeronautics and
Space Administration

Lyndon B. Johnson Space Center
Houston, Texas 77058

JANUARY 1978

LACIE PHASE III ACCURACY ASSESSMENT

THIRD INTERIM REPORT

APPROVED BY

D. E. Pitts, Manager
LACIE Accuracy Assessment

J. D. Erickson, Chief
Research, Test, and Evaluation Branch

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

January 1978

CONTENTS

Section | Page

Appendix

Section		Page

TABLES

Table		Page

FIGURES

ABBREVIATIONS

AA	Accuracy Assessment.
ACC	adjustable crop calendar.
agromet	agricultural/meteorological.
AI	analyst/interpreter
biowindow or biophase	biological window, biological phase — a Landsat data acquisition period that is related to the biostages of wheat development. The LACIE approach is based on the judgment that wheat can be separated adequately from other crops by machine analysis of up to four acquisitions of Landsat data during the growing season. The biowindow may be updated if there is a significant lag or advancement in the current crop calendar. The sequence chosen generally includes acquisitions during the following biowindows: 1. Crop establishment — from field preparation to jointing (biostage 1.0 to 3.0). 2. Green — from jointing to heading (biostage 3.0 to 4.0). 3. Heading — from heading to soft dough (biostage 4.0 to 5.0). 4. Mature — from soft dough to harvest (biostage 5.0 to 7.0).
biostage	biological stage — the specific stage of development of a crop which can be recognized by a major change in plant structure; i.e., emergence after germination, jointing, heading, soft dough, ripening, and harvest, which are represented by integers on the Robertson Biometeorological Time Scale.
blind sites	LACIE sample segments chosen at random for which ground truth is obtained in order to test classification performance. The identity of the blind sites is withheld from the CAMS analysts so that these segments will be treated the same as the other segments.
BMTS	Biometeorological Time Scale.
CAMS	Classification and Mensuration Subsystem.

PRECEDING PAGE BLANK NOT FILMED

CAR	CAS Annual Report.
CAS	Crop Assessment Subsystem.
CCEA	Center for Climatological and Environmental Assessment — an organization of the National Oceanic and Atmospheric Administration, Columbia, Missouri.
classification	in computer-aided analysis of remotely sensed data, the process of assigning data points to various classes by a testing process in which the spectral properties of each unknown data point are compared with spectral properties typical of these classes.
classification error	a measure of the degree to which the LACIE classification either overestimates or underestimates the wheat acreage in a specific area.
CMR	CAS Monthly Report.
CRD	Crop Reporting District — a geographical area used by the U.S. Department of Agriculture for the collection and reporting of agricultural information; each district consists of several counties.
crop calendar	a calendar depicting the biostages of the major crop types within a specified region during a calendar year.
crop calendar adjustment	an adjustment made to the historical crop calendar on the basis of current meteorological data.
CUR	CAS Unscheduled Report.
CV	coefficient of variation (standard deviation divided by the mean).
DAPTS	Data Acquisition, Preprocessing, and Transmission Subsystem.
Group II segment	LACIE segment in a county that historically produces small quantities of wheat/small grains; samples are allocated with probability proportional to size.
IE	Information Evaluation.
IMR	IE Monthly Report.
ITS	intensive test site — a LACIE test segment in the United States or Canada on which detailed crop information is collected by using ground and airborne equipment.

JSC	Lyndon B. Johnson Space Center of NASA, Houston, Texas.
LACIE	Large Area Crop Inventory Experiment.
Landsat	Land Satellite — formerly called ERTS (Earth Resources Technology Satellite); operates in a circular, Sun-synchronous, near-polar orbit of Earth at an altitude of approximately 915 kilometers; orbits Earth about 14 times a day and views the same scene at least every 18 days.
LEC	Lockheed Electronics Company, Inc.
MSE	mean squared error.
MSS	Multispectral Scanner System or multispectral scanner — the remote sensing instrument on Landsat that measures reflected sunlight in various spectral bands or wavelengths.
NASA	National Aeronautics and Space Administration.
NOAA	National Oceanic and Atmospheric Administration.
90/90 criterion	criterion that the LACIE U.S. Great Plains at-harvest production estimate be within 10 percent of the true value with a probability of at least 0.9.
PPS	probability proportional to size.
Sample segments	the 5- by 6-nautical-mile areas used as samples in LACIE to make acreage estimates. They are selected by a sampling strategy which is described in appendix A of this report.
USDA	U.S. Department of Agriculture.
USDA/ASCS	USDA Agricultural Stabilization and Conservation Service.
USDA/FAS	USDA Foreign Agricultural Service.
USDA/SRS	USDA Statistical Reporting Service.
U.S. Great Plains (USGP) (USSGP) (USNGP)	The U.S. Great Plains (USGP), an area encompassing the nine states of Colorado, Kansas, Minnesota, Montana, Nebraska, North and South Dakota, Oklahoma, and Texas; it is divided geographically into (1) the U.S. southern Great Plains (USSGP), which includes Colorado, Kansas, Nebraska, Oklahoma and Texas, and (2) the U.S. northern Great Plains (USNGP), which includes Minnesota, Montana, and North and South Dakota.

USGP-7 — Seven winter wheat states of the U.S. Great Plains region. These include all of the USGP states except North Dakota and Minnesota.

YES — Yield Estimation Subsystem.

INTRODUCTION

The Large Area Crop Inventory Experiment (LACIE) is an interagency endeavor of the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), and the United States Department of Agriculture (USDA). Its purposes are (1) to demonstrate the economical benefit to be obtained by using remotely sensed data from the Land Satellite (Landsat) for agricultural applications, (2) to test the capability of a system utilizing remote sensing in conjunction with climatological, meteorological, and conventional data to produce timely estimates of the production of a major world crop prior to harvest, and (3) to validate the technology and procedures for such a system.

In accordance with the objectives of LACIE, the Accuracy Assessment (AA) effort is designed to check the accuracy of the products from the experimental operations throughout the growing season and thereby determine if the procedures used are adequate to accomplish the above objectives.

1.1 OBJECTIVES

The objectives of AA are as follows:

a. To determine whether the accuracy goal of the LACIE estimate of wheat production for a region or country is being met. The LACIE accuracy goal is a 90/90 at-harvest criterion for wheat production. This specifies that the at-harvest wheat production estimate for the region or country be within 10 percent of the true production with a probability of at least 0.9.

b. To determine the accuracy and reliability of early season estimates and estimates made at regular intervals throughout a crop season prior to harvest. This includes a determination of the degree to which the 90/90 criterion is supported at these intervals during the crop season.

c. To investigate the various sources of error in the LACIE estimates of wheat production, area, and yield, to quantify and relate these error sources to causal elements in the LACIE estimation process, and to recommend procedures for reducing the error.

1.2 ACCURACY ASSESSMENT ACTIVITIES

In order to satisfy its objectives, AA carries out several types of evaluations and the results are presented in (1) monthly quick-look reports; (2) a number of interim reports leading up to a final report, and (3) certain special reports The following paragraphs contain the descriptions of the AA evaluations presented in the three types of reports.

1.2.1 ACTIVITIES REPORTED IN THE QUICK-LOOK REPORTS

The quick-look reports contain an evaluation by AA of the LACIE estimates reported in the Crop Assessment Subsystem (CAS) monthly reports (CMR's) and the CAS unscheduled reports (CUR's). The quick-look reports are released one week following the release of a CMR or a CUR. The CMR's and CUR's contain the official LACIE estimates of wheat production, area, and yield, and the corresponding statistics. The true wheat production, area, and yield for the particular region or country are, of course, unknown. Therefore, to ascertain the accuracy of the LACIE estimates, comparisons are made with a reference standard. In the United States, the reference standard consists of the most recent (at the time of the comparison) estimates released by the Statistical Reporting Service of the USDA (USDA/SRS). In foreign countries, the reference consists of the most recent estimates released by the Foreign Agricultural Service of the USDA (USDA/FAS). The AA quick-look reports contain a comparison of the LACIE estimates of wheat production, area, and yield with the corresponding reference standard, as well as significance tests of no difference at the region or country level. The relative difference calculated at the zone level (state in the U.S.) is used to indicate problem areas; available blind site results are given and an intensive test site (ITS) example is presented.

1.2.2 ACTIVITIES REPORTED IN THE INTERIM AND FINAL REPORTS

The interim reports are released at regular intervals throughout the crop season. They contain the results of the previous quick-look reports, a discussion of the 90/90 criterion as it applies to the region for which the LACIE estimates of wheat production are available, and the results of investigations of error sources[1] in the LACIE wheat production estimate including the blind site and ITS analyses. Also, any recommendations for improvement made by AA are documented in the interim and final reports.

Each interim report is built up from the previous one by including data that became available during the interim period. Technical comments on each report are solicited from a variety of sources and are used to upgrade subsequent reports. Early-season and mid-season evaluations are made in the first and second interim reports; late-season and at-harvest evaluations are made in the third and fourth interim reports.

The fourth interim report also serves as a draft for the final report, which contains material which is similar to the interim reports but covers the entire year.

The above schedule was followed in Phase II. In Phase I there were no interim reports and the Phase I final report was incorporated into the Phase II final report.

1.2.3 ACTIVITIES REPORTED IN AA UNSCHEDULED REPORTS

From time to time, special investigations are carried out that are of interest to LACIE but which are not required on a regular basis such as those mentioned above. These investigations are reported in AA unscheduled reports.

1.3 PROCEDURES USED IN OBTAINING LACIE PHASE III ESTIMATES

This report consists of evaluations of LACIE estimates of production, area, and yield for the U.S. Great Plains (USGP) region and for the U.S.S.R.; these estimates were released in the CAS reports for LACIE Phase III. During Phase III several changes were made in the aggregation procedures used by CAS. This section describes the procedures used in the various CAS reports. Some of the changes imposed by CAS during Phase III altered the Phase III monthly estimates.

[1] A detailed description of the error sources in LACIE is given in appendix A.

The initial Phase III CAS report for the USGP was released February 8, 1977, prior to the availability of the Phase III allocation of sample segments. Thus, estimates published in the February CMR are the result of the aggregation of segments from the LACIE Phase II allocation.

On April 6, the second Phase III CAS report was released. The estimates in this report were based on the Phase III allocation, but only segments which were available as of the cutoff date of the February 8 CMR were used. Thus, the only difference between the results in the February 8 CMR and the April 6 CUR is that in the former the Phase II allocation was used, whereas in the latter the Phase III allocation was used. The April 6 CUR was updated on April 22 with the release of a CUR which was based on all of the acquisitions from the Phase III allocation that were available at that date.

In LACIE Phase III, CAS developed an objective thresholding procedure to eliminate acquisitions prior to emergence. This procedure was tested and was demonstrated to reduce the magnitude of the underestimate throughout the season. Thus, in addition to the regular estimates, CAS also generated the threshold estimates in the June and July CMR's. Further, the threshold estimate replaced the regular LACIE estimate in the August, September, and October CMR's.

In September, CAS further modified the data with a procedure called screening, whereby segments were stratified according to historic county wheat proportions. In the screening procedure, CAMS proportion estimates which disagreed with their corresponding historic county proportions by a large margin (stipulated by the procedure) were excluded from the aggregation.

As a result of the investigation of the overestimation problem in South Dakota, which was initiated immediately following the release of the July 11 CMR, a redesignation of sample segments into winter wheat, spring wheat, and mixed wheat segments was instituted in August for the mixed wheat states of Montana and South Dakota. Previously, both winter and spring wheat estimates were made for each segment in Montana and South Dakota, resulting in winter wheat estimates for segments containing little or no winter wheat and spring wheat estimates for segments containing little or no spring wheat. Under the redesignation, if a county containing allocated segments contributed 1 percent or more to the state winter wheat production, its segments were designated as winter wheat segments. The same rule applied for spring wheat. This divides the counties into three groups: pure winter, pure spring and mixed. Further, counties in the pure spring and pure winter wheat groups were subsequently designated mixed if the within-county contribution for either crop type to total wheat for the county was between 25 and 75 percent. This procedure was also applied to the oblasts in the mixed wheat region of the U.S.S.R.

Table 1-1 is a summary of the procedures and allocations used in the various Phase III CMR's for the USGP.

In the first U.S.S.R. CAS report, released on August 5, 1977, the estimates for production, area, and yield were obtained using the conventional aggregation procedure.

In the second U.S.S.R. CMR, released on September 7, 1977, the official estimates were also obtained using the conventional procedure, but in addition some unofficial "modified" estimates were given. The modified estimates were obtained using a procedure which was the same as the conventional procedure except that acquisitions obtained after May 1, 1977 were thresholded (i.e., not used) unless prior acquisitions were also available, or unless usable acquisitions from biostage 6 or 7 were available.

In the third U.S.S.R. CMR, released on October 5, 1977, the official estimates were obtained using the modified procedure.

TABLE 1-1.— CAS ALLOCATION AND PROCEDURE CHANGES FOR USGP DURING PHASE III OF LACIE

CAS Report Date	Allocation	Procedure
February 8, 1977	Phase II	Conventional
April 6, 1977	Phase III acquisitions available as of February 8 report	Conventional
April 22, 1977	Phase III - all classifications available to date	Conventional
June 7, 1977	Phase III	Conventional (official); Thresholding (unofficial)
July 11, 1977	Phase III	Conventional (official); Thresholding (unofficial)
August 10, 1977	Phase III - Montana and South Dakota sites redesignated	Thresholding (official)
September 9, 1977	Phase III - Montana and South Dakota sites redesignated	Thresholding; screening
October 11, 1977	Phase III - USGP sites redesignated	Thresholding; screening

2. SUMMARY

This report discusses the evaluations of the LACIE production, area, and yield estimates released in the February 7, April 6, April 22, May 9, June 7, July 11, August 10, September 9, and October 11, 1977, CAS U.S. Great Plains reports. Also discussed are the estimates released in the CAS U.S.S.R. reports of August 5, September 7, and October 5, 1977. In the first three U.S. reports, the LACIE area estimates were compared with the USDA/SRS estimates of planted area and the LACIE yield estimates were compared with a "derived yield," obtained by dividing the USDA/SRS production estimate by the corresponding estimate of the planted area. The LACIE estimates released in the May and later reports were compared with the corresponding USDA/SRS monthly estimates of harvested area, yield, and production.

An accuracy of 90/85 was achieved with the October estimates which had a relative bias of -9.9 percent and a coefficient of variation (CV) of 5.2 percent for the total wheat production in the USGP. That is, the probability is 0.9 that the LACIE estimate was within ±15 percent of true wheat production for the USGP.

The LACIE total wheat production estimates for the USGP region are available only in the August, September, and October CAS reports. In all three instances the LACIE estimate was significantly smaller than that of the USDA/SRS, primarily because the LACIE spring wheat production was underestimated. All three USNGP spring wheat estimates were significantly smaller than their USDA/SRS counterparts, while there were no significant differences between the LACIE and USDA/SRS USGP-7 winter wheat estimates from June (when LACIE statistics first became available) through October.

The LACIE spring wheat production underestimates in August, September, and October are the result of area underestimates for spring wheat in the USNGP region. The LACIE estimates were significantly smaller than the USDA/SRS estimates for all 3 months.

In the blind site investigation, it was found that for the USNGP spring wheat blind sites the average of the LACIE proportion estimates was significantly smaller at the 10-percent level than the average of the dot-count ground-truth wheat proportions.

Winter wheat area estimates were generally in excess of their USDA/SRS counterparts. Small relative differences at the USGP-7 level in June and July resulted from overestimates in Colorado and South Dakota, cancelling the underestimate in Oklahoma. The redesignation of segments eliminated the problem in South Dakota, and the thresholding procedure appears to have solved the underestimation problem in Oklahoma. A special investigation into the South Dakota winter wheat overestimate is contained in section 6.3 of this report.

The winter wheat blind site study showed that the average proportion estimates are significantly different from the average dot-count ground-truth proportions at the USSGP and USGP-7 levels.

In Phase III, sampling appears to contribute slightly more to the variability of the area estimator than does classification; however, the CV for the total wheat area due to sampling in the USGP is only 1.9 percent, which is well within the sampling accuracy goal of 2.3 percent. Also, there is less variability in the winter wheat area estimates than in the spring wheat area estimates.

The LACIE estimate of the total wheat yield for the USGP was consistently below that of the USDA/SRS with the relative difference varying between -9.9 and -11.3 percent during the crop year. This underestimate resulted from underestimates for both winter wheat (October relative difference = -9.0 percent) and spring wheat (October relative difference = -14.1 percent) for the USGP-7 and USNGP regions, respectively. The CV's for the LACIE total wheat (USGP) yield estimate were not available until October. A test of the difference between the October LACIE and USDA/SRS yield estimates for the USGP region showed that difference to be significant at the 10-percent level.

The last section of this report consists of a comparison between the LACIE and the U.S.S.R. Task Force estimates of production, area, and yield for U.S.S.R. winter, spring, and total wheat. This analysis revealed steady improvement in the comparison of the production estimates during August, September, and October. By October, there were no significant differences (at the 10-percent level) between the LACIE and the U.S.S.R. Task Force winter, spring, or total wheat production estimates.

3. ASSESSMENT OF PRODUCTION ESTIMATION

This section contains an evaluation of LACIE performance relative to meeting the 90/90 criterion. It also includes a comparison of LACIE and USDA/SRS production estimates for winter wheat, spring wheat, and total wheat.

3.1 THE 90/90 CRITERION

The LACIE accuracy goal for the USGP region is a 90/90 at-harvest criterion for wheat production. This specifies that the at-harvest wheat production estimate for the USGP region be within 10 percent of the true production with a probability of at least 0.90 for any given year.

Let $\hat{P}$ be the LACIE at-harvest estimate of wheat production for the USGP and let P be the true wheat production for the USGP. The 90/90 criterion may be expressed by the following probability statement:

$$\Pr[|\hat{P} - P| \leq 0.1P] \geq 0.90 \qquad (3\text{-}1)$$

It is reasonable to assume for large sample sizes that $\hat{P}$ is normally distributed with mean P + B and variance $\sigma_{\hat{P}}^2$, where B is the bias of the estimator $\hat{P}$. Under this assumption, it is shown in appendix A that equation (3-1) is equivalent to

$$\Phi\left[\frac{0.1 - 1.1\frac{B}{P + B}}{CV(\hat{P})}\right] - \Phi\left[\frac{-0.1 - 0.9\frac{B}{P + B}}{CV(\hat{P})}\right] \geq 0.90 \qquad (3\text{-}2)$$

where Φ represents the cumulative standard normal distribution and $CV(\hat{P})$ is the CV of the estimator $\hat{P}$ defined by

$$CV(\hat{P}) = \frac{\sigma_{\hat{P}}}{E(\hat{P})} = \frac{\sigma_{\hat{P}}}{P + B} \qquad (3\text{-}3)$$

The term $\frac{B}{P + B}$ is called the relative bias of $\hat{P}$.

Inference as to whether the LACIE accuracy goal has been met is made by estimating $\frac{B}{P + B}$ and $CV(\hat{P})$ and then ascertaining whether equation (3-2) is satisfied. Now, $CV(\hat{P})$ is estimated by $\frac{\hat{\sigma}_{\hat{P}}}{\hat{P}}$, where $\hat{\sigma}_{\hat{P}}$ is an estimate of the standard deviation of $\hat{P}$, and $\hat{P}$ is an unbiased estimate of P + B.

Assuming that the USDA/SRS wheat production estimate is the true wheat production P, then $\frac{B}{P + B}$ could be estimated simply by $\frac{\hat{P} - P}{\hat{P}}$.

With the October estimate of relative bias (-9.9 percent) and CV (5.2 percent), the 90/90 goal was not achieved. However, an accuracy of 90/85 was achieved. That is, the probability that the LACIE estimate is within ±15 percent of the true wheat production for the USGP is 0.9.

3.2 COMPARISON OF LACIE AND USDA/SRS PRODUCTION ESTIMATES

Table 3-1 and figure 3-1 show how well LACIE performed relative to the USDA/SRS estimates throughout the crop year. The nine dates for which data are provided correspond to the CAS reports of February 8, April 6, April 22, May 9, June 7, July 11, August 10, September 9, and October 11, 1977. Winter wheat estimates for the USGP-7 states (seven of the nine states of the USGP) are available for each of the above report dates, whereas spring wheat estimates for the four U.S. northern Great Plains (USNGP) states were generated only for the reports of August 10, September 9, and October 11, 1977.

For each major region, a test was performed to determine if the LACIE estimate was significantly different from the corresponding USDA/SRS estimate. The test results are given in the last column of table 3-1. The testing procedure used is described in appendix A.

Because of software problems, statistics were not available for the LACIE production estimates until after the release of the May 9 CMR. Therefore, coefficients of variation (CV's) and tests of significance were available only for those estimates released after May 9, 1977.

TABLE 3-1.— COMPARISON OF LACIE AND USDA/SRS PRODUCTION ESTIMATES

Region	n/M	Production				[c]Relative difference		Value of test statistic
		[b]USDA/SRS	LACIE					
		Estimate (bu × 10^3)	Estimate (bu × 10^3)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
February 8, 1977								
WINTER WHEAT								
Colorado	21/32	60280	49772	a	33	-21.1	37.0	
Kansas	65/84	356400	194220	a	17	-83.5	-26.9	
Nebraska	31/35	99000	90058	a	23	-9.9	39.2	
Oklahoma	27/40	132600	64391	a	29	-105.9	-41.1	
Texas	34/49	98400	56762	a	28	-73.5	-26.9	
[d]USSGP	178/240	746680	455167	a	11	-64.0	-4.9	a
Montana	25/60	79300	73799	a	a	-7.5	a	
S. Dakota	12/33	13920	28513	a	a	51.2	a	
[e]MW states	37/93	93220	102312	a	a	8.9	a	a
[f]USGP-7	215/333	839900	557480	a	a	-50.7	a	a
April 6, 1977								
WINTER WHEAT								
Colorado	27/32	60280	48659	a	33	-23.9	37.0	
Kansas	93/121	356400	187644	a	17	-89.9	-26.9	
Nebraska	48/67	99000	88444	a	23	-11.9	39.2	
Oklahoma	40/46	132600	63918	a	29	-107.5	-41.1	
Texas	27/38	98400	63305	a	28	-55.4	-26.9	
[d]USSGP	235/304	746680	451970	a	11	-65.2	-4.9	a
Montana	40/80	79300	60723	a	a	-30.6	a	
S. Dakota	22/56	13920	46978	a	a	70.4	a	
[e]MW states	62/136	93220	107701	a	a	13.4	a	a
[f]USGP-7	297/440	839900	559672	a	a	-50.1	a	a

n = number of segments used.
M = number of segments allocated.

[a]Data not available.

[b]USDA/SRS prediction through April 22 released on December 22, 1976.

[c]Relative difference = $\left(\frac{\text{LACIE} - \text{USDA/SRS}}{\text{LACIE}} \times 100\right)\%$.

[d]U.S. southern Great Plains region.

[e]The mixed wheat states, Montana and S. Dakota.

[f]Seven-state winter wheat region of U.S. Great Plains.

*The LACIE estimate is significantly different from the USDA/SRS estimate at the 10-percent level.

[N]The LACIE estimate is not significantly different from the USDA/SRS estimate at the 10-percent level.

[g]The pure spring wheat states, Minnesota and N. Dakota.

[h]U.S. northern Great Plains region.

[i]U.S. Great Plains region.

TABLE 3-1.— Continued.

Region	n/M	Production					Relative difference[c]		Value of test statistic
		USDA/SRS[b]	LACIE						
		Estimate (bu × 10^3)	Estimate (bu × 10^3)	CV (%)	1976 CV (%)		1977 (%)	1976 (%)	
April 22, 1977									
WINTER WHEAT									
Colorado	27/32	60280	49037	a	33		-22.9	37.0	
Kansas	94/121	356400	190941	a	17		-86.7	-26.9	
Nebraska	48/67	99000	96579	a	23		-2.5	39.2	
Oklahoma	41/46	132600	64413	a	29		-105.9	-41.1	
Texas	29/38	98400	63516	a	28		-54.9	-26.9	
USSGP[d]	239/304	746680	464486	a	11		-60.8	-4.9	a
Montana	40/80	79300	65712	a	a		-20.7	a	
S. Dakota	22/56	13920	46057	a	a		69.8	a	
MW states[e]	62/136	93220	111769	a	a		16.6	a	a
USGP-7[f]	301/440	839900	576255	a	a		-45.8	a	a
May 9, 1977									
WINTER WHEAT									
Colorado	28/32	54960	70357	a	31		21.9	24.4	
Kansas	109/121	384000	286373	a	12		-34.1	-6.8	
Nebraska	48/67	103700	99038	a	19		-4.7	14.6	
Oklahoma	45/46	162500	95560	a	21		-70.1	-43.8	
Texas	34/38	101200	83068	a	17		-21.8	19.2	
USSGP[d]	264/304	806360	634396	a	8		-27.1	-1.6	a
Montana	41/80	75600	85751	a	a		11.8	a	
S. Dakota	24/56	15000	58836	a	a		74.5	a	
MW states[e]	65/136	90600	144587	a	a		37.3	a	a
USGP-7[f]	329/440	896960	778982	a	a		-15.1	a	a

TABLE 3-1.— Continued.

Region	n/M	Production					Relative difference[c]		Value of test statistic
		USDA/SRS[b]	LACIE						
		Estimate (bu × 10^3)	Estimate (bu × 10^3)	CV (%)	1976 CV (%)		1977 (%)	1976 (%)	
June 7, 1977									
WINTER WHEAT									
Colorado	28/32	56640	72456	21.9	28		21.8	31.7	
Kansas	112/121	396000	308387	11.5	11		-28.4	14.4	
Nebraska	50/67	106750	108793	16.2	17		1.9	24.4	
Oklahoma	45/46	169000	96550	14.0	17		-75.0	-34.4	
Texas	34/38	110000	91965	14.2	17		-19.6	16.5	
USSGP[d]	269/304	838390	678151	6.9	7		-23.6	11.4	-3.42*
Montana	41/80	75600	91417	23.2	192		17.3	-569.8	
S. Dakota	28/56	13600	67685	38.3	46		79.9	34.1	
MW states[e]	69/136	89200	159102	21.1	63		43.9	-147.1	-2.08*
USGP-7[f]	338/440	927590	837254	7.0	8		-10.8	1.7	-1.54[N]
July 11, 1977									
WINTER WHEAT									
Colorado	30/32	54280	66516	19.7	30		18.4	6.0	
Kansas	111/121	381300	339348	10.9	11		-12.4	3.7	
Nebraska	52/67	106750	111903	15.7	16		4.6	27.3	
Oklahoma	42/46	169000	104907	13.6	18		-61.1	-64.3	
Texas	34/38	115000	91691	13.9	17		-25.4	-22.2	
USSGP[d]	269/304	826330	714365	a	7		15.7	-3.7	a
Montana	58/80	75600	81983	17.2	53		7.8	-211.2	
S. Dakota	39/56	16320	123196	22.6	27		86.8	63.1	
MW states[e]	97/136	91920	205179	a	27		55.2	-46.7	a
USGP-7[f]	366/440	918250	919544	6.4	7		0.1	-7.9	0.02[N]

TABLE 3-1.— Continued.

Region	n/M	Production				[c]Relative difference		Value of test statistic
		[b]USDA/SRS	LACIE					
		Estimate (bu × 10^3)	Estimate (bu × 10^3)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
August 10, 1977								
WINTER WHEAT								
Colorado	31/32	54280	68682	18.8	29	21.0	3.2	
Kansas	105/121	350550	357263	10.8	10	1.9	3.1	
Nebraska	40/67	106750	109960	17.0	16	2.9	26.5	
Oklahoma	44/46	175500	110463	13.4	18	-58.9	-54.0	
Texas	33/38	117500	87579	17.7	18	-34.2	-28.2	
[d]USSGP	253/304	804580	733947	6.8	7	-9.6	-4.2	-1.41[N]
Montana	51/80	75600	72678	15.4	36	-4.0	-73.2	
S. Dakota	18/56	18360	36621	42.5	26	49.9	56.2	
[e]MW states	69/136	93960	109299	17.6	23	14.0	-15.4	0.80[N]
[f]USGP-7	322/440	898540	843247	6.4	7	-6.6	-5.6	-1.03[N]
SPRING WHEAT								
Minnesota	40/58	130954	71199	18.1	42	-83.9	-120.8	
N. Dakota	63/103	238250	157751	14.4	17	-51.0	-20.6	
[g]SW states	103/161	369204	228950	12.3	16	-61.3	-40.4	-4.98*
Montana	35/80	50050	24634	22.8	29	-103.2	-116.2	
S. Dakota	29/56	58168	45103	18.3	18	-29.0	44.6	
MW states	65/136	108218	69737	14.3	17	-55.2	-26.6	-3.86*
[h]USNGP	167/297	477422	298686	10.0	13	-59.8	-37.8	-5.98*
TOTAL WHEAT								
Montana	62/80	125650	97312	14.3	20	-29.1	-88.0	
S. Dakota	38/56	76528	81724	18.5	14	6.4	51.0	
MW states	100/136	202178	179036	11.5	12	-12.9	-19.8	-1.12[N]
USNGP	203/297	571382	407986	9.0	11	-40.0	-32.7	-4.44*
[i]USGP	456/601	1375962	1141933	5.4	6	-20.5	-15.3	-3.80*

TABLE 3-1.— Continued.

Region	n/M	Production				Relative difference[c]		Value of test statistic
		USDA/SRS[b]	LACIE					
		Estimate (bu × 10^3)	Estimate (bu × 10^3)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
September 9, 1977								
WINTER WHEAT								
Colorado	25/32	54280	68675	17.9	29	21.0	8.5	
Kansas	107/121	350550	360616	10.6	10	2.8	3.7	
Nebraska	44/67	106750	99264	14.9	16	-7.5	13.5	
Oklahoma	38/46	175500	121671	12.4	18	-44.2	-56.7	
Texas	30/38	117500	91594	15.9	18	-28.3	-27.2	
USSGP[d]	244/304	804580	741820	6.5	7	-8.5	-6.6	-1.31[N]
Montana	39/80	78400	95206	14.1	30	17.7	-53.7	
S. Dakota	13/56	18360	28130	40.2	26	34.7	57.0	
MW states[e]	52/136	96760	123336	14.2	21	21.5	-7.0	1.51[N]
USGP-7[f]	296/440	901340	865156	6.0	7	-4.2	-6.6	-0.07[N]
SPRING WHEAT								
Minnesota	37/58	130954	78744	18.7	29	-66.3	-68.7	
N. Dakota	60/103	228720	200529	13.1	12	-14.1	-14.9	
SW states[g]	97/161	359674	279273	11.6	11	-28.8	-27.1	-2.48*
Montana	30/80	48070	39357	18.6	25	-22.1	-86.5	
S. Dakota	30/56	55968	44969	17.3	19	-24.5	32.3	
MW states	60/136	104038	84326	12.6	15	-23.4	-26.4	-1.86*
USNGP[h]	157/297	463712	363599	9.4	10	-27.5	-27.0	-2.93*
TOTAL WHEAT								
Montana	53/80	126470	134563	13.7	15	6.0	-65.5	
S. Dakota	36/56	74328	73098	17.2	13	-1.7	46.1	
MW states	89/136	200798	207661	10.8	10	3.3	-14.7	0.3[N]
USNGP	186/297	530472	476935	9.0	10	-17.5	-22.8	-1.9*
USGP[i]	430/601	1365052	1228755	5.3	5	-11.1	-13.6	-2.1*

TABLE 3-1.— Concluded.

Region	n/M	Production				Relative difference[c]		Value of test statistic
		USDA/SRS[b]	LACIE					
		Estimate (bu × 10^3)	Estimate (bu × 10^3)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
October 11, 1977								
WINTER WHEAT								
Colorado	23/31	54280	77070	17.6	29	29.6	8.5	
Kansas	108/121	350550	365465	10.5	10	4.1	3.7	
Nebraska	40/56	106750	106120	13.3	16	0.6	13.5	
Oklahoma	39/46	175500	119208	12.7	18	-47.2	-56.7	
Texas	28/35	117500	92885	15.0	18	-26.5	-27.2	
USSGP[d]	238/289	804580	760748	6.4	7	-5.8	-6.6	-0.91^N
Montana	42/58	78400	90411	14.3	29	13.3	-51.6	
S. Dakota	14/21	18360	26072	30.6	26	29.6	57.0	
MW states[e]	56/79	96760	116483	13.0	20	16.9	-6.1	1.30^N
USGP-7[f]	294/368	901340	877231	5.8	7	-2.7	-6.5	-0.47^N
SPRING WHEAT								
Minnesota	37/47	124714	73213	13.9	32	-70.3	-89.7	
N. Dakota	70/103	229985	211247	13.1	12	-8.9	-10.1	
SW states[g]	107/150	354699	284460	11.2	11	-24.7	-26.2	-2.21*
Montana	33/48	50665	38683	17.4	25	-31.0	-65.7	
S. Dakota	32/37	55968	39748	16.4	18	-40.8	31.9	
MW states	65/85	106633	78431	11.9	16	-36.0	-19.8	-3.03*
USNGP[h]	172/235	461332	362890	9.1	10	-27.1	-24.9	-2.98*
TOTAL WHEAT								
Montana	57/73	129065	129094	13.5	13	0.0	-56.9	
S. Dakota	38/45	74328	65820	16.3	13	-12.9	46.0	
MW states	95/118	203393	194914	10.5	9	-4.4	-11.7	-0.42^N
USNGP	202/274	558092	479373	8.8	8	-16.4	-20.9	-1.86*
USGP[i]	440/557	1362672	1240121	5.2	5	-9.9	-12.8	-1.90*

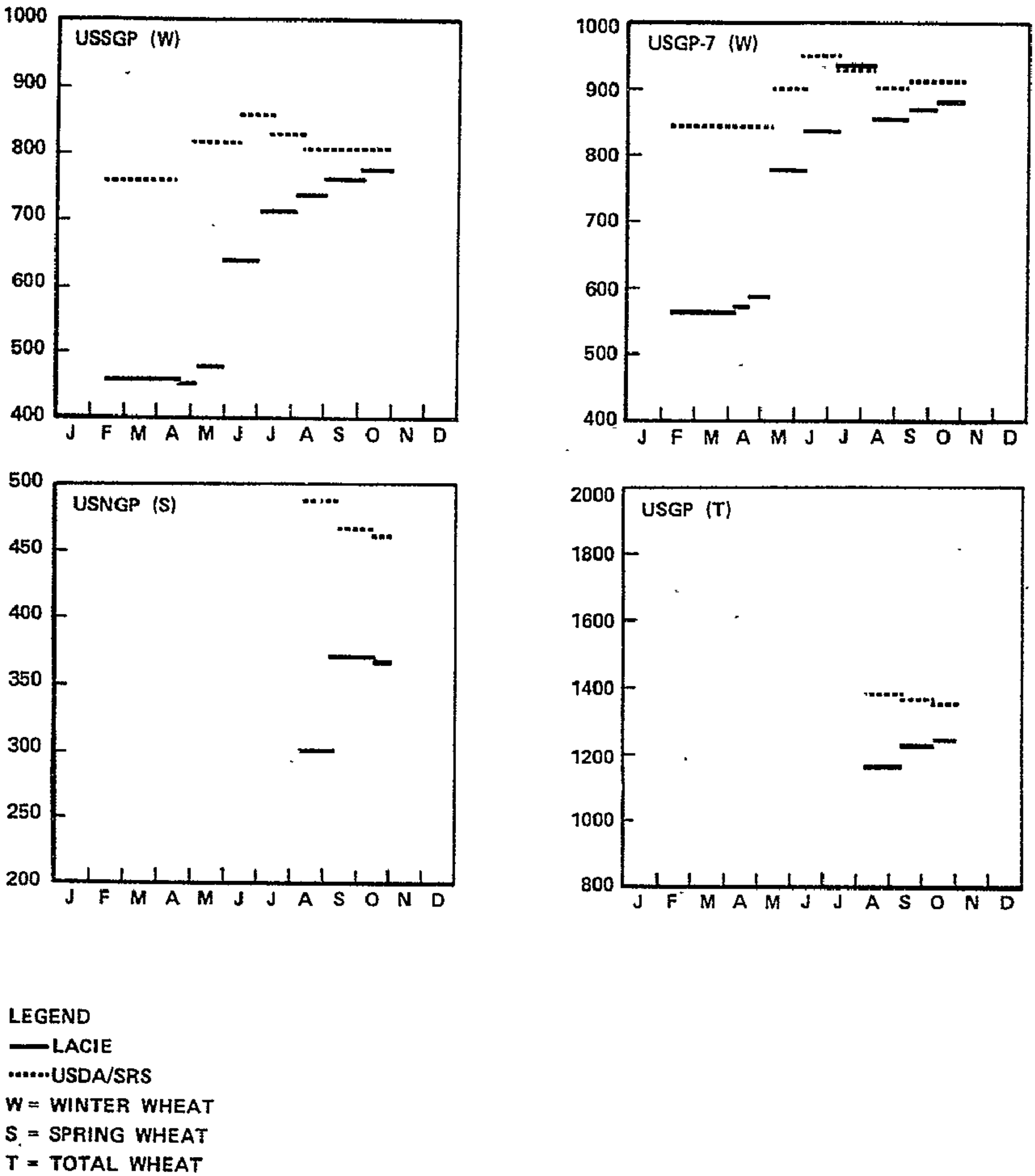

Figure 3-1.— LACIE and USDA/SRS production estimates (bushels × 10^6). (USDA/SRS estimates through April 22 released on December 22, 1976.)

At the U.S. southern Great Plains (USSGP) level, the LACIE and USDA/SRS winter wheat production estimates differed by more than 300 million bushels in February but converged steadily after July and differed by less than 44 million bushels in October. This trend is most obvious in figure 3-1. It is worth noting that the LACIE estimate has experienced steady growth since the first aggregation of the Phase III allocation on April 6, whereas the USDA/SRS figure increased through June but decreased in July and August. The relative difference between the LACIE and USDA/SRS production estimates at the USSGP level has decreased in magnitude with each aggregation since April 6, the first aggregation using the Phase III allocation. Statistics were not available for LACIE USSGP production estimates through May or for the month of July. The difference between the LACIE and USDA/SRS estimates of winter wheat production for the USSGP region was significant (at the 10-percent level) in June but not in August, September, or October.

The LACIE and USDA/SRS production estimates for the USGP-7 region followed a pattern very similar to that of the USSGP estimates except in the month of July, when the winter wheat area for South Dakota was grossly overestimated by LACIE, resulting in a large production overestimate. The problem was corrected in August when the LACIE estimate dropped to approximately the June level and resumed its approach toward the higher USDA/SRS estimate in September and October. Since the June 7 CMR (the first month with statistics available) there has been no significant difference (at the 10-percent level) between the LACIE and USDA/SRS winter wheat production estimates for the USGP-7 region.

The first LACIE estimates of 1977 spring wheat production were made available in the August 10 CMR. The relative difference between the LACIE and USDA/SRS estimates for the four-state USNGP spring wheat producing region decreased in magnitude in each successive CMR because of increases in the LACIE estimate and decreases in the USDA/SRS estimate. The difference between the LACIE and USDA/SRS USNGP spring wheat production estimates was significant at the 10-percent level for each of the three spring wheat aggregations as a result of a large LACIE underestimate (as compared to the USDA/SRS estimate) for Minnesota and moderate underestimates for the other three states.

The LACIE USGP total wheat production estimates in August, September, and October were significantly smaller than the corresponding USDA/SRS estimates as a result of the underestimate in the USNGP spring wheat region. The magnitude of the relative difference between the two estimates decreased steadily during the three reporting months because of increases in the LACIE estimate and decreases in the USDA/SRS estimate.

The CV's at regional levels for both winter and spring wheat production estimates gradually decreased, indicating improvement in accuracy of the LACIE production estimates.

4. ASSESSMENT OF AREA ESTIMATION

Three major subjects are discussed in this section: (1) a comparison of LACIE and USDA/SRS wheat area estimates (section 4.1); (2) a blind site investigation of proportion estimation error (section 4.2); and (3) a discussion of classification and sampling errors (section 4.3).

4.1 COMPARISON OF LACIE AND USDA/SRS AREA ESTIMATES

The LACIE and USDA/SRS area estimates are shown in figure 4-1 and table 4-1. Since the statistics published in the February, April, and May CAS reports were in error because of a software problem, statistical inferences are not given here for the data in these reports.

The LACIE winter wheat area estimate at the five-state USSGP level increased steadily during the season after recording a small decrease in the April 6 estimate (the first estimate using the Phase III allocation). Large negative relative differences recorded in February and April are due to the comparison of LACIE estimates of harvestable winter wheat area with USDA/SRS estimates of planted winter wheat area. Since May, however, the relative difference between the two estimates has ranged from -12.0 to +3.5 percent, improving steadily over the 5-month period except for October.

Included in figure 4-1 is a plot of the Oklahoma winter wheat area estimate, which recovered from a -164.1 percent relative difference in February to -9.2 percent at the end of the season in October. Before May, the relative difference was large because of low LACIE estimates and a high USDA/SRS estimate. The USDA/SRS estimate is expected to be high at this time because it is for planted (rather than harvested) wheat. In May, the relative difference improved (to -44.3 percent) due to an increase in the LACIE estimate and a decrease in the USDA/SRS estimate. The decreased USDA/SRS estimate is an estimate of harvested wheat. This estimate remained the same for the rest of the season. The LACIE estimate steadily approached this USDA/SRS estimate from May until September and then decreased slightly in October. There was no significant difference (at the 10-percent level) between LACIE and USDA/SRS USSGP

TABLE 4-1.— COMPARISON OF LACIE AND USDA/SRS AREA ESTIMATES

Region	n/M	Area				[c]Relative difference		Value of test statistic
		[b]USDA/SRS	LACIE					
		Estimate (ac × 10^3)	Estimate (ac × 10^3)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
February 8, 1977								
WINTER WHEAT								
Colorado	21/32	2740	2183	a	26	-25.5	20.0	
Kansas	65/84	13200	6719	a	12	-96.5	-63.5	
Nebraska	31/35	3300	2977	a	18	-10.8	24.4	
Oklahoma	27/40	7800	2953	a	24	-164.1	-90.0	
Texas	34/49	6150	2954	a	25	-108.2	-98.7	
[d]USSGP	178/240	33190	17786	a	9	-86.6	-46.0	a
Montana	25/60	3050	2763	a	a	-10.4	a	
S. Dakota	12/33	1160	1044	a	a	-11.1	a	
[e]MW states	37/93	4210	3807	a	a	-10.6	a	a
[f]USGP-7	215/333	37400	21594	a	a	-73.2	a	a
April 6, 1977								
WINTER WHEAT								
Colorado	27/32	2740	2135	a	26	-28.3	20.0	
Kansas	93/121	13200	6491	a	12	-103.4	-63.5	
Nebraska	48/67	3300	2892	a	18	-14.1	24.4	
Oklahoma	40/46	7800	2943	a	24	-165.0	-90.0	
Texas	27/38	6150	3294	a	25	-86.7	-98.7	
[d]USSGP	235/304	33190	17755	a	9	-86.9	-46.0	a
Montana	40/80	3050	2274	a	a	-34.1	a	
S. Dakota	22/56	1160	1721	a	a	32.6	a	
[e]MW states	62/136	4210	3995	a	a	-5.4	a	a
[f]USGP-7	297/440	37400	21750	a	a	-72.0	a	a

n = number of segments used.
M = number of segments allocated.

[a]Data not available.

[b]USDA/SRS prediction through April 22 released on December 22, 1976.

[c]Relative difference = $\left(\frac{\text{LACIE - USDA/SRS}}{\text{LACIE}} \times 100\right)\%$.

[d]U.S. southern Great Plains region.

[e]The mixed wheat states, Montana and S. Dakota.

[f]Seven-state winter wheat region of U.S. Great Plains.

*The LACIE estimate is significantly different from the USDA/SRS estimate at the 10-percent level.

[N]The LACIE estimate is not significantly different from the USDA/SRS estimate at the 10-percent level.

[g]The pure spring wheat states, Minnesota and N. Dakota.

[h]U.S. northern Great Plains region.

[i]U.S. Great Plains region.

TABLE 4-1.— Continued.

Region	n/M	Area					^c^Relative difference		Value of test statistic
		[b]USDA/SRS	LACIE						
		Estimate (ac × 10^3)	Estimate (ac × 10^3)	CV (%)	1976 CV (%)		1977 (%)	1976 (%)	
				April 22, 1977					
WINTER WHEAT									
Colorado	27/32	2740	2189	a	26		-25.2	20.0	
Kansas	94/121	13200	6794	a	12		-94.3	-63.5	
Nebraska	48/67	3300	3072	a	18		-7.4	24.4	
Oklahoma	41/46	7800	3061	a	24		-154.8	-90.0	
Texas	29/38	6150	3517	a	25		-74.9	-98.7	
[d]USSGP	239/304	33190	18633	a	9		-78.1	-46.0	a
Montana	40/80	3050	2274	a	a		-34.1	a	
S. Dakota	22/56	1160	1721	a	a		32.6	a	
[e]MW states	62/136	4210	3995	a	a		-5.4	a	a
[f]USGP-7	301/440	37400	22627	a	a		-65.3	a	a
				May 9, 1977					
WINTER WHEAT									
Colorado	28/32	2290	3093	a	24		26.0	32.3	
Kansas	109/121	12000	10190	a	6		-17.8	-15.0	
Nebraska	48/67	3050	3169	a	13		3.8	19.2	
Oklahoma	45/46	6500	4506	a	16		-44.3	-48.8	
Texas	34/38	4400	4262	a	14		-3.2	18.9	
[d]USSGP	264/304	28240	25220	a	6		-12.0	-3.2	a
Montana	41/80	2800	2973	a	a		5.8	a	
S. Dakota	24/56	750	2261	a	a		66.8	a	
[e]MW states	65/136	3550	5234	a	a		32.2	a	a
[f]USGP-7	329/440	31790	30453	a	a		-4.4	a	a

TABLE 4-1.— Continued.

Region	n/M	Area				Relative difference[c]		Value of test statistic
		[b]USDA/SRS	LACIE					
		Estimate (ac × 10^3)	Estimate (ac × 10^3)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
June 7, 1977								
WINTER WHEAT								
Colorado	28/32	2360	3065	15.8	23	23.0	36.6	
Kansas	112/121	12000	10915	5.8	6	-9.9	-2.0	
Nebraska	50/67	3050	3610	12.1	12	15.5	28.1	
Oklahoma	45/46	6500	4875	9.0	14	-33.3	-39.8	
Texas	34/38	4400	4529	11.9	15	2.8	14.4	
[d]USSGP	269/304	28310	26994	4.2	5	-4.9	3.9	-1.17[N]
Montana	41/80	2800	3253	19.2	193	13.9	518.9	
S. Dakota	28/56	680	2601	34.0	43	73.9	10.3	
[e]MW states	69/136	3480	5854	18.5	65	40.6	-146.5	2.19*
[f]USGP-7	338/440	31790	32848	4.8	6	3.2	-4.9	0.67[N]
July 11, 1977								
WINTER WHEAT								
Colorado	30/32	2360	2962	13.2	25	20.3	23.3	
Kansas	111/121	12300	11764	5.0	6	-4.6	-2.8	
Nebraska	52/67	3050	3475	12.4	11	12.2	27.4	
Oklahoma	42/46	6500	5264	8.5	15	-23.5	-56.5	
Texas	34/38	4600	4511	11.6	15	-2.0	-8.9	
[d]USSGP	269/304	28810	27976	3.9	5	-3.0	-4.5	-0.77[N]
Montana	58/80	2800	3097	12.3	52	9.6	-189.3	
S. Dakota	39/56	680	4629	12.6	23	85.3	29.8	
[e]MW states	97/136	3480	7726	9.0	25	55.0	-60.7	6.11*
[f]USGP-7	366/440	32290	35701	3.6	5	9.6	-9.4	2.67*

TABLE 4-1.— Continued.

Region	n/M	Area				[c]Relative difference		Value of test statistic
		[b]USDA/SRS	LACIE					
		Estimate (ac × 10^3)	Estimate (ac × 10^3)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
August 10, 1977								
WINTER WHEAT								
Colorado	31/32	2360	3059	11.7	24	22.9	22.3	
Kansas	105/121	12300	12385	4.9	5	0.7	-1.5	
Nebraska	40/67	3050	3423	14.0	11	10.9	26.6	
Oklahoma	44/46	6500	5543	8.2	15	-17.3	-46.3	
Texas	33/38	4700	4311	16.1	16	-9.0	-9.0	
[d]USSGP	253/304	28910	28721	4.1	5	-0.7	-3.2	-0.17[N]
Montana	51/80	2800	2746	9.6	35	-2.0	-58.0	
S. Dakota	18/56	680	1353	39.0	23	49.7	29.8	
[e]MW states	69/136	3480	4099	14.4	22	15.1	-19.7	1.05[N]
[f]USGP-7	322/440	32390	32819	4.0	5	1.3	-5.0	0.33[N]
SPRING WHEAT								
Minnesota	40/58	3202	2238	15.3	40	-43.1	-119.8	
N. Dakota	63/103	9530	6761	8.6	14	-41.0	-41.4	
[g]SW states	103/161	12732	8999	7.5	13	-41.5	-55.2	-5.53*
Montana	35/80	2185	1369	18.2	28	-59.6	-105.4	
S. Dakota	29/56	2332	2167	14.2	12	-7.6	5.5	
MW states	65/136	4517	3536	11.2	12	-27.7	-32.4	-2.47*
[h]USNGP	167/297	17249	12535	6.2	10	-37.6	-49.5	-6.06*
TOTAL WHEAT								
Montana	62/80	4985	4115	8.1	19	-21.1	-75.6	
S. Dakota	38/56	3012	3520	13.2	13	14.4	15.4	
MW states	100/136	7997	7635	18.5	11	-4.7	-26.0	-0.25[N]
USNGP	203/297	20729	16634	13.1	9	-24.6	-43.4	-1.88*
[i]USGP	456/601	49639	45355	3.3	5	-9.4	-18.7	-2.85*

TABLE 4-1.— Continued.

Region	n/M	Area: [b]USDA/SRS Estimate (ac × 10^3)	LACIE Estimate (ac × 10^3)	CV (%)	1976 CV (%)	[c]Relative difference 1977 (%)	1976 (%)	Value of test statistic
September 9, 1977								
WINTER WHEAT								
Colorado	25/32	2360	3059	10.3	24	22.9	18.6	
Kansas	107/121	12300	12501	4.5	5	1.6	-1.0	
Nebraska	44/67	3050	3105	11.4	11	1.8	11.7	
Oklahoma	38/46	6500	6074	7.2	14	-7.0	-47.9	
Texas	30/38	4700	4513	14.2	16	-4.1	-8.2	
[d]USSGP	244/304	28910	29252	3.6	5	1.2	-6.2	0.33^N
Montana	39/80	2800	3597	7.3	29	22.2	-43.6	
S. Dakota	13/56	680	1039	36.3	23	34.6	28.4	
[e]MW states	52/136	3480	4636	9.9	20	24.9	-14.2	2.52*
[f]USGP-7	296/440	32390	33888	3.4	5	4.4	-7.2	1.29^N
SPRING WHEAT								
Minnesota	37/58	3202	2461	15.3	27	-30.1	-50.0	
N. Dakota	60/103	9530	8678	4.6	5	-9.8	-19.6	
[g]SW states	97/161	12732	11139	4.9	7	-14.3	-25.9	-2.92*
Montana	30/80	2185	2187	12.2	23	0.1	-79.3	
S. Dakota	30/56	2332	2160	12.9	13	-8.0	2.1	
MW states	60/136	4517	4347	8.9	12	-3.9	-28.9	-0.44^N
[h]USNGP	157/297	17249	15487	4.3	6	-11.4	-26.6	-2.65*
TOTAL WHEAT								
Montana	53/80	4985	5784	6.2	14	13.8	-57.2	
S. Dakota	36/56	3012	3199	11.2	12	5.8	12.9	
MW states	89/136	7997	8983	13.9	9	11.0	-21.4	0.79^N
USNGP	186/297	20729	20123	9.2	6	-3.0	-24.3	-0.33^N
[i]USGP	430/601	49639	49375	2.6	4	-0.5	-13.9	-0.19^N

TABLE 4-1.— Concluded.

Region	n/M	Area				[c]Relative difference		Value of test statistic
		[b]USDA/SRS	LACIE					
		Estimate (ac × 10^3)	Estimate (ac × 10^3)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
October 11, 1977								
WINTER WHEAT								
Colorado	23/31	2360	3432	9.6	24	31.2	18.6	
Kansas	108/121	12300	12669	4.2	5	2.9	-1.0	
Nebraska	40/56	3050	3325	9.5	11	8.3	11.7	
Oklahoma	39/46	6500	5950	7.7	14	-9.2	-47.9	
Texas	28/35	4700	4581	12.9	16	-2.6	-8.2	
[d]USSGP	238/289	28910	29957	3.4	5	3.5	-6.2	1.03[N]
Montana	42/58	2800	3416	7.7	28	18.0	-41.7	
S. Dakota	14/21	680	963	24.8	23	29.4	28.4	
[e]MW states	56/79	3480	4379	8.1	19	20.5	-13.3	2.53*
[f]USGP-7	294/368	32390	34336	3.2	5	5.7	-7.1	1.78*
SPRING WHEAT								
Minnesota	37/47	3202	2289	9.9	30	-39.9	-74.1	
N. Dakota	70/103	9530	9173	4.4	5	-3.9	-18.5	
[g]SW states	107/150	12732	11462	4.0	7	-11.1	-28.8	-2.78*
Montana	33/48	2185	2150	10.3	24	-1.6	-55.7	
S. Dakota	32/37	2332	1909	11.6	13	-22.2	1.4	
MW states	65/85	4517	4059	7.7	12	-11.3	-22.4	-1.47[N]
[h]USNGP	172/235	17249	15521	3.6	6	-11.1	-27.3	-3.08*
TOTAL WHEAT								
Montana	57/73	4985	5566	5.5	12	10.4	-47.5	
S. Dakota	38/45	3012	2872	9.5	12	-4.9	12.5	
MW states	95/118	7997	8438	12.0	8	5.2	-17.8	0.43[N]
USNGP	202/274	20729	19900	7.7	5	-4.2	-24.7	-0.55[N]
[i]USGP	440/557	49639	49857	2.4	4	0.4	-14.1	0.17[N]

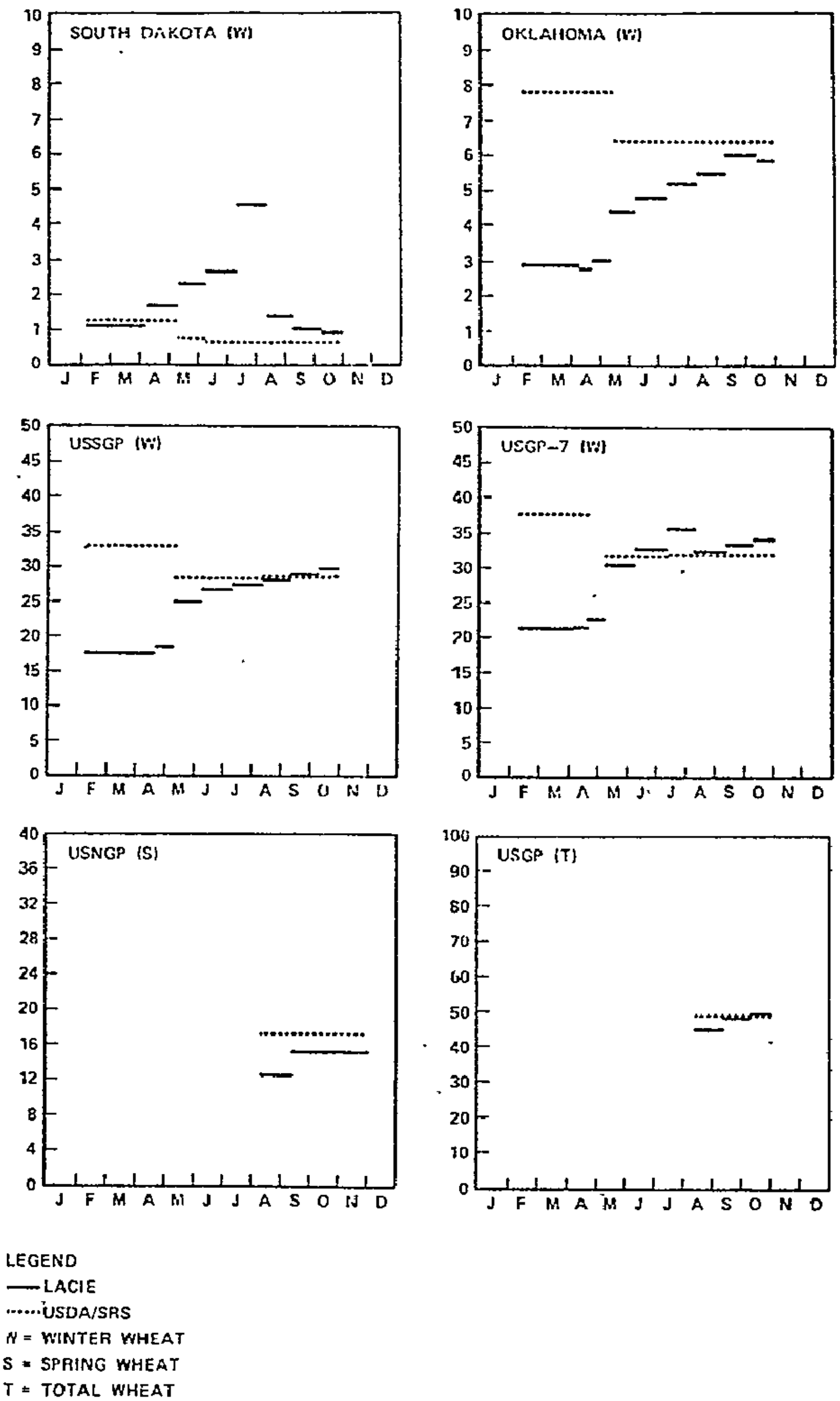

Figure 4-1.— LACIE and USDA/SRS acreage estimates (acres × 10^6). (USDA/SRS estimates through April 22 are of seeded acres, released on December 22, 1976.)

winter wheat area estimates in the months for which statistics were available (June through October).

At the USGP-7 level, the LACIE winter wheat area estimate increased steadily through July, exceeding the corresponding USDA/SRS estimate before decreasing in August because of a drop in the winter wheat area estimate for the mixed wheat states (Montana and South Dakota). The increase of the LACIE estimate through July was due primarily to overestimation in South Dakota (see figure 4-1). The LACIE area estimate for that state grew to almost seven times the corresponding USDA/SRS estimate in July. An investigation of this large overestimation problem was conducted and is reported in section 6.3. As a result of this investigation, the sample segments were redesignated by crop type. The redesignation of sample segments in the August aggregation reduced the LACIE estimate 70 percent and took the relative difference in South Dakota from its July level of 85.3 percent to 49.3 percent.

The differences between the LACIE and USDA/SRS winter wheat area estimates for the USGP-7 region were significant at the 10-percent level in July (because of the South Dakota overestimate) and also in October.

The LACIE USNGP spring wheat area estimate was significantly different from that of the USDA/SRS for each of the three reporting periods (August, September, and October CMR's). Underestimates (as compared to USDA/SRS estimates) were recorded for each of the four states in each of the three CMR's, although an improvement in the comparison was recorded in the September CMR, possibly due to the screening procedure investigated by CAS in that aggregation.

The LACIE and USDA/SRS total wheat area estimates of September and October were not significantly different at the 10-percent level, although those of August were significantly different. The improvement in September can be attributed primarily to the improvement in the LACIE spring wheat area estimate discussed above. The relative difference for the USGP total wheat area estimate decreased in magnitude from -9.4 percent in August to +0.4 percent in October. The area CV's at regional levels decreased gradually. This indicates improvement in the accuracy of the LACIE area estimates.

4.2 BLIND SITE INVESTIGATION OF PROPORTION ESTIMATION ERROR

This section contains a discussion of the wheat proportion estimation error using the blind site wheat estimates and the corresponding dot-count ground-truth proportion estimates for harvested wheat obtained by sampling the ground truth at 400 specified dots (or pixels).

4.2.1 PROPORTION ESTIMATION ERROR

Blind site results for winter wheat and spring wheat are shown in figure 4-2 and tables 4-2 and 4-3. The CAMS proportions used are from the April 22, July 11, and October 11, 1977, CAS reports. The estimates in these reports were chosen because they were the latest proportion estimates from each interim reporting period. Figure 4-2 shows plots of the proportion estimation error $\hat{X} - X$ versus the dot-count ground-truth proportion X, where $\hat{X}$ is the ratioed-down wheat proportion estimate. Plots for the USGP-7 winter wheat producing region are included for the April 22, July 11, and October 11 CAS reports. A plot for the USGP spring wheat producing region is included for the October 11 CAS report. Points lying above the horizontal line $\hat{X} - X = 0$ correspond to overestimates of wheat proportions and points lying below the line correspond to underestimates.

The tendency for CAMS to underestimate by a greater margin for segments with larger proportions of wheat is exhibited by the plots for both winter and spring wheat. It is evident, though, from the three winter wheat plots, that this tendency became less pronounced as the season progressed. This gradual improvement is due primarily to the maturation and eventual harvest of the wheat crop, although those allocation and aggregation modifications described in section 1 of this report also improved the accuracy of the estimates.

Table 4-2 contains the results of the statistical analysis of the data for the April 22 CUR, the July 11 CMR, and the October 11 CMR. The following factors are listed:

- The averaged wheat proportion estimate $\bar{\hat{X}}$
- The averaged dot-count ground-truth wheat proportion estimate $\bar{X}$

TABLE 4-2.— WINTER WHEAT BLIND SITE RESULTS

Region	n/M	$\overline{\hat{X}}$	$\overline{X}$	$\overline{D}$	$S_{\overline{D}}$	90% confidence limits for μ_D
April 22, 1977						
Colorado	11/32	14.4	22.2	-7.8	1.9	(-11.2,-4.4)*
Kansas	19/121	16.0	27.6	-11.6	2.6	(-16.1,-7.1)*
Nebraska	17/67	19.7	16.3	3.4	2.8	(-1.5,8.3)N
Oklahoma	16/46	17.9	36.7	-18.8	3.7	(-25.3,-12.3)*
Texas	5/38	20.1	30.2	-10.2	4.8	(-20.4,0.0)N
USSGP	68/304	17.4	26.2	-8.8	1.7	(-11.6,-6.0)*
Montana	9/80	7.9	14.3	-6.4	1.9	(-9.9,-2.9)*
South Dakota	3/56	1.6	2.0	-0.4	0.8	(-1.9,1.1)N
USGP-7	86/440	14.7	22.3	-7.7	1.4	(-10.0,-5.4)*
July 11, 1977						
Colorado	11/32	16.0	21.4	-5.4	2.2	(-9.4,-1.4)*
Kansas	22/121	21.4	27.1	-5.7	2.0	(-9.1,-2.3)*
Nebraska	19/67	14.1	15.1	-1.0	1.5	(-3.6,1.6)N
Oklahoma	16/46	27.8	36.0	-8.2	2.9	(-13.3,-3.1)*
Texas	6/38	20.3	25.8	-5.5	3.0	(-11.5,0.5)N
USSGP	74/304	20.1	25.0	-5.0	1.1	(-6.8,-3.2)*
Montana	13/80	11.4	14.2	-2.8	1.1	(-4.8,-0.8)*
South Dakota	5/56	3.5	2.8	0.8	1.2	(-1.8,3.4)N
USGP-7	92/440	17.9	22.3	-4.4	0.9	(-5.9,-2.9)*
October 11, 1977						
Colorado	9/31	18.8	22.0	-3.2	2.0	(-6.9,0.5)N
Kansas	21/121	26.3	29.1	-2.8	1.0	(-4.5,-1.1)*
Nebraska	16/56	15.3	17.5	-2.2	1.3	(-4.5,0.1)N
Oklahoma	14/46	34.8	38.2	-3.4	2.8	(-8.4,1.6)N
Texas	6/35	21.6	25.8	-4.2	2.3	(-8.8,0.4)N
USSGP	66/289	24.0	26.9	-3.0	0.8	(-4.3,-1.7)*
Montana	14/58	13.5	13.8	-0.3	1.0	(-2.1,1.5)N
South Dakota	3/21	3.0	3.2	-0.2	0.4	(-1.4,1.0)N
USGP-7	83/368	21.5	23.8	-2.4	0.7	(-3.6,-1.2)*

LEGEND:

n = number of blind sites available
M = number of sample segments allocated

$\overline{\hat{X}}$ = average of wheat proportion estimates
$\overline{X}$ = average of dot-count ground-truth wheat proportion estimates for harvested wheat

$\overline{D} = \overline{\hat{X}} - \overline{X}$

$S_{\overline{D}}$ = standard error of $\overline{D}$

μ_D = population $\overline{D}$

N = μ_D not significantly different from zero at the 10-percent level

* = μ_D significantly different from zero at the 10-percent level

TABLE 4-3.— SPRING WHEAT BLIND SITE RESULTS

[October 11, 1977, CMR]

Region	n/M	$\hat{\bar{X}}$	$\bar{X}$	$\bar{D}$	$S_{\bar{D}}$	90% confidence limits for μ_D
Minnesota	13/47	18.2	21.7	-3.5	2.3	(-7.6,0.6)^N
North Dakota	20/103	21.0	25.1	-4.1	1.5	(-6.7,-1.5)*
Montana	10/48	11.2	14.6	-3.4	2.1	(-7.2,0.4)^N
South Dakota	10/37	8.4	11.0	-2.6	2.2	(-6.6,1.4)^N
USNGP(S)	53/235	16.1	19.6	-3.5	1.0	(-5.2,-1.8)*

LEGEND:

* = μ_D significantly different from zero at the 10-percent level

N = μ_D not significantly different from zero at the 10-percent level

n = number of blind sites available

M = number of sample segments allocated

$\hat{\bar{X}}$ = average of wheat proportion estimates

$\bar{X}$ = average of dot-count ground-truth wheat proportion estimates for harvested wheat

$\bar{D} = \hat{\bar{X}} - \bar{X}$

$S_{\bar{D}}$ = standard error of $\bar{D}$

μ_D = population $\bar{D}$

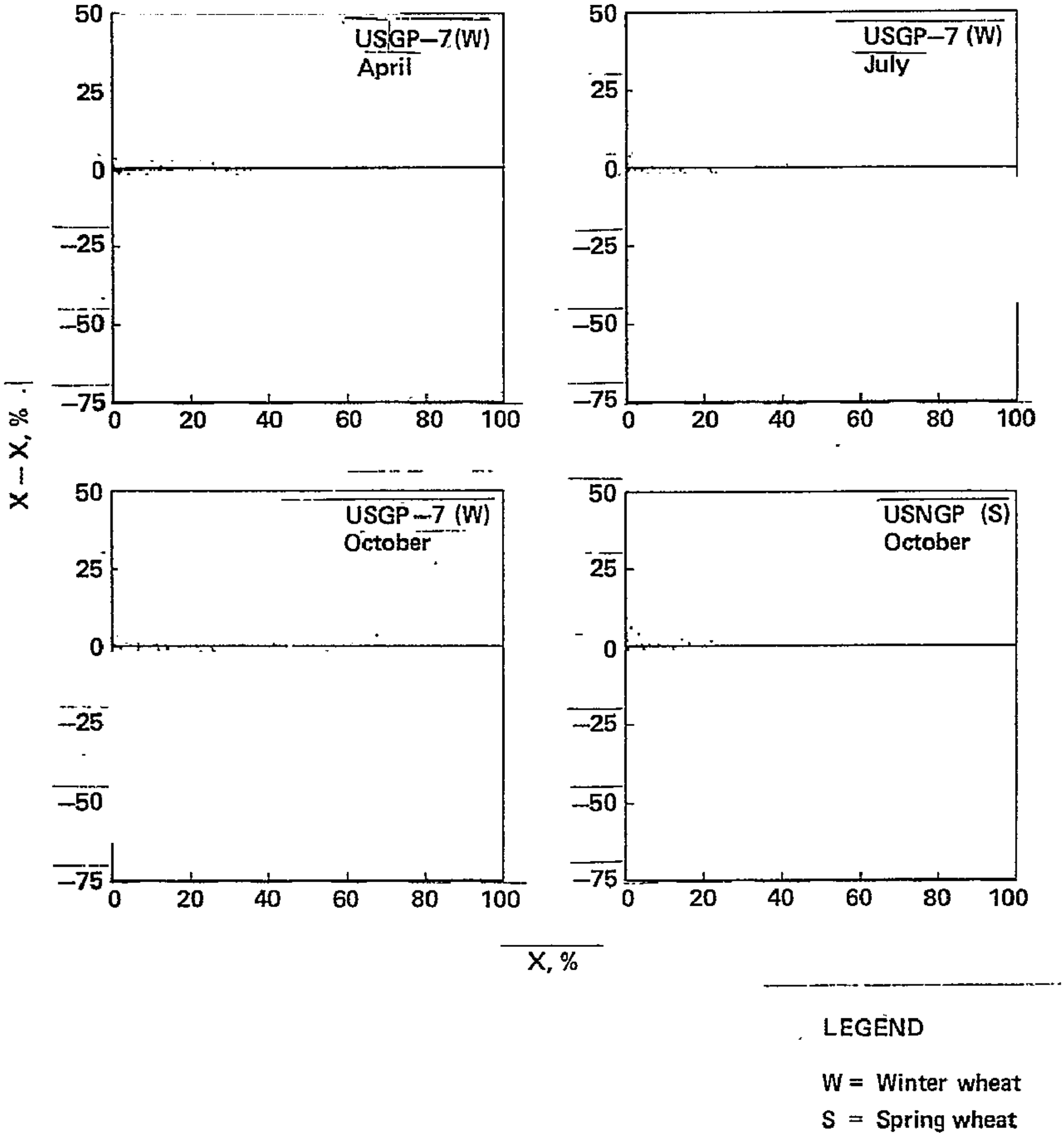

Figure 4-2.— Plot of proportion estimation errors versus ground truth proportions for blind sites.

- The averaged difference $\bar{D} = X - X$
- The standard error of the averaged difference $S_{\bar{D}}$
- The 90-percent confidence limits for the population averaged difference μ_D

The formulas for calculating these factors are given in appendix A.

To determine whether the average difference for a particular state or region is significantly different from zero, one may simply check whether the corresponding confidence interval contains zero. If it does, the averaged difference is not significantly different from zero; i.e., there is insufficient evidence to conclude that a bias exists due to proportion estimation error. If the confidence interval does not contain zero, the hypothesis for no bias is rejected. The test is performed at the 10-percent level of significance.

In the April 22 blind site results (table 4-2), the proportion estimation error was significantly different from zero in both the USSGP and USGP-7 regions and for all states of the USGP-7 region except Nebraska, Texas, and South Dakota. The average difference in Oklahoma was a particularly large negative value. All averaged differences in the April 22 investigation were negative except for Nebraska. A negative average difference indicates that the average LACIE proportion estimate is less than the average dot-count ground-truth proportion estimate.

Blind site results of the July 11 investigation revealed average differences smaller in magnitude than those for April 22 in both the USSGP and USGP-7 regions and in every state except South Dakota. Earlier in this section, it was noted that as the season progressed the tendency for LACIE to underestimate for segments with larger proportions of wheat was less pronounced. All average differences, except for South Dakota, were negative in the July 11 investigation. The average differences were significantly different from zero in both regions and in all states except Nebraska, Texas, and South Dakota.

Results of the October winter wheat blind site investigation show that the differences between $\hat{X}$ and $\overline{X}$ were again smaller in magnitude that those recorded in July in both winter wheat regions and in every winter wheat state except Nebraska. The differences were significant only in the USSGP and USGP-7 regions and in the state of Kansas. All average differences were negative, indicating underestimation of wheat proportions.

The large underestimation problem in Oklahoma, as shown in the blind site results for April 22 and July 11 data, was remedied in the October 11 blind site results and is no longer considered significant.

Figure 4-2 contains plots for USGP spring wheat blind sites for the October CMR. The USNGP(S) plot shows the tendency for LACIE to underestimate by a greater margin for larger ground-truth proportions, similar to the winter wheat plots.

Table 4-3 is a summary of the spring wheat blind site study corresponding to the USNGP(S) plot. The average classification errors were negative for all four spring wheat states and were significantly different from zero in North Dakota and for the USNGP region.

4.3 SAMPLING AND CLASSIFICATION ERRORS

This study was performed for the purpose of (1) measuring the contributions of classification and sampling errors to the within-stratum area variance and (2) estimating the CV's of the area estimates due to classification and sampling errors.

To estimate the within-stratum area variances due to classification and sampling errors, one first constructs the following three basic regression models:

- True segment proportion versus historical stratum proportion
- LACIE segment proportion versus ground-truth segment proportion
- LACIE segment proportion versus historical stratum proportion

These regression models are used to obtain, respectively, the estimates for (1) the variance contribution due to sampling (often called sampling variance) and the variance of the residuals resulting from regressing the current true stratum proportion onto the historical stratum proportion, (2) the variance contribution due to classification (often called classification variance), and (3) the classification and sampling variances. The maximum likelihood estimation technique, assuming normality, is then used to obtain the optimal estimates for sampling and classification variances. A detailed description of this method is presented in appendix A (section A.3.1.5.1).

When the above-mentioned variance estimates are obtained, the ratio ρ of the within-stratum sampling variance estimate to the total within-stratum area variance estimate can be calculated easily. Assuming that this ratio applies to each zone and each higher region, the variances of the large area estimate due to classification and sampling are given by

$$\hat{\eta}^2 = (1 - \rho)\hat{V}^2 \quad (4\text{-}1)$$

and

$$\hat{\upsilon}^2 = \rho\hat{V}^2 \quad (4\text{-}2)$$

where $\hat{n}^2$, $\hat{v}^2$, and $\hat{V}^2$ denote the classification variance, the sampling variance, and the area variance, respectively, for the large area estimate. Consequently, the estimated CV of a large area estimate $\hat{A}$ due to classification is given by

$$\hat{CV}(\hat{A}/C) = \frac{\hat{n}}{\hat{A}} \qquad (4\text{-}3)$$

and the estimated CV of a large area estimate due to sampling is given by

$$\hat{CV}(\hat{A}/S) = \frac{\hat{v}}{\hat{A}} \qquad (4\text{-}4)$$

where $\hat{CV}(\hat{A}/C)$ and $\hat{CV}(\hat{A}/S)$ are often casually referred to as the classification CV and sampling CV, respectively.

Estimates of these variances and CV's for the October LACIE estimates are tabulated below.

Crop	Within-stratum area variance	Variance component		Percentage error		Classification CV (%)	Sampling CV (%)
		Due to classification	Due to sampling	Due to classification	Due to sampling		
Winter wheat USGP-7	104.1	41.6	62.5	40	60	2.0	2.5
Spring wheat USNGP	65.6	26.2	39.4	40	60	2.3	2.8
Total wheat USGP	100.4	39.6	60.8	40	60	1.5	1.9

These results show that the sampling CV is larger than the classification CV for winter, spring, and total wheat estimates. The indication is that sampling contributes slightly more error to the area variance than does classification. Moreover, winter wheat has smaller CV's for both classification and sampling than does spring wheat; i.e., there is less variability in the winter wheat area estimates than in the spring wheat area estimates. For the USGP region, the sampling CV for the total wheat area estimate is 1.9 percent, which is well within the sampling accuracy goal of 2.3 percent.

5. ASSESSMENT OF YIELD ESTIMATION

This section consists of a comparison of the LACIE and USDA/SRS yield estimates and an assessment of the crop calendar model accuracy.

5.1 COMPARISON OF LACIE AND USDA/SRS YIELD ESTIMATES

Figure 5-1 and table 5-1 present the LACIE and USDA/SRS yield estimates. For the first three CAS reports of Phase III (February 8, April 6, and April 22, 1977), no yield estimates were available from USDA/SRS. AA, for purposes of comparison, "derived" USDA/SRS yield estimates by dividing the USDA/SRS production estimates by the corresponding estimates of planted area. These yield estimates remained unchanged through April at both regional levels (USSGP and USGP-7) since revised estimates were not released by the USDA/SRS until May.

The LACIE estimates of winter wheat yield for the USSGP and USGP-7 regions remained relatively constant throughout Phase III. The LACIE USSGP estimates ranged from 24.9 to 25.6 bushels per acre while the USGP-7 estimates ranged from 25.5 to 25.8 bushels per acre. LACIE estimates were consistently below the corresponding USDA/SRS estimates from May through October in both regions. The differences between the LACIE and USDA/SRS estimates of yield for the USSGP and USGP-7 regions were significant at the 10-percent level in every month for which statistics were available (June through October) except August.

For the winter wheat states, relative differences for Oklahoma and Texas through Phase III indicated a large underestimate when compared to the USDA/SRS estimates. The trend term problem and the questionable precipitation variable in the Center for Climatological and Environmental Assessment (CCEA) yield models, either together or individually, may have contributed to the large underestimation of yield. Specifically, the trend term, which depends on a multitude of factors including irrigation, has been assumed to be zero since 1960 (when the trend curve leveled off) for the Oklahoma and Texas/Oklahoma panhandle models. However, irrigation practices (largely concentrated in the panhandle) began in Texas after 1960. At any given time the

TABLE 5-1.— COMPARISON OF LACIE AND USDA/SRS YIELD ESTIMATES

Region	Yield				[c]Relative difference		Value of test statistic
	[b]USDA/SRS	LACIE					
	Estimate (bu/ac)	Estimate (bu/ac)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
February 8, 1977							
WINTER WHEAT							
Colorado	22.0	22.8	a	21	3.5	21.3	
Kansas	27.0	28.9	a	12	6.6	22.4	
Nebraska	30.0	30.2	a	14	0.7	19.6	
Oklahoma	17.0	21.8	a	17	22.0	34.5	
Texas	16.0	19.2	a	19	16.7	36.2	
[d]USSGP	22.5	25.6	a	7	12.1	28.3	a
Montana	26.0	26.7	a	a	2.6	a	
S. Dakota	12.0	27.3	a	a	56.0	a	
[e]MW states	22.1	26.9	a	a	17.8	a	a
[f]USGP-7	22.5	25.8	a	a	12.8	a	a
April 6, 1977							
WINTER WHEAT							
Colorado	22.0	22.8	a	21	3.5	21.3	
Kansas	27.0	28.9	a	12	6.6	22.4	
Nebraska	30.0	30.6	a	14	2.0	19.6	
Oklahoma	17.0	21.7	a	17	21.7	34.5	
Texas	16.0	19.2	a	19	16.7	36.2	
[d]USSGP	22.5	25.5	a	7	11.8	28.3	a
Montana	26.0	26.7	a	a	2.6	a	
S. Dakota	12.0	27.3	a	a	56.0	a	
[e]MW states	22.1	27.0	a	a	18.1	a	a
[f]USGP-7	22.5	25.7	a	a	12.5	a	a

[a]Data not available.

[b]USDA/SRS estimates through April 22 derived from estimates of seeded acres and production released on December 22, 1976.

[c]Relative difference $= \left(\frac{\text{LACIE - USDA/SRS}}{\text{LACIE}} \times 100\right)\%$.

[d]U.S. southern Great Plains region.

[e]The mixed wheat states, Montana and S. Dakota.

[f]Seven-state winter wheat region of U.S. Great Plains.

*The LACIE estimate is significantly different from the USDA/SRS estimate at the 10-percent level.

[N]The LACIE estimate is not significantly different from the USDA/SRS estimate at the 10-percent level.

[g]The pure spring wheat states, Minnesota and N. Dakota.

[h]U.S. northern Great Plains region.

[i]U.S. Great Plains region.

TABLE 5-1.– Continued.

Region	Yield				[c]Relative difference		Value of test statistic
	[b]USDA/SRS	LACIE					
	Estimate (bu/ac)	Estimate (bu/ac)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
April 22, 1977							
WINTER WHEAT							
Colorado	22.0	22.4	a	21	1.8	21.3	
Kansas	27.0	28.1	a	12	3.9	22.4	
Nebraska	30.0	31.4	a	14	4.5	19.6	
Oklahoma	17.0	21.0	a	17	19.0	34.5	
Texas	16.0	18.1	a	19	11.6	36.2	
[d]USSGP	22.5	24.9	a	7	9.6	28.3	a
Montana	26.0	28.9	a	a	10.0	a	
S. Dakota	12.0	26.8	a	a	55.2	a	
[e]MW states	22.1	28.0	a	a	21.1	a	a
[f]USGP-7	22.5	25.5	a	a	11.8	a	a
May 9, 1977							
WINTER WHEAT							
Colorado	24.0	22.8	a	20	-5.3	-11.7	
Kansas	32.0	28.1	a	10	-13.9	7.0	
Nebraska	34.0	31.3	a	14	-8.6	-6.0	
Oklahoma	25.0	21.2	a	14	-17.9	3.2	
Texas	23.0	19.5	a	13	-17.9	0.6	
[d]USSGP	28.6	25.2	a	6	-13.5	1.6	a
Montana	27.0	28.8	a	a	6.3	a	
S. Dakota	20.0	26.0	a	a	23.1	a	
[e]MW states	25.5	27.6	a	a	7.6	a	a
[f]USGP-7	28.2	25.6	a	a	-10.2	a	a

TABLE 5-1.— Continued.

Region	Yield				[c]Relative difference		Value of test statistic
	[b]USDA/SRS	LACIE					
	Estimate (bu/ac)	Estimate (bu/ac)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
June 7, 1977							
WINTER WHEAT							
Colorado	24.0	23.6	16.9	17	-1.7	-7.8	
Kansas	33.0	28.3	10.6	9	-16.6	16.1	
Nebraska	35.0	30.1	6.6	13	-16.3	-5.1	
Oklahoma	26.0	19.8	5.1	10	-31.3	3.9	
Texas	25.0	20.3	4.9	12	-23.2	2.7	
[d]USSGP	29.6	25.1	4.0	5	-17.9	7.6	-2.98*
Montana	27.0	28.1	14.2	12	3.9	-8.3	
S. Dakota	20.0	26.0	19.2	15	23.1	26.5	
[e]MW states	25.6	27.2	11.0	9	5.9	0	0.54[N]
[f]USGP-7	29.2	25.5	3.9	5	-14.5	6.4	-2.88*
July 11, 1977							
WINTER WHEAT							
Colorado	23.0	22.5	14.8	17	-2.2	-22.2	
Kansas	31.0	28.8	9.7	9	-7.6	6.1	
Nebraska	35.0	32.2	9.3	12	-8.7	0	
Oklahoma	26.0	19.9	10.7	10	-30.7	-4.8	
Texas	25.0	20.3	10.1	12	-23.2	-12.3	
[d]USSGP	28.7	25.5	5.5	5	-12.5	0.8	-2.27*
Montana	27.0	26.5	12.1	9	-1.9	-7.6	
S. Dakota	24.0	26.6	16.9	15	9.8	47.4	
[e]MW states	26.4	26.6	a	9	0.8	8.7	a
[f]USGP-7	28.4	25.8	5.3	5	-10.1	1.1	-1.91*

TABLE 5-1.— Continued.

Region	Yield				[c]Relative difference		Value of test statistic
	[b]USDA/SRS	LACIE					
	Estimate (bu/ac)	Estimate (bu/ac)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
August 10, 1977							
WINTER WHEAT							
Colorado	23.0	22.5	14.8	17	-2.2	-24.3	
Kansas	28.5	28.8	9.7	9	-1.0	4.5	
Nebraska	35.0	32.1	9.7	12	-9.0	0	
Oklahoma	27.0	19.9	10.7	10	-35.7	-5.3	
Texas	25.0	20.3	11.8	20	-23.2	-17.6	
[d]USSGP	27.8	25.6	5.7	5	-8.6	-0.8	-1.51[N]
Montana	27.0	26.5	12.1	9	-1.9	-9.6	
S. Dakota	27.0	27.1	18.5	14	-0.4	37.5	
[e]MW states	27.0	26.7	9.8	8	-1.1	3.4	-0.11[N]
[f]USGP-7	27.7	25.7	5.2	5	-7.8	-0.7	-1.50[N]
SPRING WHEAT							
Minnesota	40.9	31.8	10.4	11	-28.6	-0.3	
N. Dakota	25.0	23.3	12.1	11	-7.3	14.8	
[g]SW states	29.0	25.4	10.0	9	-14.2	9.5	-1.42[N]
Montana	22.9	18.0	14.0	9	-27.2	-5.4	
S. Dakota	24.9	20.8	11.6	14	-19.7	41.4	
MW states	24.0	19.7	8.9	9	-21.8	4.5	-2.45*
[h]USNGP	27.7	23.8	8.0	7	-16.4	7.6	-2.05*
TOTAL WHEAT							
Montana	25.2	23.6	a	4	-6.8	-6.8	
S. Dakota	25.4	23.2	a	5	-9.5	42.0	
MW states	25.3	23.4	a	4	-8.1	4.8	a
USNGP	27.6	24.5	a	6	-12.7	7.4	a
[i]USGP	27.7	25.2	a	4	-9.9	2.6	a

TABLE 5-1.— Continued.

Region	Yield				[c]Relative difference		Value of test statistic
	[b]USDA/SRS	LACIE					
	Estimate (bu/ac)	Estimate (bu/ac)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
September 9, 1977							
WINTER WHEAT							
Colorado	23.0	22.5	14.8	17	-2.2	-12.2	
Kansas	28.5	28.8	9.7	9	1.0	4.5	
Nebraska	35.0	32.0	9.3	12	-9.4	2.1	
Oklahoma	27.0	20.0	10.2	10	-35.0	-6.2	
Texas	25.0	20.3	11.1	5	-23.2	-17.6	
[d]USSGP	27.8	25.4	5.5	5	-9.4	-0.4	-1.71*
Montana	28.0	26.5	12.1	9	-5.7	-7.0	
S. Dakota	27 0	27.1	18.5	14	0.4	39.9	
[e]MW states	27.8	26.6	10.2	8	-4.5	6.2	-0.44[N]
[f]USGP-7	27.8	25.5	5.1	5	-9.0	0.4	-1.76*
SPRING WHEAT							
Minnesota	40.9	32.0	11.0	11	-27.8	-12.5	
N. Dakota	24.0	23.1	12.3	11	-3.9	4.1	
[g]SW states	28.2	25.1	10.4	9	-12.4	-1.1	-1.19[N]
Montana	22.0	18.0	14.0	9	-22.2	-4.0	
S. Dakota	24.0	20.8	11.6	13	-15.4	30.4	
MW states	23.0	19.4	8.8	8	-18.6	1.9	-2.11*
[h]USNGP	26.9	23.5	8.1	7	-14.5	-0.4	-1.79*
TOTAL WHEAT							
Montana	25.4	23.3	a	5	-9.0	-5.2	
S. Dakota	24.7	22.9	a	5	-7.9	38.1	
MW states	25.1	23.1	a	4	-8.7	5.4	
USNGP	27.0	23.7	a	7	-13.9	1.5	a
[i]USGP	27.5	24.7	a	4	-11.3	0.4	a

TABLE 5-1.— Concluded.

Region	Yield				Relative difference[c]		Value of test statistic
	USDA/SRS[b]	LACIE					
	Estimate (bu/ac)	Estimate (bu/ac)	CV (%)	1976 CV (%)	1977 (%)	1976 (%)	
October 11, 1977							
WINTER WHEAT							
Colorado	23.0	22.5	14.8	17	-2.2	-12.2	
Kansas	28.5	28.8	9.7	9	1.0	4.5	
Nebraska	35.0	31.9	9.4	12	-9.7	2.1	
Oklahoma	27.0	20.0	10.2	10	-35.0	-9.3	
Texas	25.0	20.3	11.4	5	-23.2	-17.6	
USSGP[d]	27.8	25.4	5.6	5	-9.4	-0.4	-1.68*
Montana	28.0	26.5	12.1	9	-5.7	-7.0	
S. Dakota	27.0	27.1	18.5	14	0.4	39.9	
MW states[e]	27.8	26.6	10.2	8	-4.5	6.2	-0.44[N]
USGP-7[f]	27.8	25.5	5.1	5	-9.0	0.4	-1.76*
SPRING WHEAT							
Minnesota	38.9	32.0	10.8	11	-21.6	-8.9	
N. Dakota	24.1	23.0	12.4	11	-4.8	7.0	
SW states[g]	27.9	24.8	10.5	9	-12.5	2.2	-1.19[N]
Montana	23.2	18.0	14.0	9	-28.9	-6.3	
S. Dakota	24.0	20.8	11.6	13	-15.4	30.8	
MW states	23.6	19.3	9.1	8	-22.3	2.3	-2.45*
USNGP[h]	26.7	23.4	8.5	7	-14.1	1.9	-1.66*
TOTAL WHEAT							
Montana	25.9	23.2	14.5	5	-11.6	-6.6	
S. Dakota	24.7	22.9	18.9	5	-7.9	38.1	
MW states	25.4	23.1	15.9	4	-10.0	5.4	-0.63[N]
USNGP	26.9	24.1	11.6	6	-11.6	3.0	-1.00[N]
USGP[i]	27.5	24.9	5.7	4	-10.4	1.1	-1.82*

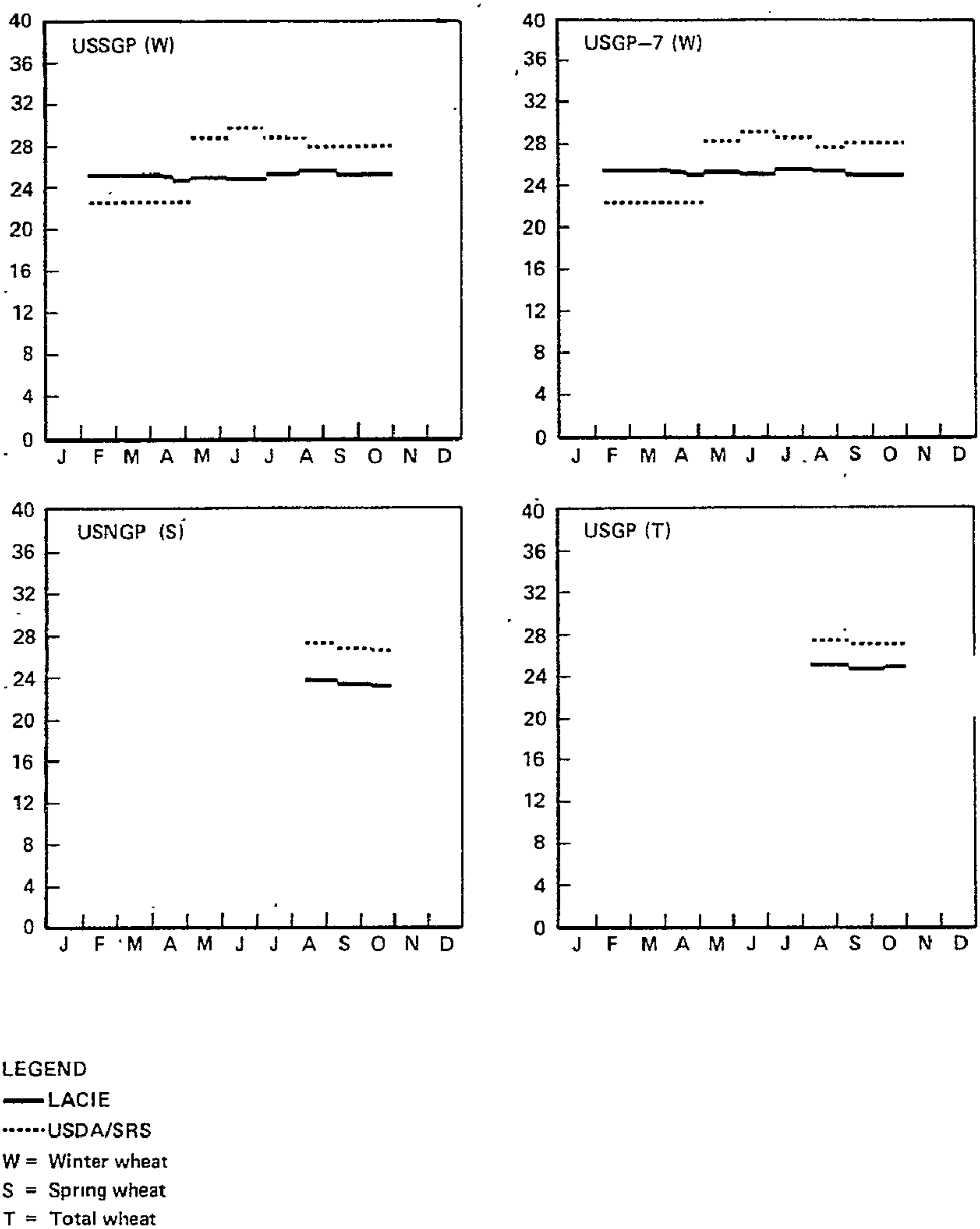

Figure 5-1.— LACIE and USDA/SRS yield estimates (bushels/acre). [USDA/SRS yield estimates through April 22 derived from predicted production and seeded acres estimates released on December 22, 1976.]

yield may vary around the trend curve according to the weather. The three Texas models show that the yields do fall above the CCEA trend for 1977. This indicates that the weather has been good, yet the LACIE estimate is still approximately 20 percent below the USDA/SRS estimate. In addition to the trend factor the May precipitation variable has demonstrated inconsistency in the models. The fact that above-normal rainfall detracted from yield in the Texas low plains and Oklahoma models, but added to yield in the panhandle model, may indicate that the models are not reflecting plant response to the full range of weather over the two states.

For the mixed wheat states, the difference between the LACIE and USDA/SRS estimates of winter wheat yield was not significant in every month for which statistics were available.

The LACIE estimates of yield for the four USNGP spring wheat producing states were consistently below their USDA/SRS counterparts, although the relative difference decreased in magnitude in each successive report. The difference between the LACIE and USDA/SRS yield estimates for the USNGP region was significant at the 10-percent level for each of the three months in which LACIE estimates were available (August, September, and October), although that of the October estimate was only marginally significant. All state-level LACIE spring wheat yield estimates were below their USDA/SRS counterparts.

The LACIE total wheat yield estimates for the USGP region were below the corresponding USDA/SRS estimates in August, September, and October. Total wheat yield statistics were not available in August and September CMR's. The difference between the LACIE and USDA/SRS October total wheat yield estimates for the USGP region was significant at the 10-percent level. The differences between LACIE and SRS yield estimates in Oklahoma, Texas, Minnesota, and Montana were the principal causes for LACIE's failure to meet the 90/90 criteria.

5.2 CROP CALENDAR MODEL ACCURACY

Crop growth stage estimation based on current year weather conditions serves two vital components of LACIE: CAMS and the Yield Estimation Subsystem (YES). Initially, CAMS relies on the crop growth information early in the year to determine whether the wheat is sufficiently emerged to be detectable. Once the Robertson model predicts the crop to have emerged (Robertson stage 2.0) analysis of the segment for wheat percentage is initiated. The winter wheat crop is monitored also to ascertain if it has emerged from dormancy. In some northerly regions of the winter wheat producing states of the USGP, crop estimates are not attempted during dormancy because the canopy is too sparse. The next major growth period of interest to CAMS is the period after dormancy to heading, when the analyst relies on the Robertson crop stage to ascertain the approximate expected intensity of the wheat vegetation signature in comparison to other spring-planted crops. Heading to senescence or maturity is another key stage in the separation of wheat from other vegetation. During this stage, the appearance of the wheat is significantly different from other vegetation types. Senescence to harvest and postharvest are very important to the analyst because the Landsat acquisitions during this period permit him to verify his early-season identifications of wheat. (Wheat, other small grains, and grasses mature and are harvested during this period.)

This very general description of the crop calendar function in CAMS aids in qualitatively understanding the effect of growth stage prediction errors. For example, if the Robertson model predicts full emergence at a date earlier than crops are fully emerged (growth model is ahead of actual progress), CAMS will analyze the segment in a period when some amount (depending on the magnitude of the growth model prediction error) of the wheat is incompletely emerged. Since incompletely emerged wheat fields will go undetected by the analyst, the growth model prediction error can result in a negative bias in the segment proportion estimate. In all cases, if the model predictions run too far ahead of the actual growth stage, the analyst will anticipate an onset of changing signatures within the segment, which will not occur at the predicted rate. Thus, if the growth model predicts 90-percent senescence within the segment and the analyst bases his labeling decision on this fact, certain fields could be discarded as being nonwheat because a senescent signature was expected and the analyst did not observe a change.

Inasmuch as the interactions between the growth model prediction errors and CAMS errors are not fully understood and their relationships to each other remain unquantified, substantial prediction errors in the model could result in substantial errors in analyst labeling.

The currently implemented operational yield models in LACIE do not depend on the crop growth model. However, the response of wheat yield to meteorological conditions is known to depend quite strongly on the growth stage at which these conditions are present. For example, high temperatures after wheat maturity do not affect yields in the same way they do during heading. The second-generation yield models being evaluated for LACIE in Phase III depend on the crop growth models; the effects of certain meteorologically related variables are weighted differently, depending on the estimated growth stage of the plant. Thus, errors in the growth model can strongly influence the yield estimation error; e.g., if high temperatures are experienced the last 2 weeks in May in an area where heading is occurring and the growth model (running fast) is predicting that the crop is ripe, the second-generation yield models will fail to predict the actual reduction in yield.

As stated, the relationship between the growth model prediction errors and the yield estimation errors is not completely understood, and the effects have not been quantified.

The accuracy assessment effort within LACIE has designed an evaluation of the crop growth models, utilizing ground-acquired information from intensive test sites (ITS's) in the yardstick region. This evaluation was conducted over eight winter wheat ITS's in Kansas and Texas during Phase II and was expanded in Phase III to include 22 ITS's throughout the United States and 11 ITS's in Canada (figures 5-2 and 5-3).

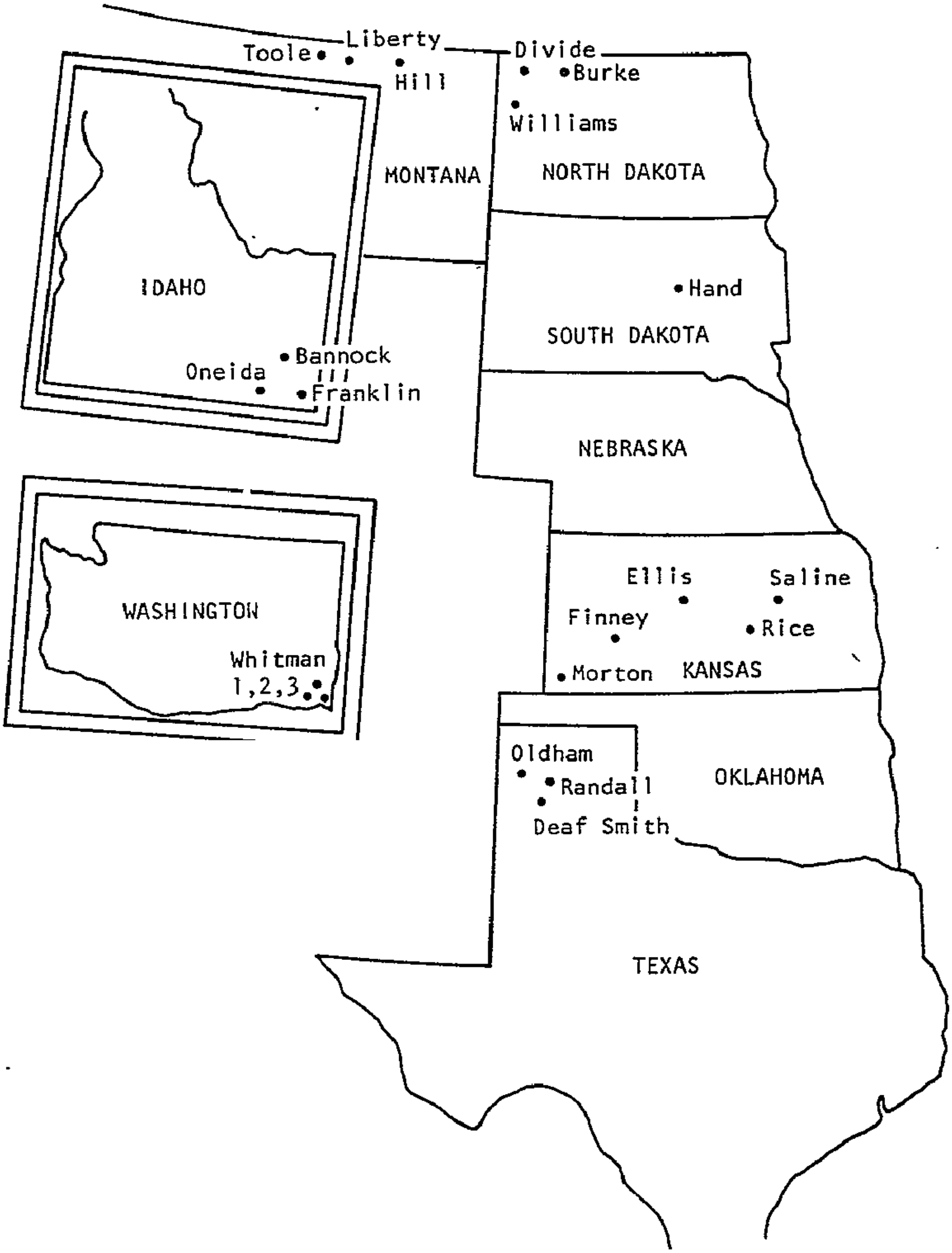

Figure 5-2.— Map of U.S. wheat-producing areas showing intensive test sites.

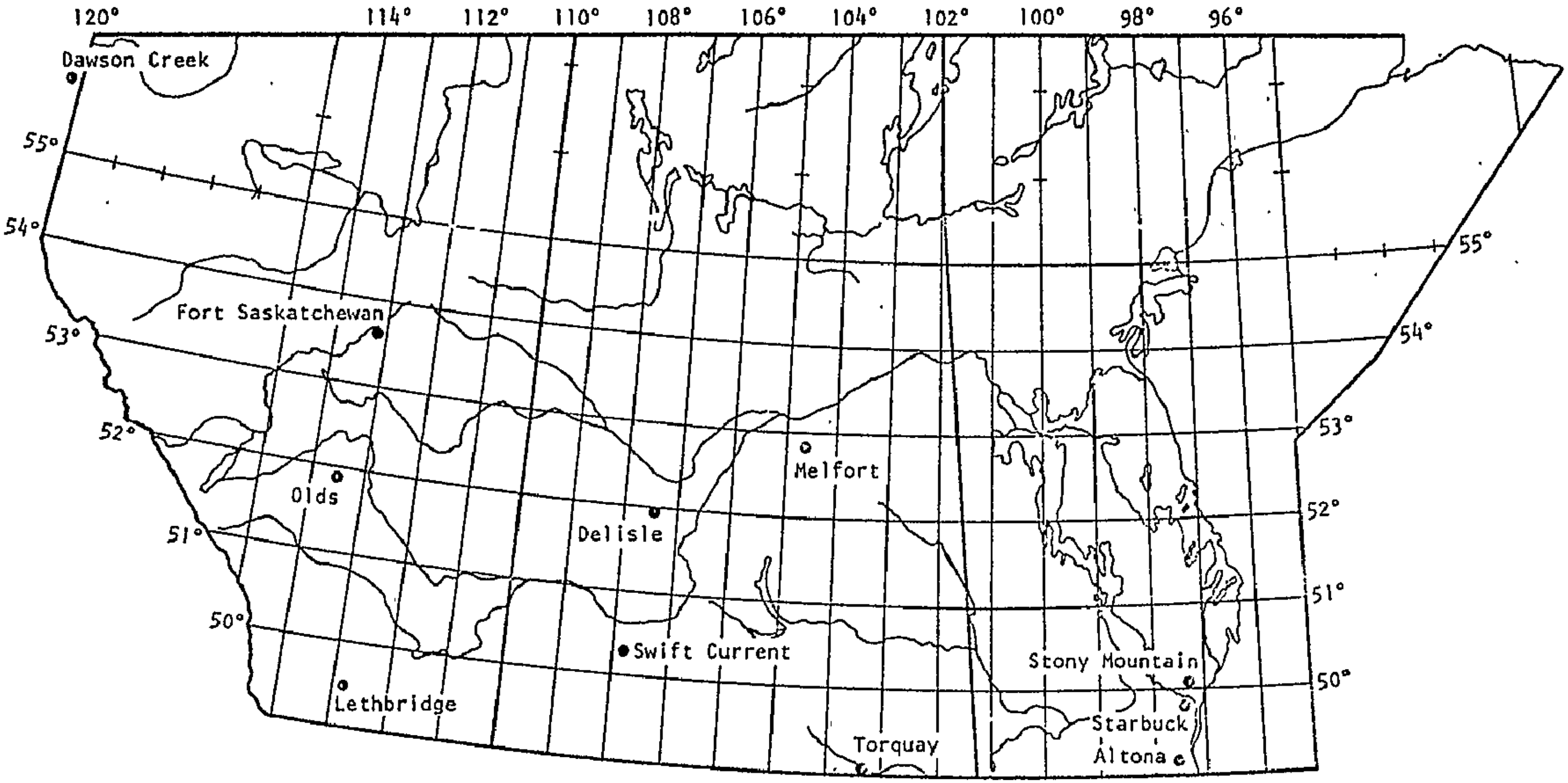

Figure 5-3.— Map of Canada showing intensive test sites.

Within each of these ITS's, the average ground-observed growth stage for the wheat crop is calculated from periodic field-by-field observations obtained by personnel from the USDA Agricultural Stabilization and Conservation Service (ASCS). ASCS personnel record detailed information regarding each field on the form shown in figure 5-4. The observer specifies the growth stage of each field to be one of the 10 stages listed on this form. All sites are visited each 18 days by ASCS field personnel, except for the Finney County, Kansas, and Hand County, South Dakota "supersites," which are visited every 9 days. The 11 ITS's in Canada are monitored each 18 days by personnel from the Canadian Agriculture Department.

The crop calendar model used by LACIE is a modification of the Biometeorological Time Scale (BMTS) developed by Robertson. The Robertson BMTS estimates the stages for the progress of wheat crop development from planting to harvest (table 5-2). Daily maximum and minimum temperatures and day length are variables used to implement this model, which is often referred to as the Adjustable Crop Calendar (ACC).

All of the growth stages defined by Robertson in the BMTS model development are not easily observable by field personnel. For example, BMTS stage 3.0 (jointing) can be observed only by plant dissection. Thus, a different set of stages has been developed for ground observations. The ground-observed growth stage of each ITS must be developed by relating the ITS growth-stage observations to the related BMTS stage. After planting, the earliest stage at which there is no ambiguity in this relationship is at heading. The BMTS stage 3.0 (jointing) is known to occur after tillering and before booting, which are observable by ground personnel. Thus, jointing is estimated by extrapolating between these observations. An error as large as a few days is customary in relating ground observations to BMTS stages. It should be kept in mind that heading is the most valid comparison as the results of the ACC are reviewed.

The ACC is published biweekly in a meteorological summary for all regions being examined by LACIE. The BMTS stages of wheat are based on inputs from

TABLE 5-2.— ROBERTSON BMTS AND OBSERVED ITS WHEAT PHENOLOGICAL STAGES

Stage	Robertson BMTS	ITS growth stage code	Description
Planted	1.0	01	Planted
		02	Planted, no emergence
Emergence	2.0	03	Emergence
Jointing	3.0	04	Tillering, prebooting, pre-budding
	3.5	05	Booted or budded
Heading	4.0	06	Beginning to head or flower
	4.5	07	Fully headed or flowered
Soft dough	5.0	08	Beginning to ripen
Ripening	6.0	09	Ripe to mature
Harvest	7.0	10	Harvest

each reporting meteorological station. These estimates are then utilized to develop BMTS contours as shown in figure 5-5. The ITS BMTS estimate is then determined from its location on this contour map and compared to that determined by ground observations. Such a comparison is shown in figures 5-6 and 5-7 for two ITS's. The standard deviation ($\pm 1\sigma$) of these ground-observed estimates on a field-to-field basis is also shown in these figures. Note in the Oldham County, Texas, example that the ground-computed stage contains the ACC-estimated stage within one standard deviation in the periods from mid-jointing (3.5) to soft dough (5.0). Before 3.5 and after 5.0, the ACC was ahead of the ground truth by a few days and more than one standard deviation. However, in most cases, the ACC BMTS estimate was somewhat more accurate than assuming a normal or average growth stage. In Finney County, Kansas, the historic data indicated approximately as well as the BMTS, and both were relatively close to the ground-observed information.

GROUND TRUTH PERIODIC OBSERVATION FORM

TEST SITE # 40 (WHITMAN (2), WASHINGTON)
OBSERVATION 09

	MONTH	DAY	
LANDSAT PASS DATE	8	24	77
OBSERVATION DATE	8	24	77

RAINFALL SINCE LAST OBSERVATION _____ IN.

LAND USE CODES
100-SPRING WHEAT
200-BARLEY
300-OATS
400-WINTER WHEAT
500-GRASSES/PASTURE
600-OTHER CROPS
601-RAPESEED
602-RYE
604-FLAX
607-CORN
617-SOYBEANS
618-COTTON
700-SUMMER FALLOW
900-UNKNOWN CROPS

GROWTH STAGES
01-NOT PLANTED
02-PLANTED NO EMERGENCE
03-EMERGENCE
04-TILLERING, PREBOOT, PREBUD
05-BOOTED OR BUDDED
06-BEGINNING TO HEAD OR FLOWER
07-FULLY HEADED OR FLOWERED
08-BEGINNING TO RIPEN
09-RIPE MATURE
10-HARVESTED
11-DOES NOT APPLY

GROUND COVER (%)
1- 0- 19
2-20- 39
3-40- 59
4-60- 79
5-80-100

SURFACE MOISTURE CONDITIONS
1-DRY
2-DAMP
3-WET
4-STANDING WATER

WEED GROWTH
1-NEGLIGIBLE
2-SLIGHT
3-MODERATE
4-HEAVY

FIELD OPERATIONS
01-BARE GROUND
02-BARE DISKED/CULTIVATED
03-BARE PLOWED
04-BARESEEDED
05-STANDING STUBBLE
06-STUBBLE DISKED/CULTIVATED
07-STUBBLE PLOWED
08-STUBBLE SEEDED
09-BURNED
10-GRAZED
11-WINDROWED OR SWATHED
12-MOWED OR COMBINED
13-STACKED OR BALED
14-OTHER

GROWTH/YIELD DETRACTANTS
01-SALINITY
02-INSECTS
03-DISEASE
04-DROUGHT
05-MOISTURE
06-WIND
07-HAIL
08-FROST
09-BIRDS
10-POT HOLES
11-UNEVEN STAND
12-WEEDS
13-WINTERKILL
14-LODGING
15-OTHER

STAND QUALITY
1-POOR
2-BELOW AVERAGE
3-AVERAGE
4-ABOVE AVERAGE
5-EXCELLENT
6-DOES NOT APPLY

FIELD NO.	ACREAGE	LAND USE CODE	GROWTH STAGE (CIRCLE ONE)	GROUND COVER (CIRCLE ONE)	PLANT HEIGHT (INCHES)	SURFACE MOISTURE (CIRCLE ONE)	WEED GROWTH (CIRCLE ONE)	FIELD OPERATIONS (CIRCLE ONE)	GROWTH/YIELD DETRACTANTS (CIRCLE ONE)	STAND QUALITY RATING (CIRCLE ONE)	COMMENTS
43	233.7	414	10	5	40	1	1	05, 07		6	Y
20	511.6	700	11	1		1	1	02		6	N
31	455.0	414	10	5	40	1	1	05		6	N
30	306.6	700	11	1		1	1	02		6	N
29	401.0	414	10	5	40	1	1	05		6	N
27	89.1	700	11	1		1	1	02		6	N
25	377.1	414	10	5	40	1	1	05		6	N
8	304.8	414	10	5	40	1	1	05		6	N
49	265.0	700	11	1		1	1	02		6	N
50	160.0	614	11	1		1	1	02		6	Y
2	126.0	400	10	5	40	1	1	05		6	N
6	108.9	400	10	5	40	1	1	05		6	N
5	61.5	700	11	1		1	1	02		6	N

Figure 5-4.— ASCS Ground Truth Periodic Observation form.

ORIGINAL PAGE IS OF POOR QUALITY

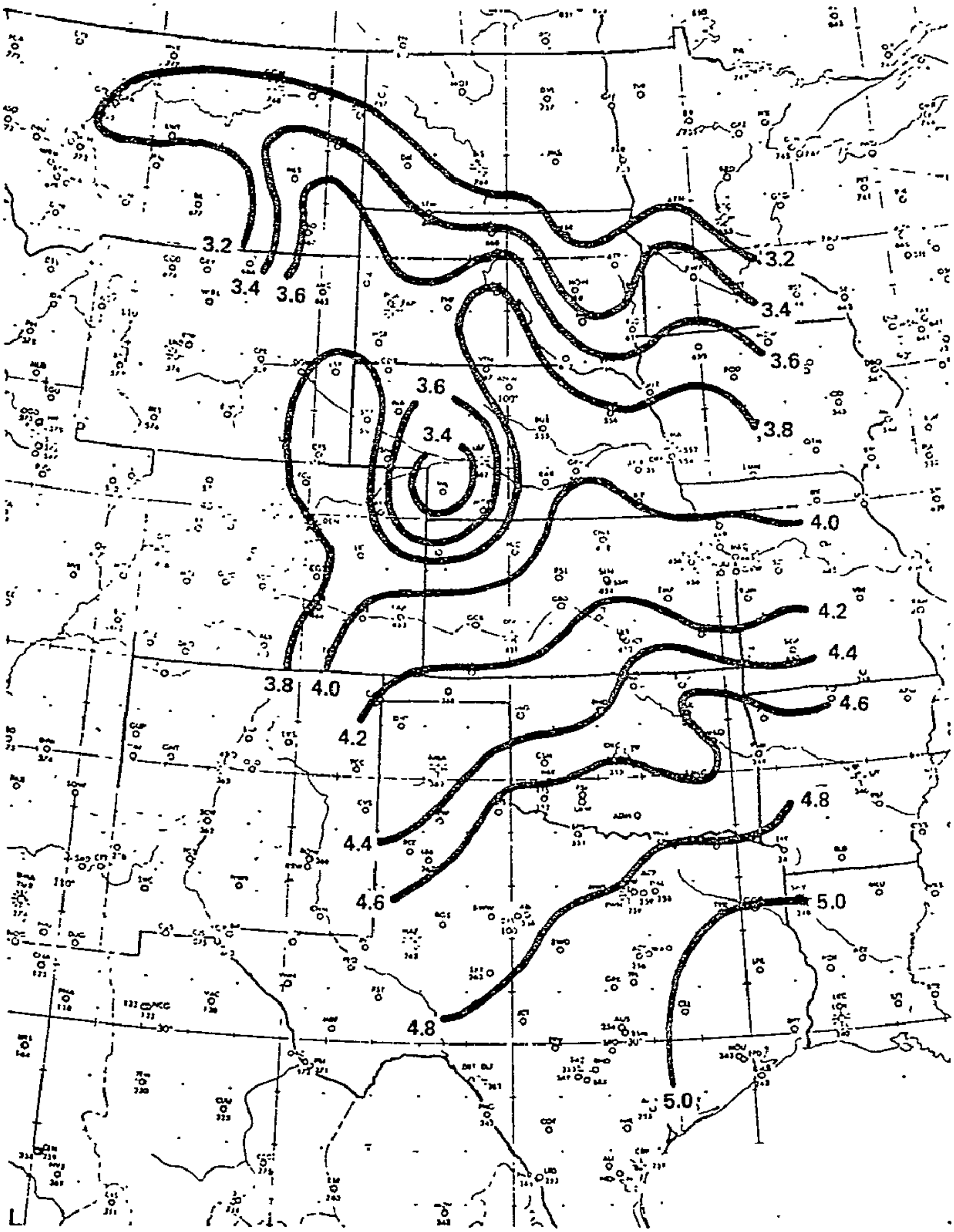

Figure 5-5.- Winter wheat BMTS isolines as predicted by the LACIE ACC meteorological data through May 1, 1977.

CRD 11, TEXAS, WINTER WHEAT, 1976-77

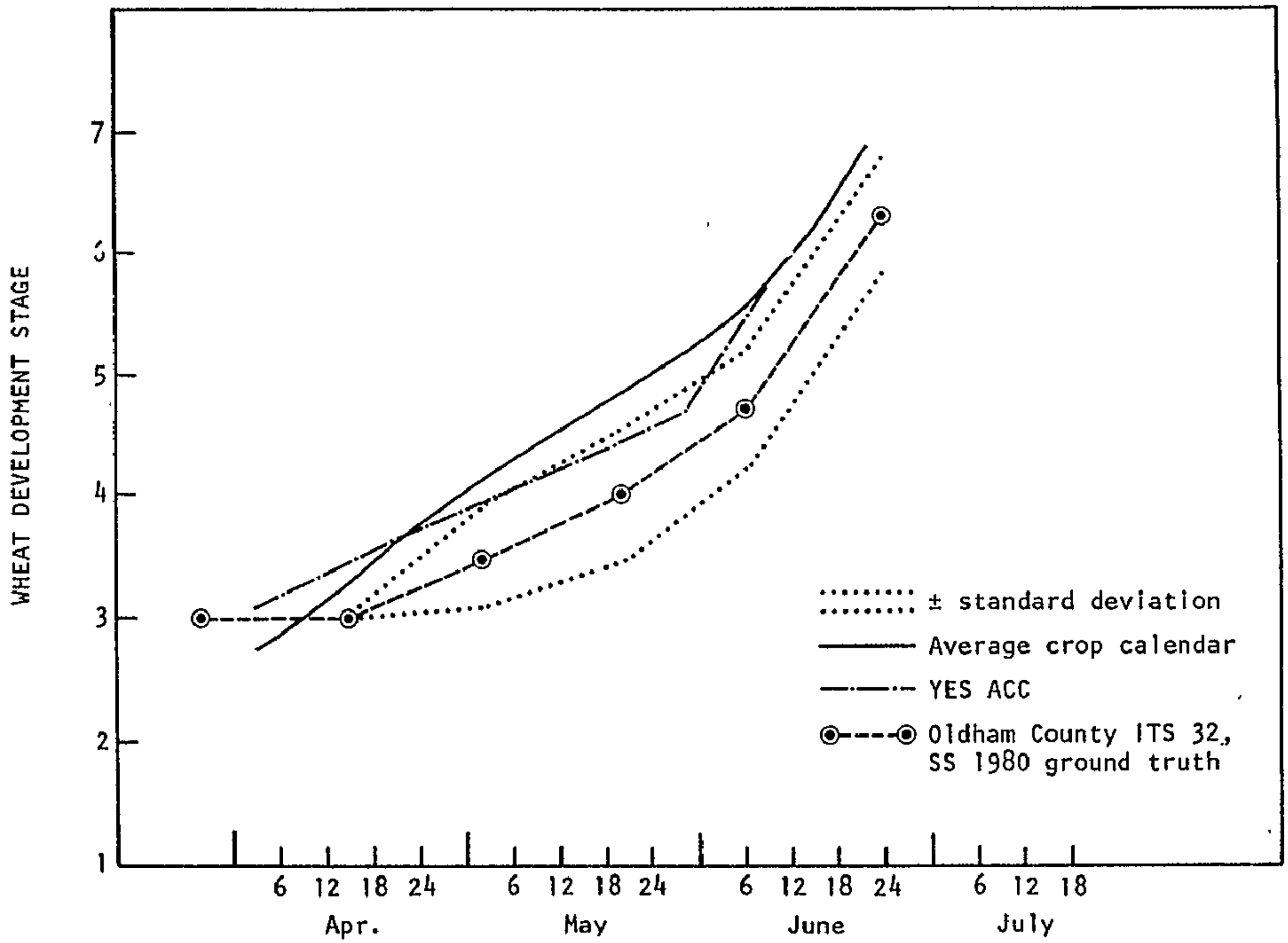

Figure 5-6.— Comparison of observed and predicted crop calendar stages for Oldham County, Texas.

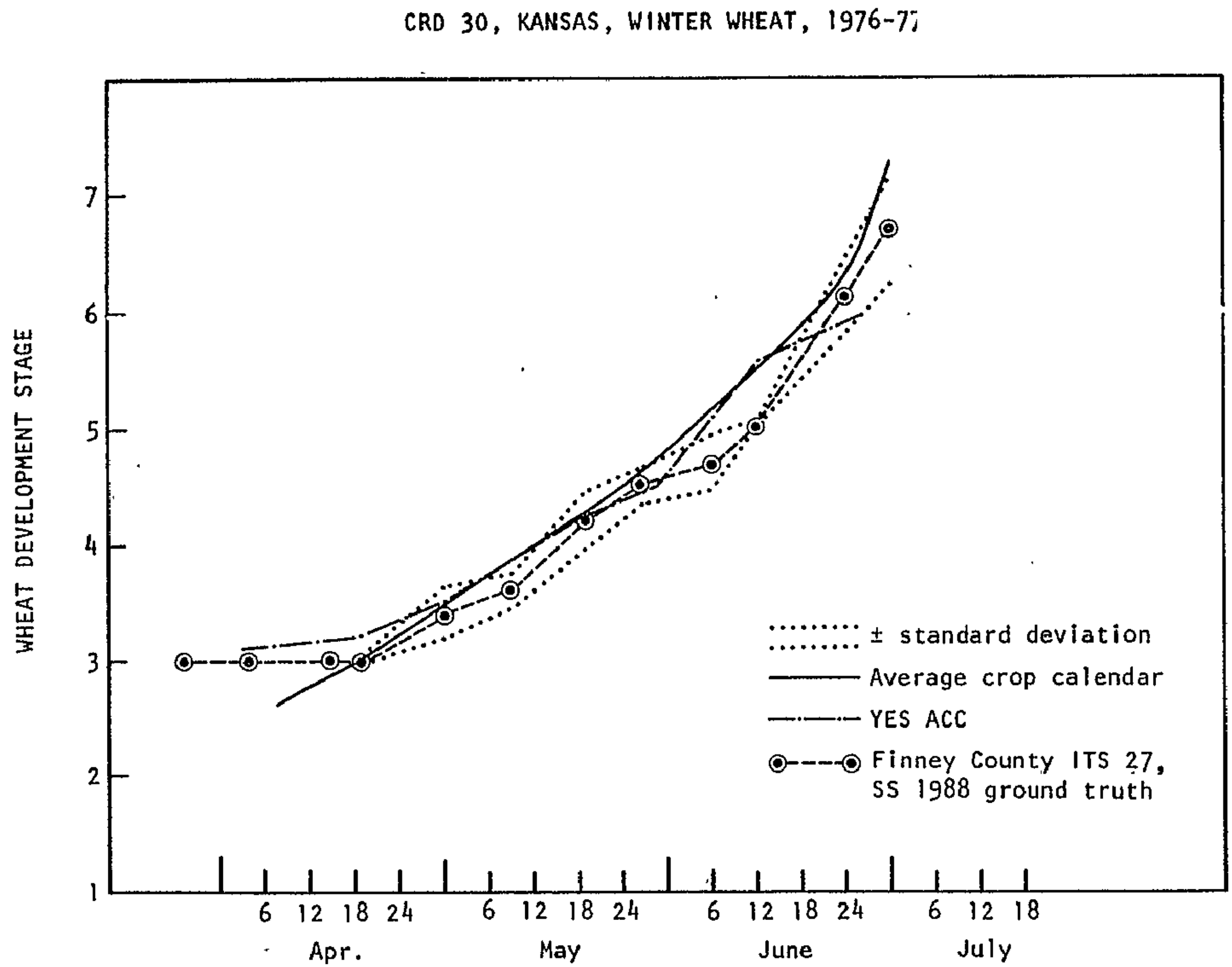

Figure 5-7.— Comparison of observed and predicted crop calendar stages for Finney County, Kansas.

ORIGINAL PAGE IS OF POOR QUALITY

Tables 5-3, 5-4, and 5-5 display the differences in days at which each of the BMTS stages was estimated by ground observations and by the LACIE ACC. At heading, the standard deviation of the ground observations is about 6 to 9 days. A difference between the ground-observed and ACC estimates larger than $\pm 1\sigma$ occurred in only three of the ITS's in the U.S. While statistical analyses of these data have not been concluded at this writing, it would appear that the computed differences between the ground-observed and ACC-estimated BMTS stages are not significant in terms of the experimental error. However, some trends were noted. In the winter wheat region, the ACC was consistently ahead of the ground observations at BMTS stages 5.0 and 6.0 (soft dough and ripening) and at jointing.

While these results do not conclusively demonstrate crop calendar inadequacies, several issues must be addressed before the ACC technology can be considered adequate. For CAMS, the analyst must know, early in the season, the expected spectral appearance of the wheat canopy. This signature, however, is related not only to the wheat growth stage but also to other factors; e.g., whether the field is irrigated and if it was fallowed the previous year, and the soil color. Thus, a signature model incorporating the ACC parameter as input would be a more desirable product from the analyst's point of view. Another major issue to be addressed is understanding just how crop calendar errors affect labeling accuracy. As mentioned at the beginning of this section, these effects are only qualitatively understood at present.

Whatever the ACC model requirements, the model can be improved for winter wheat by developing an additional model to predict the actual planting date. Currently, the LACIE ACC is "started" (i.e., the clock is set to 1.0 and meteorological data are fed to the model) on a date determined to be the historical average planting date for the Crop Reporting District (CRD) in which the segment is situated. Since this average planting date can vary considerably from one year to the next, a sizable error can be introduced into growth stage estimation before dormancy for winter wheat. In tests where the ACC has been "started" based on the ground-observed planting date, the ACC BMTS estimates have been more accurate prior to dormancy.

TABLE 5-3.— COMPARISON OF LACIE ACC WITH OBSERVED STAGES IN THE WINTER WHEAT ITS'S

[Monitoring ACC data (in days) between ITS and ACC development stages]

ITS, county/state	Jointing		Heading		Soft dough	Ripening
	3.0	3.5	4.0	4.5	5.0	6.0
Randall/Tex.	3	7	5	4	8	8
Deaf Smith/Tex.	(a)	(a)	(a)	(a)	(a)	(a)
Oldham/Tex.	-4	17	17	9	9	8
Finney/Kans.	4	5	-3	3	8	-5
Rice/Kans.	-12	0	-5	-14	0	7
Ellis/Kans.	-11	-3	-8	-15	1	-11
Saline/Kans.	4	0	-3	-3	6	11
Morton/Kans.	2	0	1	0	5	8
Shelby/Ind.	10	-1	-3	-1	-4	2
Madison/Ind.	10	6	1	0	8	5
Boone/Ind.	10	9	2	0	2	5
Oneida/Idaho	-11	-7	-7	-7	-5	(a)
Franklin/Idaho	(a)	(a)	(a)	1	4	(a)
Bannock/Idaho	15	3	0	-1	8	(a)
Whitman (1)/Wash.	(a)	(a)	(a)	(a)	(a)	(a)
Whitman (2)/Wash.	-5	10	-3	-9	2	7
Whitman (3)/Wash.	(a)	(a)	(a)	(a)	(a)	(a)
Hill/Mont.	3	-8	-9	-10	5	(a)
Liberty/Mont.	(b)	(b)	(b)	(b)	(b)	(b)
Hand (1)/S. Dak.	17	5	-5	0	(a)	(a)
Hand (2)/S. Dak.	17	(a)	(a)	(a)	(a)	(a)
Toole/Mont.	-4	-8	-6	-9	-8	(a)

[a]No data.

[b]No winter wheat.

TABLE 5-4.— COMPARISON OF LACIE ACC WITH OBSERVED STAGES IN THE SPRING WHEAT ITS'S

ITS, county/state	Jointing		Heading		Soft dough	Ripening
	3.0	3.5	4.0	4.5	5.0	6.0
Hand (1)/S. Dak.	-10	-5	-2	-8	(a)	(a)
Hand (2)/S. Dak.	(a)	(a)	(a)	(a)	(a)	(a)
Burke/N. Dak.	(a)	(a)	(a)	(a)	(a)	(a)
Divide/N. Dak.	(a)	(a)	(a)	(a)	(a)	(a)
Williams/N. Dak.	(a)	5	2	4	12	(a)
Hill/Mont.	10	12	6	6	15	(a)
Liberty/Mont.	-19	(a)	(a)	(a)	(a)	(a)
Toole/Mont.	2	(a)	-1	6	(a)	(a)
West Polk/Mont.	-7	-5	-2	6	(a)	(a)

[a] No data.

TABLE 5-5.— COMPARISON OF LACIE ACC WITH OBSERVED STAGES IN THE CANADIAN ITS'S

ITS, town/province	Jointing		Heading		Soft dough	Ripening
	3.0	3.5	4.0	4.5	5.0	6.0
Ft. Sask./Alta.	-1	0	-7	(a)	(a)	(a)
Olds/Alta.	10	7	(a)	(a)	(a)	(a)
Lethbridge/Alta.	12	13	10	(a)	(a)	(a)
Melfort/Sask.	9	9	7	(a)	(a)	(a)
Delisle/Sask.	11	5	0	(a)	(a)	(a)
Swift Current/Sask.	9	5	-4	(a)	(a)	(a)
Torquay/Sask.	7	3	-2	(a)	(a)	(a)
Stony Mt./Man.	6	3	1	2	(a)	(a)
Starbuck/Man.	4	0	-3	-3	(a)	(a)
Altona/Man.	3	-1	-8	-9	(a)	(a)
Dawson Creek/B.C.	-5	(a)	(a)	(a)	(a)	(a)

[a]No data.

6. ACCURACY ASSESSMENT SPECIAL STUDIES

This section presents the results of special studies which were done by various AA groups during the LACIE Phase III. These special studies include (1) dot labeling errors for ITS's, (2) effect of analyst, acquisition history and bias correction, (3) investigations of the winter wheat area overestimation problem in South Dakota, (4) comparison of ratioed and direct wheat aggregations for North Dakota, and (5) effect of the objective thresholding procedure.

6.1 ITS STUDY OF DOT LABELING ERRORS

By mid-September of 1977 there were 108 acquisitions from 16 ITS's in the USGP states. Reliable classifications and ground truth data were available to permit the tabulation and grouping of labeling errors for type-2 dots (bias correction dots) from the latest classification of 13 of these ITS's.

In table 6-1 the errors were divided into errors of omission, errors of commission, and errors associated with border/edge pixels. The errors of omission and commission were subdivided into two categories according to whether the pixel did or did not follow the normal wheat development sequence.

The errors of omission where the pixel did follow the normal wheat development sequence were then further subdivided in accordance with wheat development in the ITS, i.e., whether the development of most of the wheat fields in the ITS was in accordance with, behind, or ahead of the development expected from the Robertson biostage as determined from the ACC.

The errors of commission where the pixel did follow the normal wheat development sequence were subdivided into nonwheat and volunteer wheat pixels. The identification of volunteer wheat as wheat is considered to be an error because it is not usually harvested.

A total of 978 type-2 dots were labeled by analysts for the 13 ITS's and 94 (9.6 percent) were incorrectly labeled. Errors of omission (64 of 326, or 19.6 percent) exceeded errors of commission (30 of 652, or 4.6 percent).

TABLE 6-1.— ERRORS OF OMISSION AND COMMISSION

Reason	Errors of omission								Errors of commission						Border edge pixel		Errors		
	Did the dot follow the normal wheat development sequence?																		
	Yes						No		Yes				No						
	Normal[a] development		Developed late[a]		Developed early[a]				But not wheat		Volunteer wheat							Total	Total dots
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	%
Necessary acquisitions not available									6	6.4							6	6.4	0.6
Poor stand of wheat			3	3.2					1	1.1							4	4.3	0.4
Late planting, emergence, or development	10	10.6	24	25.5									2	2.1			36	38.3	3.7
Difficulties caused by narrow strip fields					2	2.1											2	2.1	0.2
Clerical error[b]	13	13.8											4	4.3			17	18.1	1.7
Confused with other crops	9	9.6							2	2.1	4	4.3	1	1.1			16	17.0	1.6
Border edge pixel													3	3.2			3	3.2	0.3
Unknown			1	1.1	2	2.1							6	6.4	1	1.1	10	10.6	1.0
Total	32	34.0	28	29.8	4	4.3			9	9.6	4	4.3	16	17.0	1	1.1	94	100.0	9.6

[a]Development refers to the development of most of the wheat fields in the ITS relative to the Robertson biostage of the adjustable crop calendar.

[b]Clerical error:
1. Wrong acquisition requested for classification. Analyst simply wrote the wrong number inadvertently.
2. Pixel misidentified by mistake. Same signature on other pixels was consistently identified as nonwheat.

The largest single cause of labeling errors was late planting, emergence, or development of small grains. This category of error accounted for 36 (38.3 percent) of the total errors. Other major causes of errors were analyst errors (17 errors, or 18.1 percent), and confusion with other crops (16 errors, or 17.0 percent).

Most errors were categorized with approximately equal weight given to each of the following reasons:

a. Omission errors because the signatures were unusual in an otherwise normal temporal development pattern.

b. Omission errors because the wheat fields were well behind the normal development pattern according to the ACC.

c. Commission errors because the wheat signature was confused with other signatures for a variety of reasons, mainly confusion with native vegetation or hay.

Similarly, the three greatest causes of error were: late emergence, inconsistencies in labeling, and confusion with other crops.

Table 6-2 shows the types and causes of labeling errors for each of the 13 ITS's. Only 4 of the 13 ITS's had more than 10 labeling errors and in each one many of the errors had the same cause. In the Hand County #2 segment, all eight errors of omission were normally developed wheat which was confused with native pasture. In Morton County all 13 omission errors were late emerged or developed wheat which was confused with hay. Ten of the errors of omission in Rice County were normally developed wheat with an unusual color signature (such as purple, magenta, or brown). Six of the errors of commission in Randall County were nonwheat pixels which appeared to be wheat on available acquisitions. (Additional acquisitions were needed.)

6.2 EFFECTS OF ANALYST INTERPRETER (AI), ACQUISITION HISTORY, AND BIAS CORRECTION ON PROPORTION ESTIMATION ERROR

The Image 100 processor and data from eight U.S. blind sites were used in an experiment wherein each site was analyzed by three AI's to give a procedure 1

TABLE 6-2.— LABELING OF ERRORS AT SEGMENT LEVEL

Segment	Robertson biostage	LACIE biowindow	True wheat, %	Wheat dots			Other dots			No. incorrect labels	Incorrect labels, %	Necessary acquisitions not available	Poor stand of wheat	Late planting, emergence, or development	Difficulties caused by narrow strip fields	Clerical errors[b]	Confused with other crops	Border/edge pixels	Unknown
					Incorrect labels			Incorrect labels											
				No.	No.	%	No.	No.	%										
1687 Hand #1, SD	2	1	15.9	26	2	7.7	72	8	11.1	10	10.2					4	4		2
1986 Hand #2, SD	[a]4-3	[a]3-2	1.3	9	8	88.9	89	3	3.4	11	11.2						10		1
1961 Morton, KS	3	2	42.8	35	13	37.1	64	0	0	13	13.1			13					
1962 Saline, KS	5	3	62.9	36	1	2.8	24	4	16.7	5	8.3					1		3	1
1963 Rice, KS	6	4	35.0	23	11	47.8	37	0	0	11	18.3			10			1		
1964 Elvis, KS	5	3	48.3	23	10	43.5	35	0	0	10	17.2			3	2	3			2
1988 Finney, KS	3	2	42.9	21	4	19.0	31	1	3.2	5	9.6		3			2			
1970 Liberty, MT	1	1	3.3	8	0	0	91	2	2.2	2	2.0			2					
1971 Hill, MT	4	3	32.1	33	6	18.2	57	1	1.8	7	7.8			3		3	1		
1978 Randall, TX	5	3	43.5	23	3	13.0	30	9	30.0	12	22.6	6		2		1			3
1979 Deaf Smith, TX	2	1	27.7	20	2	10.0	78	1	1.3	3	3.1			2					1
1980 Oldham, TX	6	4	30.0	31	4	12.9	22	1	4.5	5	9.4		1	1		3			
1987 West Polk, MN	4	3	60.1	38	0	0	22	0	0	0	0								
Total				326	64	19.6	652	30	4.6	94	9.6	6	4	36	2	17	16	3	10

[a]Mixed wheat.

[b]Clerical error:
1. Wrong acquisition requested for classification. Analyst simply wrote the wrong number inadvertently.
2. Pixel misidentified by mistake. Same signature on other pixels was consistently identified as nonwheat.

"raw" and a "bias-corrected" estimate of the proportion of small grains in each segment. The segments were of two types; namely, those having acquisitions in all four biophases and those having only early season acquisitions. The segments were chosen at random from the blind sites for which detailed ground truth was available.

The objectives of the experiment were: (1) to evaluate the performance of procedure 1 in terms of absolute proportion estimation error and its repeatibility over AI's, (2) to make comparisons between "bias-corrected" and "raw" procedure 1 estimates, and (3) to determine if the performance was better when acquisitions from all biostages were used than when only the early season acquisition was used.

The third objective could not be achieved properly because of the small number of segments used (four of each type). It was later estimated that to make effective comparisons of this type in a fully nested design, one would need about 10 times as many segments. The efficiency of the test could be improved if the same segments were analyzed first using only early-season acquisitions and then using all acquisitions; however, there would be potential biasing problems in such replication if the same AI analyzed the segment under both the early-season and full-season conditions. If different AI's performed the analysis, the resulting large variability would destroy the power of the test just as the large segment variability destroyed it in the experiment reported here.

Table 6-3 shows the absolute proportion estimation error $|\hat{X} - X|$ where X is the ground truth small grains proportion and $\hat{X}$ is the analyst's estimate of X for the various treatment combinations. Averages are blocked off from the basic data; for example, the average absolute error for AI "B" on early-season segments was 11.6 for the raw estimate and 11.8 for the bias-corrected estimate. The average absolute error on all segments was 7.9 for raw estimates and 11.1 for bias-corrected estimates. The average absolute error for all three AI's was 12.8 for raw early-season estimates, 6.3 for raw full-season estimates, and 9.5 for all eight segments with the raw estimate. The grand mean was 10.0.

TABLE 6-3.— IMAGE 100 — PROCEDURE 1 DATA

[$|\hat{X} - X|$ (small grains)]

Acquisition history	Raw				Bias correction				Overall average
	Analyst			Average	Analyst			Average	
	A	B	C		A	B	C		
Early season only	16.5	10.8	2.0		18.9	8.7	16.8		
	11.4	18.5	21.3		5.6	18.3	19.7		
	9.7	14.6	30.3		8.0	11.9	19.5		
	8.4	2.5	7.0		1.6	8.2	1.5		
Average	11.5	11.6	15.2	12.8	8.5	11.8	14.4	11.6	12.2
Full season	0.8	1.4	0.9		1.4	1.4	2.0		
	5.2	10.6	31.7		9.7	32.9	32.6		
	1.3	0.3	15.1		7.2	5.0	14.0		
	1.7	4.7	2.4		2.7	2.5	2.5		
Average	2.2	4.3	12.5	6.3	5.3	10.5	12.8	9.5	7.9
Overall average	6.9	7.9	13.8	9.5	6.9	11.1	13.6	10.5	10.0

The most obvious features of table 6-3 are the large variability between AI's and between segments. If this variation is taken to be typical, then future experiments should be designed so that segments and AI's are "crossed" with treatments as much as possible.

Analysis of variance was used to test for the effects of AI's, time (i.e., early-season versus all acquisitions), method (raw versus bias correction), and their interactions. The results are shown in table 6-4. They lead to the following conclusions:

a. The large disparity between data from various AI's was not consistent over segments; i.e., an AI would do better on one segment than on another one.

b. There was no significant difference between methods; i.e. the use of bias correction just traded one random error for another one of comparable magnitude.

c. Any test involving "times" was not significant. As stated earlier, these tests had extremely low power because of insufficient numbers of segments.

6.3 INVESTIGATIONS OF THE WINTER WHEAT AREA OVERESTIMATION PROBLEM IN SOUTH DAKOTA

The LACIE winter wheat area estimate for July was below the USDA/SRS estimate at the USSGP level and above the USDA/SRS estimates for the mixed wheat states and the USGP. This was primarily because of the large overestimate in South Dakota as shown in table 6-5. A relative difference of 85.3 percent was reported in South Dakota, indicating a large overestimate. In fact, the LACIE area estimate for South Dakota was approximately seven times greater than the corresponding USDA/SRS estimate. An investigation of this problem was conducted by various elements of LACIE.

Several factors contributed to the overestimation of winter wheat area in South Dakota. One factor was the CAMS overestimation of winter wheat area in marginal wheat areas. A second factor was the sensitivity of the aggregation model to overestimation in such areas. Also, it is possible that part of the

TABLE 6-4.— ANALYSIS OF VARIANCE

Source	Degree of freedom	SS[a]	MS[b]	F[c]
T (TIME)	1	215.48	215.48	0.70[1]
M (METH)	1	11.50	11.51	0.41[2]
TM (TM)	1	56.55	56.55	2.01[3]
A (AI)	2	379.05	189.52	2.53
AT	2	34.51	17.25	0.23
AM	2	29.91	14.71	0.68
TAM	2	16.50	8.25	0.38
S/T	6	1840.09	306.68	4.09*
MS/T	6	168.94	28.16	1.30
AS/T	12	899.63	74.97	3.46X*
MAS/T	12	259.65	21.64	

[a]SS - Sum of Squares.

[b]MS - Mean square.

[c]F - F-value.

LEGEND:

1 - Assume $\sigma^2 AT = 0$
2 - Assume $\sigma^2 AM = 0$
3 - Assume $\sigma^2 TAM = 0$
*P < .025
X - Conservative test (inflated denominator)

TABLE 6-5.— COMPARISON OF SOUTH DAKOTA AND MONTANA WINTER WHEAT ESTIMATES USING REDESIGNATED SEGMENTS WITH USDA/SRS AND LACIE ESTIMATES

	No. segments (acq/alloc)	Area (ac × 10^3)	S.E.[a] (ac × 10^3)	CV (%)	RD[b] (%)	$\sqrt{MSE}$[c]
SOUTH DAKOTA						
USDA/SRS		680				
LACIE	39/56	4629	583	12.6	85.3	3991.8
Redes. LACIE	20/26	1323	713	53.9	48.6	960.1
MONTANA						
USDA/SRS		2800				
LACIE	58/80	3097	380	12.3	9.6	482.3
Redes. LACIE	46/58	2902	386	13.3	3.5	399.2

[a]S.E. = Standard error.

[b]RD = Relative difference.

[c]MSE = Variance + $(\text{bias})^2$, where bias is estimated by (LACIE - SRS).

overestimation might have been due to the use of CRD-level ratios of winter wheat to winter small grains to determine the winter wheat proportion. To check this, aggregation using the 1975 USDA/SRS county-level ratios was performed and a large overestimation still resulted. Use of the CRD ratios resulted in an acreage estimate about seven times larger than the USDA/SRS estimate, compared to an estimate six times larger using the county ratios. This seemed to indicate that overestimation of winter small grains proportions by CAMS was the greater problem.

The aggregation logic in the CAS system is especially sensitive to proportion estimation errors in marginal areas. As an example, consider the collection of Group II counties in CRD 90 in South Dakota. The epoch year winter wheat area data and number of segments allocated to each of these counties were as follows:

County	No. segments acquired/allocated	1974 Census winter wheat area (acre × 10^3)
C_1	1/1	21
C_2	0/0	371
C_3	0/1	443
C_4	1/1	79
C_5	0/0	0
C_6	1/1	271
C_7	0/0	939
C_8	1/1	375
TOTAL	4/5	2 499

With the four acquired segments, the LACIE estimate of winter wheat area for this Group II collection is given by:

$$\frac{1}{4}\left[\frac{2499}{21}\ \hat{C}_1 + \frac{2499}{79}\ \hat{C}_4 + \frac{2499}{271}\ \hat{C}_6 + \frac{2499}{375}\ \hat{C}_8\right]$$

$$= 29.75\ \hat{C}_1 + 7.91\ \hat{C}_4 + 2.31\ \hat{C}_6 = 1.67\ \hat{C}_8$$

where $\hat{C}_i$ is the estimate of winter wheat area in the *ith* county (i=1,...,8) as determined from the winter wheat proportion estimate of the segment in the *ith* county. Note that even a small overestimation of the $\hat{C}_i$, particularly $\hat{C}_1$ and $\hat{C}_4$, could lead to a gross overestimate for the collection. In the July CMR, the estimate for the segment in county C_1 was 38 times larger than the historical county winter wheat proportion with obvious results — extreme overestimation for the collection.

Observing the 1974 area, or production, of winter wheat for these counties, it is obvious that none of them should have received a segment to estimate winter wheat area. They did so because a new allocation was performed for Phase III based on total small grains. This resulted in the allocation of sample segments to areas containing small grains but little or no winter wheat in South Dakota. Also, the new allocation required both a winter wheat and a spring wheat estimate for segments in areas designated as mixed; i.e., containing both winter wheat and spring wheat. Several counties in South Dakota, a mixed region, contained a significant portion of spring wheat but not winter wheat. This resulted in CAMS having to make winter wheat estimates where there was likely to be no winter wheat.

To avoid the overestimation of area in sparse wheat (winter or spring) regions a redesignation of the segments as winter (W), spring (S), or mixed (M) was made for the remainder of Phase III. This had no great impact on the CAS aggregation procedures. No spring wheat estimates had been made for those segments designated as pure winter segments and vice versa. The estimates for these counties were made using the Group III estimator, or the Group II estimator if the county was a Group II collection containing a county for which an estimate was made. Of course, both spring wheat and winter wheat estimates were made for those segments designated as mixed.

Currently, action is being taken to redesignate Group III counties in the USGP using the epoch year wheat production rather than the small grains production used in the Phase III allocation. This will indicate counties that received segments but should not have and counties that did not receive segments but should have.

Aggregations have been performed using the above-mentioned designations, based on segment data from the July 11 CMR, to obtain state winter wheat area estimates for South Dakota and Montana. When bias is a problem, as is the case here, the preferred statistic for comparison of two estimators is the mean squared error (MSE) or its square root, where the MSE is given by the variance of the estimator plus the square of the bias of the estimator. Assuming the USDA/SRS estimate to be the true value, the bias is estimated by the difference between the LACIE and the USDA/SRS estimates. The results are presented in table 6-5. It is apparent that the redesignation of segments improved the LACIE estimate considerably, particularly in South Dakota where there is very little winter wheat, according to USDA/SRS data. Although an increase in the standard error is noted using redesignations, the large reduction in bias resulted in a 76-percent decrease in the square root of the mean squared error for South Dakota. In Montana, there are many more acres of winter wheat than in South Dakota and the use of redesignated segments resulted in only a slight improvement, a 17-percent reduction in the square root of the mean squared error.

6.4 COMPARISON OF RATIOED AND DIRECT WHEAT AGGREGATIONS

For the CMR's of August, September, and October, two types of proportion estimates were made by CAMS for the segments in North Dakota. First, as usual, CAMS estimated the spring small grains percentage for each segment. These estimates were passed to CAS and ratioed down to spring wheat proportions before aggregation. In addition, CAMS estimated spring wheat proportions directly for these same segments. These estimates were also aggregated by CAS.

The results of the two aggregations are shown in table 6-6 along with the corresponding USDA/SRS estimates. The CV's for the direct wheat estimates are slightly smaller than those for the ratioed wheat estimates in all three months. However, the relative differences for August and September are larger (in absolute value) for the ratioed wheat estimates. In October the relative difference for the direct wheat estimate was larger. In August both estimates were significantly different from the USDA/SRS estimate. In September the

TABLE 6-6.– COMPARISON OF RATIOED AND DIRECT SPRING WHEAT (AGGREGATION) AREA ESTIMATES FOR NORTH DAKOTA

Month of estimate	USDA/SRS area estimate (ac × 10^3)	LACIE				RD(%)		Value of test statistic	
		Estimate (ac × 10^3)		CV(%)					
		Ratioed wheat	Direct wheat	Ratioed wheat	Direct wheat	Ratioed wheat	Direct wheat	Ratioed wheat	Direct wheat
August	9530	6761	7525	8.6	9.6	-41.0	-26.6	-4.8*	-2.8*
September	9530	8678	9828	4.6	5.2	-9.8	3.0	-2.1*	0.6^N
October	9530	9173	10604	4.4	4.8	-3.9	10.1	-0.9^N	2.1*

*μ_D is significantly different from zero at the 10-percent level.

$^N\mu_D$ is not significantly different from zero at the 10-percent level.

direct wheat estimate was not significantly different from the USDA/SRS estimate and in October the ratioed wheat estimate was not significantly different from the USDA/SRS estimate.

A blind site study was performed using the ratioed and direct estimates for North Dakota from the October 11 CMR. Figure 6-1 shows plots of the proportion error $\hat{X} - X$, where $\hat{X}$ is the LACIE proportion estimate and X is the ground-truth proportion obtained using the dot-count ground-truth proportions. Table 6-7 shows that the results of the statistical calculations for the same data were closer to the dot-count ground-truth proportion estimates than were the ratioed estimates. Like the aggregation study, the blind site study indicated a higher degree of variability in the direct wheat estimates, as evidenced by the plots shown in figure 6-1.

In both studies, the October results reveal underestimation in the ratioed wheat estimate and overestimation in the direct wheat estimate.

6.5 EFFECT OF THE OBJECTIVE THRESHOLDING PROCEDURE

Investigations of the early-season estimates in Phase II disclosed the presence of an early-season bias or underestimate of harvestable wheat area. This was caused by wheatfields with insufficient canopy development, which were not detectable by Landsat. LACIE began Landsat data processing when the normal crop calendar reached biostage 2.0 (emergence) on the Robertson growth scale. As the season progressed, ground cover within the fields increased, and the LACIE area estimates converged toward the area harvested. Because of cloud cover and data drop, some segments were not acquired after complete emergence. However, wheat area estimates based on the early acquisitions for these segments were utilized to make area estimates throughout the season in the conventional aggregations. This contributed to the tendency to underestimate wheat area at harvest.

In Phase III, an objective thresholding procedure was developed to eliminate segments with incomplete emergence from consideration in the overall area

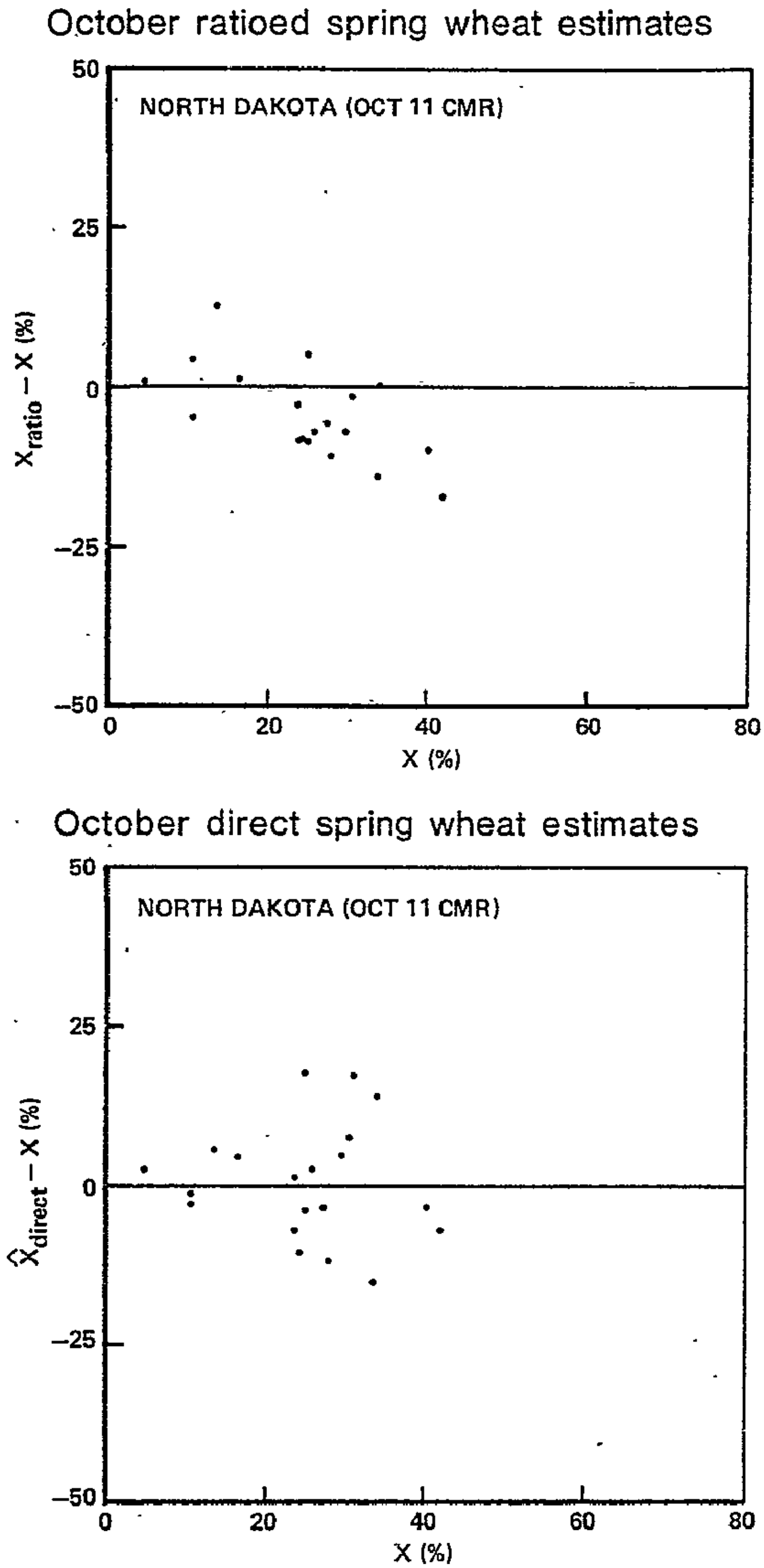

Figure 6-1.— Plots of proportion estimation errors versus dot-count ground-truth proportion estimates for the blind sites in North Dakota.

TABLE 6-7.— COMPARISON OF RATIOED AND DIRECT SPRING WHEAT BLIND SITE PROPORTION ESTIMATES (EXPRESSED IN PERCENTAGES) FOR NORTH DAKOTA

Estimate	n/M	$\bar{\hat{X}}$	$\bar{X}$	$\bar{D}$	$S_{\bar{D}}$	90% confidence interval for μ_D
Ratioed	20/103	21.0	25.1	-4.1	1.5	(-6.7,-1.5)*
Direct	20/103	25.6	25.1	0.5	1.8	(-2.6,3.6)^N

*μ_D is significantly different from zero at the 10-percent level.

$^N\mu_D$ is not significantly different from zero at the 10-percent level.

estimates. The thresholding procedure can be applied only at mid-season after several opportunities to acquire and estimate wheat percentages have occurred. This procedure was tested in Phase III and was demonstrated to decrease the magnitude of the underestimate throughout the season. Therefore, in addition to the conventional (nonthresholded) LACIE estimates, CAS also provided the thresholded estimates in the June and July CMR's.

The Application Evaluation System of LACIE established Robertson biostage 2.55, as determined from the ACC for crop year 1977, as the wheat detection threshold of the LACIE system for all winter wheat states except Texas, which utilizes biostage 2.6. These thresholds were applied to the Landsat data, and no segments acquired before the detection threshold were included in the thresholded aggregation.

In table 6-8, the LACIE thresholded and conventional estimates of winter wheat area for the seven states and for the regional levels are compared with the USDA/SRS estimates. In June, area estimates from all regions and states except Nebraska increased after the thresholding procedure was utilized. Nebraska showed a slight decrease in the area estimate. These changes in the estimates took them closer to agreement with the USDA/SRS estimates in four of the seven USGP winter wheat states but increased the relative difference at the USGP-7 level. This increase was caused by a sampling problem in the mixed wheat states. The CV's were increased only slightly by the thresholding except in South Dakota (where the greatest increase in the estimate occurred) and at the USGP-7 level. The CV for South Dakota jumped from 34 to 60 percent and that for the USGP-7 region went from 5 to 18 percent. This increase resulted from the decrease in the number of segments used for aggregation.

As shown in the July CMR, area estimates for the five USSGP winter wheat producing states changed only slightly after thresholding. The CV's for these estimates remained constant. The small observed differences between the thresholded and conventional estimates resulted from a large number of segment acquisitions after emergence and, therefore, minimal thresholding. Recorded changes were in the form of mixed increases and decreases among the seven states.

TABLE 6-8.— COMPARISON OF THRESHOLDED WITH CONVENTIONAL AREA ESTIMATES

Region	USDA/SRS	Thresholded LACIE				Nonthresholded LACIE			
	Est ($ac \times 10^3$)	Est ($ac \times 10^3$)	n/M	RD[a] (%)	CV[e] (%)	Est ($ac \times 10^3$)	n/M	RD[a] (%)	CV[e] (%)
June 7, 1977 CMR									
Colorado	2360	3500	12/32	32.6	15.7	3065	28/32	23.0	15.8
Kansas	12000	11743	82/121	-2.2	5.5	10915	112/121	-9.9	5.8
Nebraska	3050	3603	22/67	15.3	13.3	3610	50/67	15.5	12.1
Oklahoma	6500	5307	40/46	-22.5	8.0	4875	45/46	-33.3	9.0
Texas	4400	4910	26/38	10.4	13.1	4529	34/38	2.8	11.9
[b]USSGP	28310	29063	182/304	2.6	4.3	26994	269/304	-4.9	4.2
Montana	2800	4188	3/80	33.1	28.8	3253	41/80	13.9	19.2
S. Dakota	680	13759	5/56	95.1	10.8	2601	28/56	73.9	34.0
[c]MW States	3480	17947	8/136	80.6	10.7	5854	69/136	40.6	18.5
[d]USGP-7	31790	47010	190/440	32.4	4.9	32848	338/440	3.2	4.8
July 11, 1977 CMR									
Colorado	2360	2781	25/32	15.1	15.5	2962	30/32	20.3	13.2
Kansas	12300	12524	98/121	1.8	4.8	11764	111/121	-4.6	5.0
Nebraska	3050	3746	34/67	18.6	11.6	3475	52/67	12.2	12.4
Oklahoma	6500	5628	37/46	-15.5	7.5	5264	42/46	-23.5	8.5
Texas	4600	4625	29/38	0.5	12.8	4511	34/38	-2.0	11.6
[b]USSGP	28810	29304	223/304	1.7	3.8	27976	269/304	-3.0	3.9
Montana	2800	2629	44/80	-6.5	11.9	3097	58/80	9.6	12.3
S. Dakota	680	5671	32/56	88.0	13.8	4629	39/56	85.3	12.6
[c]MW States	3480	8300	76/136	58.1	10.1	7726	97/136	55.0	9.0
[d]USGP-7	32290	37604	299/440	14.1	3.7	35701	366/440	9.6	3.6

n = number of segments used.
M = number of segments allocated.
[a]Relative difference = $\left(\frac{\text{LACIE - USDA/SRS}}{\text{LACIE}} \times 100\right)\%$.
[b]U.S. southern Great Plains region.
[c]The mixed wheat states, Montana and S. Dakota.
[d]Seven-state winter wheat region of U.S. Great Plains.
[e]Coefficient of variation.

In June and July, the winter wheat areas were largely overestimated by both procedures. The overestimation was caused by an inappropriate sampling strategy for the mixed wheat areas. The detailed results of the investigation of this problem are presented in section 6.3.

Table 6-9 presents the LACIE thresholded and conventional estimates of winter wheat yield for the USGP-7 states and for the regional levels in comparison with the USDA/SRS estimates. The thresholding procedure should have no effect upon the LACIE yield estimates at the pseudo zone level. The difference between conventional and thresholded yield estimates for the state or higher levels is due to the different weighting factors for the thresholded and conventional area estimates applied at the pseudo zone level. The results presented in table 6-9 indicate that the thresholding procedure had very little effect upon the LACIE yield estimates at the state and regional levels.

In table 6-10 the LACIE thresholded and nonthresholded estimates of winter wheat production for the USGP-7 and the regional levels are compared with the USDA/SRS estimates. In June, thresholding increased all of the production estimates at both the state and regional levels. At the USSGP level, this increase resulted in an improvement in the relative difference from -23.6 percent to -15.0 percent. In Montana and South Dakota, thresholding reduced the number of usable segments to 3 and 5, respectively. This is clearly not enough segments to make a reliable estimate, as evidenced by very large increases in the estimates for these two states, particularly South Dakota, which increased by more than a factor of 5 as a result of thresholding. The thresholded estimate at the USGP-7 level was less accurate than the nonthresholded estimate mainly due to the increase in South Dakota.

In July, estimates increased as a result of thresholding in the three regions and in all states except Colorado and Montana. The number of segments thresholded at the USGP level decreased from 148 in June to 67 in July largely as a result of a number of later acquisitions becoming available in Montana and South Dakota. Thesholding decreased the relative difference at the USSGP

TABLE 6-9.— COMPARISON OF THRESHOLDED WITH CONVENTIONAL YIELD ESTIMATES

Region	USDA/SRS	Thresholded LACIE			Nonthresholded LACIE		
	Est (bu/ac)	Est (bu/ac)	RD[a] (%)	CV[e] (%)	Est (bu/ac)	RD[a] (%)	CV[e] (%)
June 7, 1977 CMR							
Colorado	24.0	23.6	-1.7	16.9	23.6	-1.7	16.9
Kansas	33.0	28.3	-16.6	10.6	28.3	-16.6	10.6
Nebraska	35.0	30.6	-14.4	9.8	30.1	-16.3	6.6
Oklahoma	26.0	19.8	-31.3	5.1	19.8	-31.3	5.1
Texas	25.0	20.2	-23.8	5.0	20.3	-23.2	4.9
[b]USSGP	29.6	25.1	-17.9	4.0	25.1	-17.9	4.0
Montana	27.0	28.1	3.9	14.2	28.1	3.9	14.2
S. Dakota	20.0	26.0	23.1	19.2	26.0	23.1	19.2
[c]MW States	25.6	26.5	3.4	11.3	27.2	5.9	11.0
[d]USGP-7	29.2	25.6	-14.1	7.8	25.5	-14.5	3.9
July 11, 1977 CMR							
Colorado	23.0	22.5	-2.2	14.8	22.5	-2.2	14.8
Kansas	31.0	28.8	-7.6	9.7	28.8	-7.6	9.7
Nebraska	35.0	32.1	-9.0	9.7	32.2	-8.7	9.3
Oklahoma	26.0	19.9	-30.7	10.9	19.9	-30.7	10.7
Texas	25.0	20.3	-23.2	10.5	20.3	-23.2	10.1
[b]USSGP	28.7	25.6	-12.1	*	25.5	-12.5	5.5
Montana	27.0	26.5	-1.9	12.1	26.5	-1.9	12.1
S. Dakota	24.0	26.6	9.8	18.9	26.6	9.8	16.9
[c]MW States	26.4	26.6	0.8	*	26.6	0.8	*
[d]USGP-7	28.4	25.8	-10.1	5.5	25.8	-10.1	5.3

[a]Relative difference = $\left(\frac{\text{LACIE - USDA/SRS}}{\text{LACIE}} \times 100\right)\%$.
[b]U.S. southern Great Plains region.
[c]The mixed wheat states, Montana and S. Dakota.
[d]Seven-state winter wheat region of U.S. Great Plains.
[e]Coefficient of variation.
*Data not available.

TABLE 6-10.— COMPARISON OF THRESHOLDED WITH CONVENTIONAL PRODUCTION ESTIMATES

Region	USDA/SRS	Thresholded LACIE				Nonthresholded LACIE			
	Est ($bu \times 10^3$)	Est ($bu \times 10^3$)	n/M	RD[a] (%)	CV[e] (%)	Est ($bu \times 10^3$)	n/M	RD[a] (%)	CV[e] (%)
June 7, 1977 CMR									
Colorado	56640	82752	12/32	31.6	21.8	72456	28/32	21.8	21.9
Kansas	396000	331765	82/121	-19.4	11.6	308387	112/121	-28.4	11.5
Nebraska	106750	110081	22/67	3.0	18.7	108793	50/67	1.9	16.2
Oklahoma	169000	104958	40/46	-61.0	13.1	96550	45/46	-75.0	14.0
Texas	110000	99394	26/38	-10.7	15.9	91965	34/38	-19.6	14.2
[b]USSGP	838390	728950	182/304	-15.0	7.1	678151	269/304	-23.6	6.9
Montana	75600	117700	3/80	35.8	28.1	91417	41/80	17.3	23.2
S. Dakota	13600	358030	5/56	96.2	62.2	67685	28/56	79.9	38.3
[c]MW States	89200	475730	8/136	81.2	47.3	159102	69/136	43.9	21.1
[d]USGP-7	927590	1204680	190/440	23.0	19.2	837254	388/440	-10.8	7.0
July 11, 1977 CMR									
Colorado	54280	62436	25/32	13.1	21.3	66516	30/32	18.4	19.7
Kansas	381300	361294	98/121	-5.5	10.8	339348	111/121	-12.4	10.9
Nebraska	106750	120392	34/67	11.3	15.0	111903	52/67	4.6	15.7
Oklahoma	169000	112045	37/46	-50.8	13.1	104907	42/46	-61.1	13.6
Texas	115000	93817	29/38	-22.6	14.8	91691	34/38	-25.4	13.9
[b]USSGP	826330	749984	223/304	-10.2	*	714365	269/304	-15.7	*
Montana	75600	69581	44/80	-8.7	16.9	81983	58/80	7.8	17.2
S. Dakota	16320	150933	32/56	89.2	23.2	123196	39/56	86.8	22.6
[c]MW States	91920	220514	76/136	58.3	*	205179	97/136	55.2	*
[d]USGP-7	918250	970498	299/440	5.4	6.6	919544	366/440	0.1	6.4

n = Number of segments used.
M = Number of segments allocated.

[a]Relative difference = $\left(\frac{\text{LACIE - USDA/SRS}}{\text{LACIE}} \times 100\right)$%.

[b]U.S. southern Great Plains region.

[c]The mixed wheat states, Montana and S. Dakota.

[d]Seven-state winter wheat region of U.S. Great Plains.

[e]Coefficient of variation.

*Data not available.

level and increased it for the mixed wheat states and for the USGP. CV's were only slightly changed by thresholding.

7. ASSESSMENT OF LACIE ESTIMATES FOR U.S.S.R.

This section presents an evaluation of LACIE estimates of production, area, and yield for the winter, spring, and total wheat crops in the U.S.S.R. The reports included in this evaluation are the CMR's of August 5, September 7, and October 5, 1977. LACIE estimates on the U.S.S.R. grain situation are compared to those provided by a USDA Interagency Task Force (herein referred to as the "Task Force") composed of the Foreign Agricultural Service (FAS), the Environmental Research Service (ERS), the Agricultural Stabilization and Conservation Service (ASCS), the Agricultural Marketing Service, and the Office of the General Sales Manager. Although LACIE estimates are provided for each region of the U.S.S.R. no comparison can be made at this level since the Task Force estimates are for the national level only.

The only Task Force estimate made available in October was that for total wheat production. Therefore, the October LACIE estimates are compared to the September Task Force estimates.

The LACIE winter wheat area estimate characteristically increases during the late season due to the confusion of hay and row crops with small grains. This inability to differentiate between small grains and other crops is caused primarily by the use of single acquisitions from specific time periods during the growing season. To avoid this confusion the thresholding procedure described in section 1.3 was utilized by CAS.

7.1 PRODUCTION ESTIMATES

The LACIE and Task Force estimates of production for winter, spring, and total wheat are shown in table 7-1. The LACIE August estimates for spring wheat do not contain the estimates for the regions of Tyumen and the Northwest since no usable acreage data existed for these regions. For September and October there were no LACIE estimates for Tyumen and the Northwest but the estimates given in the CMR's (and shown in table 7-1) include an estimate based on historic data for these regions. The August estimates for spring and total wheat are significantly different from the respective Task Force

TABLE 7-1.— PRODUCTION

Wheat crop type	Number of segments (n/M)	U.S.S.R. Task Force estimate (MT × 10^6)	LACIE		RD[a] (%)	Value of test statistic
			Estimate (MT × 10^6)	CV (%)		
August 5, 1977						
Winter	658/1157	60.0	63.0	4.4	4.8	1.09[N]
Spring	491/1412	45.0	34.6	9.2	-30.1	-3.27*
Total	1149/2569	100.0	97.6	4.3	-7.6	-1.77*
September 7, 1977						
Winter	713/1157	60.0	63.9	4.3	6.1	1.4[N]
Spring	782/1416	40.0	37.9	7.2	-5.5	-0.8[N]
Total	1495/2573	100.0	101.8	3.8	1.8	0.5[N]
October 5, 1977						
Winter	553/1149	60.0	60.8	4.6	1.3	0.3[N]
Spring	899/1377	40.0	38.3	7.0	-4.4	-0.6[N]
Total	1452/2526	100.0	99.1	3.9	-0.9	-0.2[N]

[a]RD = Relative difference = $\left(\frac{\text{LACIE - Task Force}}{\text{LACIE}} \times 100\right)\%$.

[b]The total wheat average yield is derived as the software in use does not produce a yield for total wheat.

[c]Data are not available.

[N]LACIE estimate is not significantly different from U.S.S.R. Task Force estimate at the 10-percent level.

*LACIE estimate is significantly different from U.S.S.R. Task Force estimate at the 10-percent level.

C-2

counterparts. However, with inclusion of this historic estimate of production in these two regions, the production estimates for spring and total wheat are not significantly different from the corresponding Task Force estimates at the 10-percent level. There is no significant difference between the LACIE and Task Force production estimates for winter, spring, or total wheat released in the September and October CAS reports.

The relative difference for total wheat production estimates decreased in magnitude in each successive month, from -7.6 percent in August to -0.9 percent in October.

Since the difference between the latest LACIE and Task Force total wheat production estimates is not statistically significant, it would be reasonable to assume that the bias in the production estimate is very small. With a CV of 3.9 percent and a negligible bias, the LACIE at-harvest production estimate satisfies the 90/90 criterion.

7.2 AREA ESTIMATES

The August, September, and October LACIE and Task Force area estimates are shown in table 7-2. As in the case of production, the LACIE August estimate for spring wheat does not include an estimate for the Tyumen and Northwest regions but the September and October LACIE estimates do include an estimate for these regions based on historical data. The LACIE winter and total wheat estimates for all three months are significantly different (at the 10-percent level) from the corresponding Task Force estimates. The difference for spring wheat was significant in August but not in September or October. However, if the historical area estimates of 0.7 million hectares (1.75 million acres) for the Tyumen and Northwest regions were added to the August spring wheat area estimates, the difference between the LACIE and Task Force area estimates for spring and total wheat would not be significant at the 10-percent level. The CV of the area estimate for each type of wheat in every month is small indicating the high degree of dependability in the area estimate.

TABLE 7-2.— AREA

Wheat crop type	Number of segments (n/M)	U.S.S.R. Task Force estimate (ha × 10^6)	LACIE Estimate (ha × 10^6)	LACIE CV (%)	RD[a] (%)	Value of test statistic
August 5, 1977						
Winter	658/1157	22.0	24.3	2.7	9.5	3.5*
Spring	491/1412	42.0	38.9	4.3	-8.0	-1.9*
Total	1149/2569	64.0	63.2	2.8	-1.3	-0.5[N]
September 7, 1977						
Winter	713/1157	20.8	24.6	2.7	15.4	5.7*
Spring	782/1416	41.2	41.0	2.9	-0.5	0.2[N]
Total	1495/2573	62.0	65.6	2.1	5.5	2.6*
October 5, 1977						
Winter	553/1149	20.8	22.6	3.3	8.0	2.4*
Spring	899/1377	41.2	42.6	2.6	3.3	1.3[N]
Total	1452/2526	62.0	65.2	2.0	5.1	2.6*

[a]RD = relative difference = $\left(\frac{\text{LACIE - Task Force}}{\text{LACIE}} \times 100\right)\%$.

[b]The total wheat average yield is derived as the software in use does not produce a yield for total wheat.

[c]Data are not available.

[N]LACIE estimate is not significantly different from U.S.S.R. Task Force estimate at the 10-percent level.

*LACIE estimate is significantly different from U.S.S.R. Task Force estimate at the 10-percent level.

Moreover, a CV of less than 3 percent for the spring wheat area estimate with an insignificant difference between the LACIE and Task Force estimates indicates that the LACIE spring wheat area estimate supports the 90/90 criterion.

7.3 YIELD ESTIMATES

The LACIE and Task Force estimates of yield for the U.S.S.R. are shown in table 7-3. The LACIE winter wheat estimate was not significantly different (at the 10-percent level) from the corresponding Task Force estimate in the August CMR, but the difference was significant in the September and October CMR's. The LACIE yield estimate for winter wheat was low in every month. This was due partly to the effect of area overestimates in the low-yield regions, giving more weight to these regions and thus biasing the (weighted) average yield.

The LACIE and Task Force spring wheat yield estimates were significantly different in August but not in September or October. The LACIE spring wheat yield estimate of 8.9 quintals/hectares in August is the second lowest yield of the decade; however, this yield estimate is based largely on early-season meteorological data obtained prior to June. These data indicate soil moisture shortages in major parts of the spring wheat region which have since been alleviated to some extent.

TABLE 7-3.— YIELD

Wheat crop type	U.S.S.R. Task Force estimate (ql/ha)	LACIE		RD[a] (%)	Value of test statistic
		Estimate (ql/ha)	CV (%)		
August 5, 1977					
Winter	27.0	25.9	3.4	-4.2	-1.2[N]
Spring	11.0	8.9	8.7	-23.6	-2.7*
Total	16.0	15.4[b]	c	-3.9	c
September 7, 1977					
Winter	28.8	26.0	3.6	-10.8	-3.0*
Spring	9.7	9.3	7.1	-4.3	-0.6[N]
Total	16.1	15.5[b]	c	-3.9	c
October 5, 1977					
Winter	28.8	26.8	3.6	-7.5	-2.1*
Spring	9.7	9.0	6.9	-7.8	-1.1[N]
Total	16.1	15.2[b]	c	-5.9	c

[a]RD = relative difference = $\left(\frac{\text{LACIE - Task Force}}{\text{LACIE}} \times 100\right)\%$.

[b]The total wheat average yield is derived as the software in use does not produce a yield for total wheat.

[c]Data are not available.

[N]LACIE estimate is not significantly different from U.S.S.R. Task Force estimate at the 10-percent level.

*LACIE estimate is significantly different from U.S.S.R. Task Force estimate at the 10-percent level.

APPENDIX A

PHASE III ACCURACY ASSESSMENT METHODOLOGY

APPENDIX A

PHASE III ACCURACY ASSESSMENT METHODOLOGY

A.1 INTRODUCTION

This appendix contains mathematical details of the techniques used in accuracy assessment. The methods used in comparing the LACIE estimates for acreage, yield, and production with the reference standard are presented in section A.2. The techniques used to study errors in the LACIE estimates are discussed in section A.3.

A.2 COMPARISON OF LACIE ESTIMATES WITH REFERENCE STANDARDS

The reference standards to which the LACIE estimates are compared are the USDA/SRS estimates for the United States and the USDA/FAS estimates for foreign countries. The statistic used for making these comparisons is the relative difference (RD) defined as follows:

$$RD = \left(\frac{\text{LACIE} - \text{STANDARD}}{\text{LACIE}} \right) \times 100\%$$

where LACIE stands for the LACIE estimate of wheat production, area, or yield and STANDARD represents the corresponding reference standard estimate. This definition expresses the difference between the two estimates as a percentage of the LACIE estimate.

Significance tests of no difference are made only at the region or country level for the LACIE production, area, and yield estimates for spring wheat, winter wheat, and total wheat. For a significance test, the LACIE estimate (of wheat production, area, or yield) is assumed to be normally distributed with unknown mean μ and variance σ^2_{LACIE}. A test of the hypothesis

$$H_o : \mu = \text{STANDARD}$$

versus the alternative hypothesis

$$H_A : \mu \neq \text{STANDARD}$$

is then made using this assumption. The test statistic is given by

$$Z = \frac{\text{LACIE} - \text{STANDARD}}{\hat{\sigma}_{\text{LACIE}}} \qquad \text{(A-1)}$$

which, under the null hypothesis, is approximately normally distributed with mean 0 and variance 1. The null hypothesis is rejected in favor of the alternative at the α-level of significance if

$$|Z| > z_{\alpha/2}$$

where $z_{\alpha/2}$ is the $\left(1 - \frac{\alpha}{2}\right)$ critical point of the standard normal distribution. For $\alpha = 0.10$, $z_{\alpha/2} = 1.645$, and if $|Z| > 1.645$, it is concluded that the mean of the LACIE estimator is significantly different from the reference standard estimate.

A.3 ERROR SOURCES IN LACIE

The techniques used to study errors in the estimates of acreage, yield, and production are discussed respectively in sections A.3.1, A.3.2, and A.3.3 of this appendix.

A.3.1 ACREAGE

This section contains a description of the methods used to estimate the following:

a. The errors in segment wheat proportion estimates (section A.3.1.1)

b. Wheat acreage at the state and higher levels (section A.3.1.2)

c. The variance of the wheat acreage estimates (section A.3.1.3)

d. The bias in the acreage estimates for large areas having ground truth available for a subset of their LACIE segments (section A.3.1.4)

e. The relative variances of the sampling and classification errors in stratum wheat acreage estimates (section A.3.1.5)

A.3.1.1 Error in Proportion Estimates at the Segment Level

This section describes the statistical calculations used to compare CAMS wheat proportion estimates for blind sites with the corresponding ground truth values. Let N be the number of segments allocated to a region (state or higher level) and let n be the number of blind sites selected randomly from these N segments. For a region, let $\hat{X}_i$ represent the CAMS estimate of the proportion of wheat in the *ith* segment and let X_i represent the ground truth proportion of wheat in the *ith* segment, where $i = 1, \ldots, N$. Then the average error μ_D is given by

$$\mu_D = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{X}_i - X_i\right) \qquad \text{(A-2)}$$

The estimate of μ_D is given by

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_i - X_i\right) \qquad \text{(A-3)}$$

where the summation is taken over the n blind sites. Letting $D_i = \hat{X}_i - X_i$, we may estimate the variance of $\bar{D}$ by

$$S_{\bar{D}}^2 = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}{n-1} \qquad \text{(A-4)}$$

Lower and upper confidence limits for the population average difference μ_D are given by

$$\mu_{D_L} = \bar{D} - t_{1-\alpha/2}S_{\bar{D}} \;,\quad \mu_{D_U} = \bar{D} + t_{1-\alpha/2}S_{\bar{D}} \qquad \text{(A-5)}$$

where $t_{1-\alpha/2}$ is the value of the $1-\alpha/2$ percentage point, from the Student's t distribution with (n-1) degrees of freedom, corresponding to the desired confidence level of $1-\alpha$.

The hypothesis $\mu_D = 0$ (i.e., no bias) is rejected at the α-level of significance if $\left|\bar{D}/S_{\bar{D}}\right| > t_{1-\alpha/2}$, or equivalently, if the confidence interval given by equation (A-5) does not contain zero.

A.3.1.2 Acreage Estimation

This section gives a brief summary of the methods used to estimate wheat acreage. These methods are described in detail in appendix B of the CAS Requirements Document.*

A.3.1.2.1 Background of Sample Allocation

The LACIE sample allocation in the U.S. Great Plains (USGP) region is based upon a two-stage stratified sampling scheme in which counties represent the primary sampling units (substrata) and 5- × 6-nautical-mile segments are secondary sampling units. The criterion for determining the total sample size was the ability to achieve a sampling error of 2 percent or less for the country wheat acreage estimates.

Sample segments were allocated to the counties based on relative weights derived from agriculture and wheat acreage reported in 1969 agriculture census statistics. Depending upon the relative weights, counties were designated as Group I (at least one sample segment in the county), Group II (at most one sample segment in a county), or Group III (no sample segments in the county). All Group II counties in a CRD (stratum) were combined to determine the number of segments allocated to the Group II part of the CRD.

---

*Crop Assessment Subsystem (CAS) Requirements Vol IV (Rev. B) (Change Notice, March 8, 1977), JSC-11329, LACIE C00200.

In this appendix any reference to the CAS Requirements Document indicates this specific document.

A probability proportional to size (PPS) procedure was applied to select the Group II counties in a CRD which were to receive these segments.

Once the number of segments to be allocated to each county was determined, the sample segments were selected at random within the agricultural area of the county. For further details of the LACIE sampling scheme refer to the CAS Requirements Document (JSC-11329)

A.3.1.2.2 Aggregation of Acreage Estimates

Wheat acreage estimates are made for each CRD, state, and region (group of states) in the USGP. However, no estimate is made for a state if it does not contain three or more segments satisfactorily processed by CAMS. Segment data may be lost due to the following cases of nonresponse:

a. The sample segment being obscured by cloud cover

b. Landsat data quality being insufficient to permit processing

c. Landsat data acquisition failing to register with the reference Landsat image

d. Failure of acquisition/processing procedures to provide an acceptable estimate

No replacement is allowed if a sample segment is not workable by CAMS.

A CRD acreage estimate consists of three components:

1. An acreage estimate for the Group I counties in the CRD for which segment data exist. (A Group I county is treated as a Group III county if it does not have at least one segment with an acceptable proportion estimate.)

2. An acreage estimate for the entire set of Group II counties in the CRD if there is at least one segment with an acceptable

proportion estimate in this set of counties. (Otherwise, the Group II counties are all treated as Group III counties.)

3. An acreage estimate for the Group III counties, including the Group I and Group II counties being treated as Group III counties.

The wheat acreage estimates for these three components are computed using a stratified random sampling estimator for the Group I counties, a PPS estimator for the Group II counties, and a ratio estimator for the Group III counties.*

There are three categories of Group III acreage estimates, depending on the number of segments in a CRD for which data are available. Categories 1, 2, and 3 correspond respectively to three or more segments, one or two segments, and no segments having data available. The ratio used for the Group III estimator is the ratio of historical wheat acreages for Group III counties to Group I and Group II counties. For category 1 estimates the ratio is based on historical acreages in the CRD. For category 2 and category 3 estimates the ratio is based on acreages in the state containing the CRD for which the estimate is being made.

The CRD wheat acreage estimate is obtained from the sum of the wheat acreage estimates for Group I, II, and III counties. Next, aggregation of the CRD acreage estimates gives a state wheat acreage estimate, and summation of the state acreage estimates gives the regional wheat acreage estimate. For specific aggregation formulas, see appendix B in the Cas Requirements Document.

In a mixed wheat area, separate aggregations are performed for spring and winter wheat and the total wheat acreage estimate is obtained by summing the results. This is done at the CRD and higher levels.

*For details on these standard estimation procedures, see Sampling Techniques by W.G. Cochran, Wiley, 1963.

A.3.1.3 Acreage Variance Estimation

The acreage variance estimation for a CRD requires an estimate of within-county variance for each of the Group I and Group II counties in the CRD. Often there is only one sample segment in a county and hence no direct estimate of the within-county variance is possible. Therefore, an indirect method is employed. This method uses a regression approach and is based on the assumption that the historical county proportions are well correlated with the CAMS proportions. The method consists of (1) forming homogeneous groups of counties in a state with respect to the within-county variability, (2) performing regression for the CAMS segment wheat proportion estimate onto the county historical wheat proportion, and (3) taking the residual mean square error (MSE) for an estimate of the within-county variance for each county in the group.

For estimation of a CRD acreage variance, the acreage variance components for Group I and Group II counties are estimated independently. For Group I counties it is computed according to the variance formula for a stratified random sampling scheme.[1] The appropriate inputs of county sizes, number of sample segments, and within-county variance estimates are obtained using the abovementioned procedure. Similarly, the variance formula for a PPS estimator[1] is used to compute the Group II acreage variance estimate. It requires all of the inputs mentioned in the Group I case plus the probabilities of selection of Group II counties for sample allocation. These probabilities are those utilized in determining which of the Group II counties in a CRD receive sample segments.

The acreage variance component for the Group III counties depends directly on Groups I and II variances and contributes to the CRD

[1]Cf = Sampling Techniques, by W. G. Cochran, Wiley, 1963.

acreage variance indirectly through the ratio utilized to obtain the Group III acreage estimate. The formulas used to calculate the acreage variance for the Group III counties are described in appendix B of the CAS Requirements Document. As mentioned above, there are three categories of Group III acreage estimates and each category has a different formula for the variance estimate. For category 1 the variance estimate depends on the acreage estimates for all the Group I and Group II counties in the CRD; for categories 2 and 3 it depends on the acreage estimates for all of the Group I and Group II counties in the state.

If data are available for at least three segments in each CRD in the state, the acreage variance estimate is computed by adding the variance estimates for the CRD's in the state. Otherwise, the state variance estimate is obtained using an aggregation procedure which accounts for the dependence between various CRD acreage estimates in a state.

Since the state acreage estimates are obtained independently, the acreage variance estimates at both the regional and country level are computed by adding the state acreage variance estimates.

In a mixed wheat area, separate aggregations are performed for estimating the variance of the spring and winter wheat acreage estimates at the CRD and higher levels. In each case the estimation procedure is the same as that described above for each aggregation level. The acreage variance estimates at the CRD and state levels for the total wheat case are obtained from the previously described variance formulas using total wheat acreage estimates for sample segments and the historical total wheat for counties in the area. For higher levels the total wheat acreage variance estimates are computed by taking the sum of the variance estimates for the states involved. The CRD and state level variance estimates for the total wheat case are not unbiased; therefore, the method of determining variance of a total wheat acreage estimate in a mixed wheat area is considered approximate.

A.3.1.4 Acreage Bias Estimation

The method for estimating bias described in this section is valid for any area having a sufficient number of blind sites to represent the bias. In this report it is applied at the state and higher levels.

The LACIE estimate of wheat acreage for a given area can be written

$$\hat{A} = \sum_{i=1}^{n} W_i \hat{X}_i \tag{A-6}$$

where $\hat{A}$ is the estimated wheat acreage, $\hat{X}_i$ is the wheat proportion estimate in the *ith* LACIE segment, n is the number of processed LACIE segments, and $\{W_i\}_{i=1}$ are known weights based on historical and cartographic data.*

Corresponding to $\hat{A}$ is the true acreage, A, which can be written

$$A = \sum_{i=1}^{n} W_i^* C_i \tag{A-7}$$

*The precise definition of W_i depends on whether the *ith* segment is used as part of a Group III estimate.

where C_i is the true wheat acreage for the county containing the *ith* segment and W_i^* is the value of the weight which would give perfect Group III estimates of wheat acreage for unsampled counties.

We can now write

$$\hat{X}_i = C_i + \left(X_i - C_i\right) + \left(\hat{X}_i - X_i\right)$$
$$= C_i + \delta_i + \varepsilon_i$$

where X_i is the true wheat proportion of the *ith* segment, δ_i is the sampling error and ε_i is the classification error. Since sampling is unbiased, we assume $E\left(\delta_i\right) = 0$; however, we do not assume unbiased classification. Instead, let θ be an average segment bias; i.e.,

$$E\left(\varepsilon_i\right) = \theta$$

The bias in A is defined by $E(\hat{A} - A)$, which is thus given by

$$B = E(\hat{A} - A) = E\left(\sum_{i=1}^{n} W_i \hat{X}_i - \sum_{i=1}^{n} W_i^* C_i\right)$$

$$= \sum_{i=1}^{n} W_i E\left(C_i + \delta_i + \varepsilon_i\right) - \sum_{i=1}^{n} W_i^* C_i$$

$$= \sum_{i=1}^{n} \left(W_i - W_i^*\right) C_i + \theta \sum_{i=1}^{n} W_i \qquad \text{(A-8)}$$

Note that the first term of equation (A-8) represents a bias caused by the failure of the Group III ratios to be exact;

(i.e., $W_i \neq W_i^*$), whereas the second term is the average segment bias multiplied by the sum of the W_i.

At present, only the second term of equation (A-8) will be estimated, since good county-level data are not available for estimating the first term. The second term is estimated by (1) breaking up the large area into strata (not necessarily connected) for which the bias is assumed to be approximately constant; (2) estimating θ_k by $\hat{\theta}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \left(\hat{X}_i - X_i\right)$, the average proportion error on a segment level in the *kth* stratum; and (3) aggregating $\hat{\theta}$ over the strata.

If $\hat{B}$ represents the AA estimate of bias due to classification, a 90-percent confidence interval for β, the real bias, can be constructed by

$$(\hat{B} - 1.645\hat{\sigma},\ \hat{B} + 1.645\hat{\sigma})$$

where $\hat{\sigma}^2$ is an estimate of the variance of $\hat{\beta}$.

If we assume $\mathrm{Var}\left(\varepsilon_i\right) \doteq \sigma_{ck}^2$ (a constant) within the *kth* stratum, then σ_{ck}^2 can be estimated by

$$\hat{\sigma}_{ck}^2 = \sum_{i=1}^{n_k} \frac{\left(\hat{X}_i - X_i - \hat{\theta}\right)^2}{n_k - 1}$$

and $\mathrm{Var}(\hat{B})$ can be estimated by $\hat{\mathrm{Var}}(\hat{B}) = \sum_k \hat{\sigma}_{ck}^2 \left(\sum_{i=1}^{n_k} W_{ki}\right)^2$

where W_{ki} is the weight for the *ith* segment in the *kth* stratum.

A.3.1.5 Contribution of Sampling and Classification to Acreage Estimation Error

This section describes the calculation of the contribution of sampling and classification errors to the variance of the LACIE production estimate.

A.3.1.5.1 Approach

The variance of the LACIE acreage estimate for a large area (e.g., zone) can be written

$$V^2 = \sum_i V_i \sigma_i^2$$

where σ_i^2 is the variance of the acreage estimate for the *ith* county and V_i is a weight which depends on the size of the county, the number of segments in the county, etc. (Refer to CAS Requirements Document, appendix B for details.)

The variance σ_i^2 represents a mean-squared deviation between the LACIE estimate for the county wheat proportion and the true county wheat proportion. This variance is caused mainly by two factors: sampling errors and classification errors.

In accuracy assessment, it is desirable to quantify the contribution of each of these error sources to the large area production estimate. The LACIE production estimate depends on acreage and yield estimation errors in a complicated way; hence, it is unrealistic to assume the error in the production estimate can be written as a sum of uncorrelated random variables representing acreage and yield errors. Instead, the effect of a particular error source is measured by the reduction in the LACIE production variance which would be achieved if that source were eliminated.

It will be assumed (section A.3.1.5.2) that the *ith* county acreage error variance σ_i^2 can be written $\sigma_i^2 = \sigma_c^2 + \lambda^2\sigma_s^2$, where σ_c^2 is a contribution due to classification, and $\lambda^2\sigma_s^2$ is a contribution due to sampling. To determine the effect of no classification error, the variance of the LACIE production estimate will be calculated using $\rho\sigma_i^2$ instead of σ_i^2 where ρ is an estimate of the ratio $\frac{\lambda^2\sigma_s^2}{\sigma_c^2 + \lambda^2\sigma_s^2}$. Similarly, the effect of no sampling error is estimated by replacing σ_i^2 by $(1 - \rho)\sigma_i^2$. This procedure is described in detail in section A.3.3.5 of this appendix. The following two sections describe the methods employed for estimating sampling and classification variances and the function ρ.

A.3.1.5.2 Acreage Regression Models

For counties with one sample segment, the LACIE estimate of the *ith* county wheat proportion can be written

$$\hat{X}_i = C_i + \left(X_i - C_i\right) + \left(\hat{X}_i - X_i\right)$$
$$= C_i + \varepsilon_i + \delta_i \qquad \text{(A-9)}$$

where

$\hat{X}_i$ = LACIE estimate of the wheat proportion in the sampled segment

C_i = true (current year) proportion of wheat in the county

X_i = true proportion of wheat in the sampled segment

ε_i = sampling error = $X_i - C_i$

δ_i = classification error = $Y_i - X_i$

It will be assumed that for some reasonably large area (e.g., a zone) the errors ε_i and δ_i have the following properties:

ε_i and δ_i are uncorrelated

$$E(\varepsilon_i) = 0$$

$$E(\delta_i | X_i) = \lambda^* X_i + \theta$$

$$V(\varepsilon_i) = \sigma_s^2$$

$$V(\delta_i | X_i) = \sigma_c^2$$

It is also assumed that there is a linear model relating the current year county proportions, C_i, to the historical proportions which will be denoted by Z_i; i.e.,

$$C_i = \alpha + \beta Z_i + \zeta_i \quad \text{(A-10)}$$

where $E(\zeta_i) = 0$, $V(\zeta_i) = \sigma_H^2$, $Cov(\zeta_i, \varepsilon_i) = Cov(\zeta_i, \delta_i) = 0$, and α and β are regression coefficients.

From the above assumptions and definitions, three basic regression models are obtained:

a. True segment proportion versus historical county proportion — from the definition of ε_i,

$$X_i = C_i + \varepsilon_i$$

$$= \alpha + \beta Z_i + \zeta_i + \varepsilon_i \quad \text{(A-11)}$$

It follows that

$$E(X_i) = \alpha + \beta Z_i \quad \text{(A-12)}$$

$$V(X_i) = \sigma_H^2 + \sigma_s^2 \quad \text{(A-13)}$$

b. LACIE segment proportion versus ground truth segment proportion – from the definition of δ.

$$\hat{X}_i = X_i + \delta_i \qquad \text{(A-14)}$$

It follows that

$$E\left(\hat{X}_i | X_i\right) = X_i + \lambda^* X_i + \theta \qquad \text{(A-15)}$$

$$V\left(\hat{X}_i | X_i\right) = \sigma_c^2 \qquad \text{(A-16)}$$

Writing $\lambda = 1 + \lambda^*$, one obtains

$$E\left(\hat{X}_i | X_i\right) = \lambda X_i + \theta \qquad \text{(A-17)}$$

$$V\left(\hat{X}_i | X_i\right) = \sigma_c^2 \qquad \text{(A-18)}$$

c. LACIE segment proportion versus historical county proportion – from equations (A-12) through (A-18),

$$E\left(\hat{X}_i\right) = E_{X_i}\left(E\left(\hat{X}_i | X_i\right)\right) = E_{X_i}\left(\lambda X_i + \theta\right) = \lambda\left(\alpha + \beta Z_i\right) + \theta \qquad \text{(A-19)}$$

$$V\left(\hat{X}_i\right) = E_{X_i}\left(V\left(\hat{X}_i | X_i\right)\right) + V_{X_i}\left(E\left(\hat{X}_i | X_i\right)\right) = \sigma_c^2 + \lambda^2\left(\sigma_H^2 + \sigma_s^2\right) \qquad \text{(A-20)}$$

As stated previously, one would like to estimate $\rho = \dfrac{\lambda^2 \sigma_s^2}{\sigma_c^2 + \lambda^2 \sigma_s^2}$.

None of the three regression models permits an estimate of σ_s^2 separately from σ_H^2; i.e., one can only estimate $\sigma_s^2 + \sigma_H^2$, not σ_s^2 alone. If current year county proportions C_i were available, σ_H^2 could be estimated, but since this is not the case,

$$\rho^* = \frac{\lambda^2\left(\sigma_s^2 + \sigma_H^2\right)}{\sigma_c^2 + \lambda^2\left(\sigma_s^2 + \sigma_H^2\right)}$$ will be estimated instead of ρ. If $\sigma_H^2 << \sigma_s^2$ (a reasonable assumption) then $\rho^* \approx \rho$.

A.3.1.5.3 Normality Assumptions — Maximum Likelihood Estimation of ρ^*

Suppose a given zone has m blind site segments and n ordinary (i.e., not blind site) segments, and let the blind site segments be numbered 1 to m. It is assumed that ground truth wheat proportions $\left\{X_i\right\}_{i=1}^{m}$ are available for the blind sites and LACIE estimates $\left\{\hat{X}_i\right\}_{i=1}^{m+n}$ are available for all the segments. It is also assumed that historical wheat proportions $\left\{Z_i\right\}_{i=1}^{m+n}$ are available for the counties containing the segments. If $\sigma_H^2 << \sigma_S^2$ so that $\rho \approx \rho^*$ the regression models equations (A-11 through A-20) can be used to obtain

$$E\left(X_i\right) = \alpha + \beta Z_i; \; V\left(X_i\right) = \sigma_s^2 \qquad i = 1,\cdots,m$$

$$E\left(\hat{X}_i|X_i\right) = \lambda X_i + \theta; \; V\left(\hat{X}_i|X_i\right) = \sigma_c^2 \qquad i = 1,\cdots,m$$

$$E\left(\hat{X}_i\right) = \theta + \lambda\alpha + \lambda\beta Z_i; \; V\left(\hat{X}_i\right) = \lambda^2\sigma_s^2 + \sigma_c^2 \qquad i = m+1,m+n$$

If there is one segment per county, then the errors ε_i and δ_i are independent for different values of i, and hence the likelihood function of the sample can be written

$$L = \prod_{i=1}^{m} f\left(X_i,\hat{X}_i\right) \prod_{i=m+1}^{m+n} h\left(\hat{X}_i\right) \tag{A-21}$$

where $f\left(X_i,\hat{X}_i\right)$ is the joint density of X_i and $\hat{X}_i$ for $i = 1,\cdots,m$ and $\left(h\ \hat{X}_i\right)$ is the density of $\hat{X}_i$ for $i = m+1,\cdots,m+n$.

The function $\prod_{i=1}^{m} f(X_i, \hat{X}_i)$ can be written $\prod_{i=1}^{m} f(X_i, \hat{X}_i) =$ $\prod_{i=1}^{m} f(\hat{X}_i | X_i)\ g(X_i)$ where $f(\hat{X}_i | X_i)$ is the conditional density of $\hat{X}_i$ given X_i and $g(X_i)$ is the density function of X_i.

If normality is assumed, $\prod_{i=1}^{m} f(X_i, \hat{X}_i) = \prod_{i=1}^{m} \frac{1}{\sigma_c \sqrt{2\pi}}$

$$\exp\left\{-\frac{1}{2\sigma_c^2} \sum_{i=1}^{m} (\hat{X}_i - \lambda X_i - \theta)^2\right\} \frac{1}{\sigma_s \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_s^2} \sum_{i=1}^{m} (X_i - \alpha - \beta Z_i)^2\right\}$$

and

$$\prod_{i=m+1}^{m+n} h(\hat{X}_i) = \frac{1}{\left(\lambda^2\sigma_s^2 + \sigma_c^2\right)^{1/2}\sqrt{2\pi}} \exp\left\{-\frac{1}{2\left(\lambda^2\sigma_s^2 + \sigma_c^2\right)} \sum_{i=m+1}^{m+n} (\hat{X}_i - \lambda\alpha - \theta - \lambda\beta Z_i)^2\right\}$$

Letting $Q = -2\log L - \log 2\pi$,

$$Q = m \log \sigma_s^2 + m \log \sigma_s^2 + n \log(\sigma_c^2 + \lambda^2\sigma_s^2) + \frac{D_m}{\sigma_c^2} + \frac{T_m}{\sigma_s^2} + \frac{T_n}{\sigma_c^2 + \lambda^2\sigma_s^2} \tag{A-22}$$

where

$$D_m = \sum_1^m (\hat{X}_i - \lambda X_i - \theta)^2$$

$$T_m = \sum_1^m (X_i - \alpha - \beta Z_i)^2$$

$$T_n = \sum_{i=m+1}^{m+n} (\hat{X}_i - \lambda\alpha - \theta - \lambda\beta Z_i)^2$$

One attempts to maximize L by finding a stationary point of Q:

$$-\frac{1}{2}\frac{\partial Q}{\partial \alpha} = \frac{\sum_{1}^{m} (X_i - \alpha - \beta Z_i)}{\sigma_s^2} + \frac{\sum_{m+1}^{m+n} \lambda(\hat{X}_i - \lambda\alpha - \theta - \lambda\beta Z_i)}{\sigma_c^2 + \lambda^2\sigma_s^2} = 0 \qquad \text{(A-23)}$$

$$-\frac{1}{2}\frac{\partial Q}{\partial \beta} = \frac{\sum_{1}^{m} Z_i(X_i - \alpha - \beta Z_i)}{\sigma_s^2} + \frac{\sum_{m+1}^{m+n} \lambda Z_i(\hat{X}_i - \lambda\alpha - \theta - \lambda\beta Z_i)}{\sigma_c^2 + \lambda^2\sigma_s^2} = 0$$

(A-24)

$$-\frac{1}{2}\frac{\partial Q}{\partial \theta} = \frac{\sum_{1}^{m} (\hat{X}_i - \lambda X_i - \theta)}{\sigma_c^2} + \frac{\sum_{m+1}^{m+n} (\hat{X}_i - \lambda\alpha - \theta - \lambda\beta Z_i)}{\sigma_c^2 + \lambda^2\sigma_s^2} = 0 \qquad \text{(A-25)}$$

$$-\frac{1}{2}\frac{\partial Q}{\partial \lambda} = \frac{\sum_{1}^{m} X_i(\hat{X}_i - \lambda X_i - \theta)}{\sigma_c^2} + \frac{-n\lambda\sigma_s^2 + \sum_{i=m+1}^{m+n} (\beta Z_i + \alpha)(\hat{X}_i - \lambda\alpha - \theta - \lambda\beta Z_i)}{\sigma_c^2 + \lambda^2\sigma_s^2}$$

$$+ \frac{\lambda^2\sigma_s^2 T_n}{\left(\sigma_c^2 + \lambda^2\sigma_s^2\right)^2} = 0 \qquad \text{(A-26)}$$

$$\frac{\partial Q}{\partial \sigma_c^2} = \frac{m}{\sigma_c^2} + \frac{n}{\lambda^2\sigma_s^2 + \sigma_c^2} - \frac{D_m}{\sigma_c^4} - \frac{T_n}{\left(\lambda^2\sigma_s^2 + \sigma_c^2\right)^2} = 0 \qquad \text{(A-27)}$$

$$\frac{\partial Q}{\partial \sigma_s^2} = \frac{m}{\sigma_s^2} + \frac{n\lambda^2}{\lambda^2\sigma_s^2 + \sigma_c^2} - \frac{T_m}{\sigma_s^4} - \frac{T_n\lambda^2}{\left(\sigma_c^2 + \lambda^2\sigma_s^2\right)^2} = 0 \qquad \text{(A-28)}$$

Equations (A-23) through (A-29) must be solved for the parameters α, β, θ, λ, σ_c^2, and σ_s^2. If $\hat{\alpha}$, $\hat{\beta}$, $\hat{\theta}$, $\hat{\lambda}$, $\hat{\sigma}_c^2$, and $\hat{\sigma}_s^2$ represent the solution to equations (A-23) and (A-29), then the invariance theorem for maximum likelihood estimation can be used to obtain

$$\hat{\rho} = \frac{(\hat{\lambda})^2\hat{\sigma}_s^2}{\hat{\sigma}_c^2 + (\hat{\lambda})^2\hat{\sigma}_s^2} \tag{A-29}$$

as the maximum likelihood estimate of ρ.

The equations (A-23) through (A-29) are nonlinear but can be solved using numerical techniques. Newton's Method was used to solve the equations for this report; i.e., if $u^{(k)}$ is an estimate of the solution vector $u = (\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda}, \hat{\sigma}_c^2, \hat{\sigma}_s^2)$ at the *kth* step, then

$$u^{(k+1)} = u^{(k)} - F^{-1}f\left(u^{(k)}\right) \tag{A-30}$$

where $f\left(u^{(k)}\right) = (f_1,\cdots,f_6)^T$ is the vector of the left sides of equations (A-23) through (A-29) evaluated at $u^{(k)}$ and $F = (F_{ij}) = \frac{\partial f_i}{\partial u_j}$

In practice, it was simpler to use the parameter transformations

$$r = \frac{\sigma_s^2}{\lambda^2\sigma_s^2 + \sigma_c^2} \tag{A-31}$$

and

$$s = \lambda^2\sigma_s^2 + \sigma_c^2 \tag{A-32}$$

and solve for α, β, θ, λ, r, and s. Again, the invariance theorem can be used to give

$$\hat{\rho} = \hat{\lambda}^2\hat{r}$$

A.3.1.5.4 Accuracy of $\hat{\rho}$

Since $\hat{\rho}$ is an extremely complicated function of the data, it is impossible to write down the variance of $\hat{\rho}$ for finite sample sizes m and n. However, the asymptotic variance of $\hat{\rho}$ can be estimated using the information matrix; i.e., if

$$V = (V_{ij}) = E\left\{\frac{-\partial^2 \log L}{\partial u_i \; \partial u_j}\right\}$$

and $g(u) = g\left(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda}, \hat{\sigma}_c^2, \hat{\sigma}_s^2\right)$ is a differentiable function of the parameter vector u, then the variance of g(u) is asymptotic to

$$[g'(u)]^T V^{-1} g'(u)$$

where
$$g'(u) = \left(\frac{\partial g}{\partial u_1}, \cdots, \frac{\partial g}{\partial u_6}\right)^T \qquad \text{(A-33)}$$

Thus, in our case, $g(u) = \dfrac{\lambda^2\sigma_s^2}{\lambda^2\sigma_s^2 + \sigma_c^2}$

$$g'(u) = \left[0, 0, 0, \; 2\lambda\sigma_s^2\sigma_c^2\left(\lambda^2\sigma_s^2 + \sigma_c^2\right)^{-2}, \; -\lambda^2\sigma_s^2\left(\lambda^2\sigma_s^2 + \sigma_c^2\right)^{-2}, \; -\frac{\lambda^2\sigma_c^2}{\left(\sigma_c^2 + \lambda^2\sigma_s^2\right)^2}\right] \qquad \text{(A-34)}$$

To estimate V, the observations $\{X_i\}$, $\{Y_i\}$, and $\{Z_i\}$ and the estimated parameters $\left(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda}, \hat{\sigma}_c^2, \text{ and } \hat{\sigma}_s^2\right)$ were substituted into the matrix $H = (h_{ij}) = \dfrac{\partial^2 \log L}{\partial u_i \; \partial u_j}$. Then equation (A-33) was used to obtain an approximate variance for $\hat{\rho}$.

A.3.1.5.5 Coefficients of Variation of a Large Area Estimate Due To Classification and Sampling Errors

Let $\hat{\rho}$ be the ratio of the within-county sampling variance estimate to the total within-county area variance estimate as defined in (A-29). Assuming that this ratio also applies to a large area, the variances of the large area estimate due to classification and sampling are given by

$$\hat{\eta}^2 = (1 - \hat{\rho})\hat{V}^2$$

and

$$\hat{\nu}^2 = \rho\hat{V}^2$$

where $\hat{\eta}^2$, $\hat{\nu}^2$, and $\hat{V}^2$ denote the classification variance, the sampling variance, and the acreage variance for the large area estimate, respectively. Consequently, the estimated CV of a large area estimate $\hat{A}$ due to classification is given by

$$\hat{CV}(\hat{A}|C) = \frac{\hat{\eta}}{\hat{A}}$$

and the estimated CV of large area estimate due to sampling is given by

$$\hat{CV}(\hat{A}|S) = \frac{\hat{\nu}}{\hat{A}}$$

where $\hat{CV}(\hat{A}|C)$ and $\hat{CV}(\hat{A}|S)$ are often casually referred to in LACIE as the classification CV and sampling CV, respectively.

A.3.2 YIELD

This section contains a description of the methods used to predict yields (section A.3.2.1) and to estimate yield prediction error (section A.3.2.2). In Phase II no estimate of yield bias was made.

A.3.2.1 Yield Prediction

Most of the yield predictions made in LACIE are provided by the Center for Climatic and Environmental Assessment (CCEA) of NOAA.

They are produced from multiple linear regression yield models* developed on historical weather and yield data. Usually these models cover a state but in some cases they cover part of a state or part of two states and in some cases they overlap.

In a given state there is either one yield stratum or two. In the first case the state yield prediction is that given by the CCEA model. In the second case the state yield prediction is given by:

$$Y = P/A \tag{A-35}$$

where P is the production estimate (section A.3.3.1) and A is the acreage estimate (section A.3.1.2) for the state. The yield prediction at the region or country level is also obtained from equation (A-35), with P and A in that case being the production and acreage estimates at the corresponding level.

A.3.2.2 Estimation of the Yield Prediction Error

CCEA provides estimates of the mean squared yield prediction error at the stratum level. In the CAS Requirements Document it is shown that at the state, region, or country levels the estimate of the mean squared yield prediction error for a given area (state, region, or country) is

$$\sigma^2 = \bar{Y}^2 \left[\frac{S^2}{P^2} + \frac{V^2}{A^2} - 2 \frac{\Sigma Y_i V_i^2}{P\ A} \right] \tag{A-36}$$

where

S^2 = estimated mean squared prediction error of the production estimate P for the area

V^2 = estimated variance of the acreage estimate A for the area

*Wheat Yield Models for the United States (LACIE 00431), National Aeronautics and Space Administration, Johnson Space Center, Houston, Texas, June 1975.

Y_i = yield estimate for the *ith* pseudo zone in the area

V_i^2 = estimated variance of the acreage estimate for the *ith* pseudo zone in the area

In the case where there is only one yield stratum for a state, the yield prediction error for the state is given directly by the CCEA model.

A.3.3 PRODUCTIOI

Chis section contains descriptions of the methods used to do the following:

a. Estimate wheat production (section A.3.3.1).

b. Estimate the variance in the wheat production estimate (section A.3.3.2).

c. Estimate the bias in the wheat production estimate (section A.3.3.3).

d. Evaluate whether LACIE is satisfying the 90/90 criterion (section A.3.3.4).

e. Determine the effect of errors in area, yield, sampling, and classification on the production variance (section A.3.3.5).

A.3.3.1 Production Estimation

At the CRD level the production estimate is obtained by multiplying the area estimate and the yield prediction for the CRD. The area estimate is made for the CRD itself but the yield prediction is made for a group of CRD's in a state (section A.3.2.1

The production estimates for the state and higher levels are obtained by simply adding the estimates for all the CRD's in the area.

A.3.3.2 Production Variance Estimation

Since the production estimate is the product of an acreage estimate and a yield prediction, the measure of variability in the estimate should properly be called the production mean squared prediction error. However, in this report, for simplicity this quantity will be called the production variance.

Since the yield predictions are made for a group of CRD's it is not possible to obtain independent production variance estimates at the CRD level; hence, the estimates of production variance are made only at the state and higher levels. The estimation procedures are described in detail in appendix B of the CAS Requirements document.

A.3.3.3 Production Bias Estimation

The production bias at the state level is given by

$$\begin{aligned} B_{P_i} &= E(\hat{P}_i - P_i) \\ &= E(\hat{P}_i) - P_i \qquad \text{(A-37)} \\ &= E(\hat{A}_i \hat{Y}_i) - A_i Y_i \end{aligned}$$

where A_i, Y_i, and P_i are respectively the true values of the acreage, yield, and production for the *Nth* state in question, and $\hat{A}_i$, $\hat{Y}_i$, and $\hat{P}_i$, are the corresponding estimates for these quantities. Assuming $\hat{A}_i$ and $\hat{Y}_i$ are independent, one obtains

$$B_{P_i} = E(\hat{A}_i)E(\hat{Y}_i) - A_i Y_i \qquad \text{(A-38)}$$

If one further assumes that Y_i is unbiased, then $E(\hat{Y}_i) = Y_i$, and

$$\begin{aligned} B_{P_i} &= Y_i[E(\hat{A}_i) - A_i] \qquad \text{(A-39)} \\ &= Y_i B_{A_i} \end{aligned}$$

where B_{A_i} is the acreage bias for the *ith* state. The quantities Y_i and B_{A_i} are unknown, but an estimate $\hat{B}_{P_i}$ for B_{P_i} can be obtained by using the estimates for Y_i and B_{A_i} described in sections A.3.2.1 and A.3.1.4, respectively. Thus,

$$\hat{B}_{P_i} = \hat{Y}_i \hat{B}_{A_i} \qquad \text{(A-40)}$$

The variance of $\hat{B}_{P_i}$ is given by

$$\mathrm{Var}\left(\hat{B}_{P_i}\right) = Y_i^2 \,\mathrm{Var}\left(\hat{B}_{A_i}\right) + E^2\left(\hat{B}_{A_i}\right) \mathrm{Var}(\hat{Y}_i) + \mathrm{Var}\left(\hat{B}_{A_i}\right) \mathrm{Var}(\hat{Y}_i)$$

and estimated by

$$\hat{\mathrm{Var}}\left(\hat{B}_{P_i}\right) = \hat{Y}_i^2 \,\hat{\mathrm{Var}}\left(\hat{B}_{A_i}\right) + \hat{B}_{A_i}^2 \,\hat{\mathrm{Var}}(\hat{Y}_i) - \hat{\mathrm{Var}}\left(\hat{B}_{A_i}\right) \hat{\mathrm{Var}}(\hat{Y}_i)$$

For the nine-state level, the production bias estimate $\hat{B}_P$ is

$$\hat{B}_P = \sum \hat{B}_{P_i} = \sum \hat{Y}_i \hat{B}_{A_i}$$

and the estimate of its variance is $\sum \hat{\mathrm{Var}}\left(\hat{B}_{P_i}\right)$. The relative bias of the production estimate $R(\hat{B}_P)$ is estimated by expressing the production bias as a percentage of the LACIE production estimate; that is, by

$$R(\hat{B}_P) = \frac{\sum \hat{Y}_i \hat{B}_A}{\sum \hat{A}_i \hat{Y}_i} \times 100\% \qquad \text{(A-41)}$$

A.3.3.4 Evaluating the 90/90 Criterion

Let $\hat{P}$ be the LACIE estimate of wheat production for the region or country, and let P be the true wheat production of the same region or country. The accuracy goal of the LACIE is a 90/90 at-harvest criterion for wheat production, which is given by the following probability statement.

$$\Pr\left[|\hat{P} - P| \le 0.1P\right] \ge 0.90 \qquad \text{(A-42)}$$

This states that the accuracy goal is for the LACIE estimate of wheat production to be within 10 percent of the true wheat production with a probability of at least 0.9.

It is assumed that the LACIE estimate, $\hat{P}$, is normally distributed with mean P + B and variance $\sigma^2_{\hat{P}}$, where B is the bias given by

$$B = E(\hat{P}) - P$$

Under this assumption, equation (A-42) may be written as

$$\Pr\left[\frac{-0.1 - 0.9\,\frac{B}{P+B}}{CV(\hat{P})} \le Z \le \frac{0.1 - 1.1\,\frac{B}{P+B}}{CV(\hat{P})}\right] \ge 0.90 \qquad \text{(A-43)}$$

where $Z = \frac{P - (P+B)}{\sigma\hat{P}}$ follows the standard normal distribution, N(0,1), and $CV(\hat{P})$ is the coefficient of variation of $\hat{P}$ defined by

$$CV(\hat{P}) = \frac{\sigma_{\hat{P}}}{E(\hat{P})} = \frac{\sigma_{\hat{P}}}{P+B} \qquad \text{(A-44)}$$

The term $\frac{B}{P+B}$ is called the relative bias of $\hat{P}$ and is given by

$$\frac{B}{P+B} = \frac{E(\hat{P}) - P}{E(\hat{P})}$$

It follows that the accuracy goal of LACIE is attained if

$$\Phi\left[\frac{0.1 - 1.1\,\frac{B}{P+B}}{CV(\hat{P})}\right] - \Phi\left[\frac{-0.1 - 0.9\,\frac{B}{P+B}}{CV(\hat{P})}\right] \ge 0.90 \qquad \text{(A-45)}$$

where Φ represents the cumulative standard normal distribution. The enclosed region of figure A-1 indicates combinations of $CV(\hat{P})$ and relative bias for which equation (A-40) is satisfied.

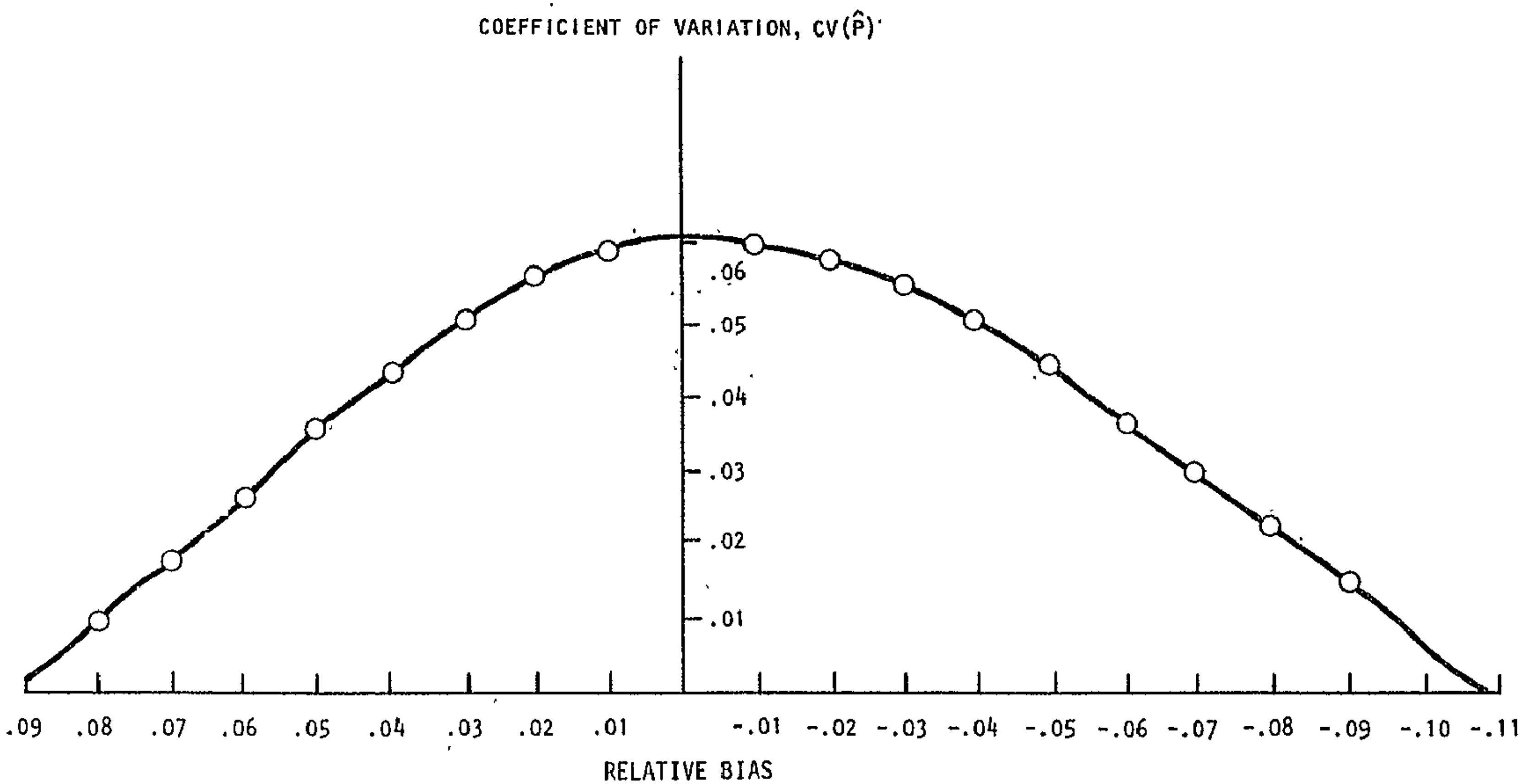

Figure A-1.—Diagram showing value of relative bias and $CV(\hat{P})$ for which 90/90 criterion is satisfied.

The estimates of $CV(\hat{P})$ are provided by CAS and the estimates of the relative bias are obtained using the method described in section A.3.3.3.

A proper evaluation of the 90/90 criterion can only be made toward the end of the season since the results for spring wheat are normally not available before August. In order to gauge how well LACIE is performing early in the season when only winter wheat data are available, a method of projecting the winter wheat results for the 5- or 7-state level to the 9-state total harvestable wheat level using Phase II results was developed. Since adequate blind site proportions are not available in the early season, the relative difference between the LACIE and USDA estimates is taken as an estimate of the relative bias. The "projected" relative difference at the 9-state level is given by the equation

$$RD_9' = -12.3 \frac{RD_{77}}{RD_{76}} \qquad \text{(A-46)}$$

where -12.3 is the Phase II final relative difference at the 9-state level, RD_{77} is the current month relative difference in 1977 for the 5- or 7-state winter wheat production estimate, and RD_{76} is the corresponding relative difference for the same month in 1976. The values for RD_{76} are given in table A-1.

Similarly, the "projected" $CV(\hat{P})$ at the 9-state level is given by

$$CV_9' = 5 \frac{CV_{77}}{CV_{76}} \qquad \text{(A-47)}$$

where 5 is the Phase II final $CV(\hat{P})$ at the 9-state level, CV_{77} is the current month $CV(\hat{P})$ for the 5- or 7-state winter wheat production estimate, and CV_{76} is the corresponding CV for the same month in 1976. The values for CV_{76} are given in table A-1. The 7-state results are used if they are available for both years.

TABLE A-1.— PHASE II CV'S AND RELATIVE DIFFERENCES

Date	Area, state	Phase II CV for production	Phase II relative difference
Feb.	5	11	-4.9
Mar. 25	5	10	-9.9
Apr. 8	5	8	-8.5
May 7	5	8	-1.6
June 8	5	7	+11.4
	7	8	+1.7
June 29	5	7	+12.7
	7	7	+4.7
July 9	5	7	-3.7
	7	7	-7.9
Aug. 11	5	7	-4.2
	7	7	-5.6
	9	6	-14.7
Sept. 9	5	7	-6.6
	7	7	-6.6
	9	5	-13.6
Oct. 8	5	7	-6.6
	7	7	-6.5
	9	5	-13.8
Dec. 17	5	7	-7.2
(final)	7	7	-7.2
	9	5	-12.3

After RD_9' and CV_9' have been calculated, inference as to whether the 90/90 criterion has been satisfied can be made by determining whether these values fall within the enclosed area in figure A-1. If they do, it is said that the current LACIE estimates *support* (rather than *satisfy*) the 90/90 criterion since the determination was based on projections which may or may not be accurate.

A.3.3.5 Effect of Errors in Acreage, Yield, Sampling, and Classification on the Production Variance

The production variance consists of two major error components: acreage and yield. The acreage error may be further subdivided into sampling and classification errors. The effect of a particular error is determined by the reduction in the production variance estimate when the error is omitted from the calculation of that estimate. If there is only one yield stratum in a zone (state), the production variance is calculated at the zone level and aggregated to higher levels. If a zone contains more than one yield stratum, it is subdivided into pseudozones, which are the intersections of the zone with the various yield strata. The production variance estimate is then calculated at the pseudozone level and aggregated to the zone and higher levels.

Suppose the zone consists of H pseudozones, $G_1, G_2, \cdots, G_H$, with acreage estimates $A_{Z1}, A_{Z2}, \cdots, A_{ZH}$ and yield predictions $Y_{Z1}, Y_{Z2}, \cdots, Y_{ZH}$, respectively. Then the estimate of the production variance at the zone level is given by the following equation, which also appears in appendix B of the CAS Requirement Document.

$$S_Z^2 = \sum_{i=1}^{H} \left(V_{Zi}^2 Y_{Zi}^2 + U_{Zi}^2 A_{Zi}^2 - V_{Zi}^2 U_{Zi}^2 \right)$$

$$+ 2 \sum_{i=2}^{H} \sum_{\ell=1}^{i-1} Y_{Zi} Y_{Z\ell} \left(\sum_{j \in G_i} \sum_{k \in G_\ell} \Psi_{jk} \right) \qquad \text{(A-48)}$$

where

U_{Zi}^2 = the estimate of the yield variance for the *ith* pseudozone

V_{Zi}^2 = the area variance estimate for the *ith* pseudozone

Ψ_{jk} = the estimated covariance between A_j in G_i and A_k in G_ℓ .

In order to determine the production variance without a given error term, equation (A-48) must be rederived with that term omitted. Let S_{ZA}^2, S_{ZY}^2, S_{ZS}^2, and S_{ZC}^2 be the state production variances without acreage, yield, sampling, and classification errors, respectively. One obtains the following expressions f these quantities.

$$S_{ZA}^2 = \sum_{i=1}^{H} \left(U_{Zi}^2 A_{Zi}^2 - V_{Zi}^2 U_{Zi}^2\right) \tag{A-49}$$

$$S_{ZY}^2 = \sum_{i=1}^{H} \left(V_{Zi}^2 Y_{Zi}^2 - V_{Zi}^2 U_{Zi}^2\right) + 2 \sum_{i=2}^{H} \sum_{\ell=1}^{i-1} Y_{Zi} Y_{Z\ell} \left(\sum_{j \in G_i} \sum_{k \in G_\ell} \Psi_{jk}\right) \tag{A-50}$$

$$S_{ZS}^2 = \sum_{i=1}^{H} \left[(1 - \hat{P}) V_{Zi}^2 Y_{Zi}^2 + U_{Zi}^2 A_{Zi}^2 - (1 - \hat{P}) V_{Zi}^2 U_{Zi}^2\right] + 2 \sum_{i=2}^{H} \sum_{\ell=1}^{i-1} Y_{Zi} Y_{Z\ell} \left(\sum_{j \in G_i} \sum_{k \in G_\ell} \Psi_{jk}\right) \tag{A-51}$$

$$s^2_{ZC} = \sum_{i=1}^{H} \left(\hat{\rho} V^2_{Zi} Y^2_{Zi} + U^2_{Zi} A^2_{Zi} - \hat{\rho} V^2_{Zi} U^2_{Zi} \right)$$

$$+ 2 \sum_{i=2}^{H} \sum_{\ell=1}^{i-1} Y_{Zi} Y_{Z\ell} \left(\sum_{j \in G_i} \sum_{k \in G_\ell} \Psi_{jk} \right) \qquad \text{(A-52)}$$

where $\hat{\rho}$ is defined in (A-29).

Let S^2_{rA}, S^2_{rY}, S^2_{rS}, and S^2_{rC} be the regional-level production variance estimates without acreage, yield, sampling, and classification errors, respectively. These estimates can be obtained from the following expressions.

$$S^2_{rA} = \sum_{Z=1}^{R} S^2_{ZA} + \sum_{Z=1}^{R} \sum_{Z'=1}^{R} S_{rZZ'} \qquad \text{(A-53)}$$

$$S^2_{rY} = \sum_{Z=1}^{R} S^2_{ZY} \qquad \text{(A-54)}$$

$$S^2_{rS} = \sum_{Z=1}^{R} S^2_{ZS} + \sum_{Z=1}^{R} \sum_{Z'=1}^{R} S_{rZZ'} \qquad \text{(A-55)}$$

$$S^2_{ZC} = \sum_{Z=1}^{R} S^2_{ZC} + \sum_{Z=1}^{R} \sum_{Z'=1}^{R} S_{rZZ'} \qquad \text{(A-56)}$$

Here R is the total number of zones in the region and $S_{rZZ'} = 0$ if *Zth* and *Z'th* zones have no yield strata in common. Otherwise,

$$S_{rZZ'} = \sum_{K=1}^{C} A_{rZK} A_{rZ'K} U^2_{rK} \qquad \text{(A-57)}$$

where

A_{rZK} = the area estimate for the pseudozone corresponding to yield stratum K in zone Z of region r

U_{rK}^2 = the squared prediction error for the *Kth* yield stratum common to zones Z and Z'

C = the number of yield strata common to the *Zth* and *Z'th* zones

The estimates of the corresponding variances for a country are obtained by adding the corresponding estimates for all the regions in the country. These computations assume that the regional production estimates are uncorrelated.

APPENDIX B

PHASE III INTENSIVE TEST SITES

APPENDIX B

PHASE III INTENSIVE TEST SITES

To accomplish the objectives of accuracy assessment, data including ground truth, aircraft photographs, and Landsat multispectral scanner imagery were gathered from 23 intensive test sites. Because of factors such as atmospheric effects and data dropout, acceptable imagery was received and processed for only 17 intensive test sites and only six of these sites were processed during more than one biowindow. These 23 sites were located in the states of Idaho, Indiana, Kansas, Montana, North and South Dakota, Texas, and Washington (table B-1). These states are from four regions: the Northwestern United States, the Great Lakes, and the Southern and Northern Great Plains.

TABLE B-1.— LACIE PHASE III INTENSIVE TEST SITES

Segment	State	County	Center coordinates		Site size		Wheat type (a)
			N. Lat.	W. Long.	Statute mile	Km	
1965	N. Dakota	Burke	48°53.2'	102°10.0'	5×6	8×9.7	S
1966	N. Dakota	Williams	48°19.2'	103°24.7'	5×6	8×9.7	S
1967	N. Dakota	Divide	48°53.6'	103°10.9'	2×10	3×16	S
1687	S. Dakota	Hand 1	44°35.0'	98°58.0'	5×6	8×9.7	S&W
1986	S. Dakota	Hand 2	44°21.0'	98°45.1'	5×6	8×9.7	S&W
1969	Montana	Toole	48°53.0'	111°46.5'	2×10	3×16	S&W
1970	Montana	Liberty	48°44.0'	110°51.0'	2×10	3×16	S&W
1971	Montana	Hill	48°42.0'	109°55.0'	2×6	3×9.7	S&W
1973	Washington	Whitman	46°50.4'	117°48.3'	3×3	4.8×4.8	S&W
1975	Idaho	Oneida	42°04.5'	112°29.5'	3×3	4.8×4.8	S&W
1976	Idaho	Franklin	42°08.0'	111°58.0'	3×3	4.8×4.8	S&W
1977	Idaho	Bannock	42°56.5'	112°25.5'	3×3	4.8×4.8	S&W
1978	Texas	Randall	35°09.5'	102°04.4'	3×3	4.8×4.8	W
1979	Texas	Deaf Smith	34°52.2'	102°22.3'	3×3	4.8×4.8	W
1980	Texas	Oldham	35°15.0'	102°32.0'	3×3	4.8×4.8	W
1981	Indiana	Shelby	39°27.6'	85°47.2'	3×3	4.8×4.8	W
1982	Indiana	Madison	40°13.5'	85°37.5'	3×3	4.8×4.8	W
1983	Indiana	Boone	40°05.7'	86°33.5'	3×3	4.8×4.8	W
1960	Kansas	Morton	37°16.0'	101°54.0'	5×6	8×9.7	W
1962	Kansas	Saline	38°41.8'	97°28.4'	3×3	4.8×4.8	W
1963	Kansas	Rice	38°17.0'	98°12.7'	3×3	4.8×4.8	W
1964	Kansas	Ellis	38°50.1'	99°13.0'	3×3	4.8×4.8	W
1988	Kansas	Finney	38°10.2'	100°43.2'	5×6	8×9.7	W
1987	Minnesota	Polk	47°49.0'	96°41.0'	5×6	8×9.7	S

[a]As indicated by ground truth: S = spring wheat; W = winter wheat; SW = spring and winter wheat.

APPENDIX C

METHOD OF DESIGNATING SEGMENTS AS SPRING, WINTER, OR MIXED

APPENDIX C

METHOD OF DESIGNATING SEGMENTS AS SPRING, WINTER, OR MIXED

The USDA/SRS winter wheat and spring wheat production estimates for each county in South Dakota and Montana for the years 1965 to 1976 were taken into consideration to determine the county contribution to the state total production for each crop type. A county-to-state contribution threshold of 1 percent was taken for each crop type. If a county, containing allocated segments, contributed 1 percent or more to the state winter wheat production, its segments were designated as winter — similarly for spring wheat. This divided the counties into three groups: pure spring, pure winter, and mixed. Further, those counties in the pure spring and pure winter groups were then designated mixed if the within county contribution for either crop type to the total wheat for the county was between 25 and 75 percent. For example, a county may have contributed more than 1 percent to state winter wheat production but less than 1 percent to state spring wheat production. However, spring wheat could make up 50 percent of the county's total wheat production. In this case, the county is designated as mixed. The resulting segment designations are in the following tables. In the group of segments not to be used, those that are asterisked are to be processed by CAMS as mixed segments for evaluation purposes but are not to be used in the aggregations.

SOUTH DAKOTA

Mixed Segments

1666	1670	1696	1805
1485	1676	1687	1808
1486	1677	1688	1697
1668	1488	1699	
1669	1686	1689	

Spring Wheat Segments

1665	1673	1498	1548	1690
1484	1674	1679	1681	1784
1667	1675	1499	1599	
1487	1489	1525	1755	
1671	1678	1680	1756	

Winter Wheat Segments

1597	1804
1683	1694
1598	1806
1803	1696

Segments Not To Be Used

1800*	1809
1801	1811*
1802*	1812
1807*	1813

MONTANA

Mixed Segments

1528	1735	1936	1741	1941
1529	1933	1737	1939	1539
1929	1934	1937	1534	1540
1732	1736	1738	1535	1942
1932	1935	1739	1536	1555
1733	1530	1938	1537	
1734	1531	1740	1538	

Spring Wheat Segments

1532	1943	1546
1533	1543	1547
1940	1544	1945
1541	1944	1946
1542	1545	1559

Winter Wheat Segments

1725	1930	1745	1549	1552
1728	1931	1948	1550	1556
1729	1742	1747	1753	1557
1730	1743	1750	1101	1104
1731	1744	1949	1102	1558

Segments Not To Be Used

1928	1553
1947	1103
1551	1554
1752*	

*To be processed by CAMS as mixed for evaluation but not for aggregation.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1978—771-083/ 699