Lockheed Electronics Company, Inc.
A SUBSIDIARY OF LOCKHEED CORPORATION
1830 NASA Road 1, Houston, Texas 77058
Tel. 713-333-5411

Ref: 644-1269
Contract NAS 9-15800
Job Order: 74-715

TECHNICAL MEMORANDUM

A REVISED SCREENING PROCEDURE FOR LARGE AREA CROP INVENTORY EXPERIMENT (LACIE) PHASE III DATA IN THE U.S. GREAT PLAINS

By

R. S. Chhikara

Approved By: B L Carroll

B. L. Carroll, Manager
Earth Observations
Systems Department

August 1979

LEC-12723

CONTENTS

1. INTRODUCTION

The screening procedure implemented in Phase III of the Large Area Crop Inventory Experiment (LACIE) for detecting Classification and Mensuration Subsystem (CAMS) segment estimates which indicated a significant change from the historical wheat acreages of their corresponding counties was based on the extreme studentized deviate statistic. However, the distribution of the statistic was assumed normal, and the method of testing many outliers present in a data set was not adequately developed at the time of its implementation. Since then, these defects have been corrected, and a screening procedure that utilizes appropriate statistical methodology has been implemented on the Crop Assessment Subsystem (CAS) Development System.

2. THE SCREENING PROCEDURE

The basic approach to screening remains the same as that given in reference 1. The variable defined by the ratio $\bar{r} = \bar{y}/X$, where $\bar{y}$ is the average of CAMS wheat proportion estimates for sample segments in a county and X is the historical wheat proportion for the county, provides an estimate of the change in wheat acreage of a county in a current year from X. In reference 1, the ratio $r = y/X$, where y is the CAMS wheat proportion estimate for a segment in the county, was considered for an estimate of this change in wheat acreage; however, when a large within-county variance exists for the CAMS segment estimates, as observed for certain counties in Colorado (ref. 2), r is likely to be an unreliable estimate for the county, as a bias is possible. Conversely, the average ratio $\bar{r}$ can be regarded as a better estimate, since the use of $\bar{r}$ eliminates the bias in a county estimate that might be caused by deletion of a subset of segments. The logarithmic transformation is applied to the values of $\bar{r}$ to maintain the normal approximation hold for the underlying distribution.

Winter and spring wheat regions are treated separately; each region is stratified by the size of historical wheat acreage at the county level. The winter wheat region is divided into the following four strata:

$$S_1 = \{X: \ 0 < X \leq 5\}$$

$$S_2 = \{X: \ 5 < X \leq 15\}$$

$$S_3 = \{X: \ 15 < X \leq 30\}$$

$$S_4 = \{X: \ X > 30\}$$

For the spring wheat region, three strata are formed:

$$S_1' = \{X: \ 0 < X \leq 5\}$$

$$S_2' = \{X: \ 5 < X \leq 25\}$$

$$S_3' = \{X: \ X > 25\}$$

There is one less stratum in the spring wheat case mainly because its region is smaller than the winter wheat region.

The statistical procedure for testing outliers is quite different from the one used in reference 1. The critical values for the test of significance do not correspond to the percentage points of the normal distribution; instead, these values are developed by using the Monte Carlo technique for the test statistic computed for the normal samples. Simulations are used because the exact distribution of the test statistic cannot be obtained. The significance test is developed to detect as many as 19 outliers in a data set. Thus far in the statistical literature, detection has been developed for a maximum of four outliers (ref. 3). This test procedure was documented in reference 4.

The screening was applied to final (Phase III) CAMS segment estimates obtained after thresholding from each of the above strata. Counties were flagged whose z values (where $z = \log \bar{r}$) were declared outliers. Consequently,

CAMS segment estimates for these counties were deleted from the CAS data base, and the counties were treated as "Group III" in the CAS aggregation.

The revised U.S. Great Plains (USGP) winter and spring wheat acreage estimates by states are presented in table 1. Also given are the official LACIE Phase III estimates. The numerical results show that there is a better agreement between the revised estimates and the U.S. Department of Agriculture/Statistical Reporting Service[1] (USDA/SRS) (end of season) estimates when compared to the agreement between the official LACIE Phase III estimates and the USDA/ESCS estimates.

3. REVISED VERSUS OFFICIAL LACIE PHASE III ACREAGE ESTIMATES

The revised and the official LACIE Phase III acreage estimates given in table 1 were obtained on the CAS Development System using the LACIE Phase III CAS data base with thresholding and screening applied to the final CAMS segment estimates. Thresholding precedes screening and is the same in both cases; thus, the difference between the two estimates is due only to the use of different screening procedures. The official estimates (column 5) correspond to the screening procedure employed previously in LACIE Phase III, and the revised estimates (column 9) correspond to the updated procedure discussed in the previous section. (There is a slight difference between the estimates given in column 5 and the officially reported LACIE estimates because of a difference in the number of significant digits to which the CAMS estimates were carried in the two CAS systems – development and operational.)

A state-by-state comparison between the two estimates shows that the revised LACIE winter wheat acreage estimates are closer to their corresponding USDA/ESCS estimates. The only exception is for the state of Oklahoma, where the difference between the two estimates was slightly larger. For the seven

[1]Now called the Economics, Statistics, and Cooperative Service (ESCS).

TABLE 1.– LACIE PHASE III ACREAGE ESTIMATES OBTAINED WITH THE SCREENING PROCEDURE APPLIED TO CAMS THRESHOLDED DATA

Region	USDA/ESCS estimate ($\times 10^3$)		LACIE estimates							
			Previous screening				Revised screening			
			No. of segments used	Estimate ($\times 10^3$)		CV[a]	No. of segments used	Estimate ($\times 10^3$)		CV
	Sq hm	Acres		Sq hm	Acres			Sq hm	Acres	
Winter wheat										
Colorado	1 032	2 550	24	1 407	3 463	10	30	1 151	2 844	12
Kansas	4 899	12 100	106	5 064	12 508	4	109	5 012	12 380	4
Nebraska	1 194	2 950	39	1 393	3 440	9	43	1 310	3 237	10
Oklahoma	2 632	6 500	42	2 300	5 682	8	43	2 279	5 630	8
Texas	1 903	4 700	29	1 816	4 485	16	26	1 989	4 913	15
USSGP[b]	11 660	28 800	240	11 975	29 578	4	251	11 742	29 004	4
South Dakota	275	680	15	367	906	35	20	363	897	44
Montana	1 134	2 800	43	1 366	3 375	8	54	1 200	2 965	9
USGP-7[c] total	13 069	32 280	298	13 709	33 860	3	325	13 306	32 866	4
Spring wheat										
Minnesota	1 304	3 222	38	953	2 354	15	38	910	2 247	18
North Dakota	3 704	9 150	73	3 733	9 200	4	72	3 733	9 220	4
Montana	915	2 260	32	882	2 178	10	35	904	2 233	12
South Dakota	946	2 336	35	787	1 944	15	29	911	2 249	12
USNGP[d] total	6 870	16 968	178	6 347	15 676	4	174	6 457	15 948	4

[a]Coefficient of variation.

[b]U.S. Southern Great Plains.

[c]USGP seven-state.

[d]U.S. Northern Great Plains.

states combined, the difference between the LACIE estimate and the USDA/ESCS estimate was reduced from 640 000 square hectometers (1.58 million acres) to 239 000 square hectometers (0.59 million acres) as a result of updating the screening procedure.

In obtaining the revised estimate, 325 CAMS segment estimates were used compared to 298 CAMS segment estimates used for the LACIE estimate at the seven-state level, a deletion of 27 segments. This is an expected result provided the assumption of a uniform change in county wheat acreages from epic year to current year holds for counties in each stratum. The outlier test procedure applied in reference 1 is conservative, as it tends to declare false outliers more often than is allowed under the 5-percent level of significance presently used. Although the revised screening flags counties and results in deleting all segments in them, compared to flagging and deleting individual segments, there should be no adverse effect.

No significant change is noticed in spring wheat estimates obtained using different screening procedures. There is a difference of 110 121 square hectometers (272 000 acres) between the revised and the official LACIE estimates, with the largest change in South Dakota. The revised spring wheat acreage estimate for South Dakota is below the ESCS estimate by 3.7 percent; the previous estimate was 16.8 percent below the ESCS estimate. There is very little change in the LACIE estimates of the three other spring wheat states of Minnesota, Montana, and North Dakota.

4. CONCLUSION

The revised screening procedure has a sound statistical basis and eliminates the two major drawbacks of the reference 1 procedure. The revision resulted in a substantial decrease in the official LACIE winter wheat acreage estimate and some increase in the spring wheat acreage estimate, bringing the two estimates into better agreement with corresponding USDA/ESCS estimates.

5. REFERENCES

1. A Segment Screening Procedure for U.S. Great Plains, an attachment to letter from R. S. Chhikara to W. McAllum, dated Sept. 9, 1977.

2. Colorado Blind Sites (Action Item RTEB SF3-134). A memorandum from Jon Erickson to the Transition Year Project Manager, dated June 15, 1978.

3. Rosner, Bernard: Percentage Points for the RST Many Outlier Procedure. Technometrics, 19, Aug. 1977.

4. Chhikara, R. S.; and Feiveson, A. H.: A Statistical Test Procedure for Detecting Multiple Outliers in a Data Set, JSC-14594, Technical Report, LEC-12910, Nov. 1978.