General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

E82-10013

AgRISTARS

FC-LO-00612
JSC-16819
CR-161030

APR 10 1981

"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Foreign Commodity Production Forecasting

March 1981

FISCAL YEAR 1980-81 IMPLEMENTATION PLAN IN SUPPORT OF TECHNICAL DEVELOPMENT AND INTEGRATION OF SAMPLING AND AGGREGATION PROCEDURES

NASA CR-161030

(E82-10013) FISCAL YEAR 1980-81 IMPLEMENTATION PLAN IN SUPPORT OF TECHNICAL DEVELOPMENT AND INTEGRATION OF SAMPLING AND AGGREGATION PROCEDURES (Lockheed Engineering and Management) 88 p HC A05/MF A01 CSCL 02C G3/43

N82-15491

Unclas
00013

Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1, Houston, Texas 77058

Lyndon B. Johnson Space Center
Houston, Texas 77058

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FC-LO-00612; JSC-16819		

4. Title and Subtitle

Fiscal Year 1980-81 Implementation Plan in Support of Technical Development and Integration of Sampling and Aggregation Procedures

5. Report Date

March 1981

6. Performing Organization Code

7. Author(s)

8. Performing Organization Report No.

LEMSCO-15168

9. Performing Organization Name and Address

Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1
Houston, Texas 77058

10. Work Unit No.

11. Contract or Grant No.

NAS 9-15800

12. Sponsoring Agency Name and Address

National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas 77058 Technical Monitor: J. D. Erickson

13. Type of Report and Period Covered

Technical Plan

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This document is a detailed plan for the development of sampling and estimation technology. The tasks described in this plan are those tasks to be addressed primarily in FY1980-81. The specific objectives of the FY1980-81 tasks are: (a) further refinements to the weighted aggregation procedure, (b) improved approaches for estimating within-stratum variance, (c) more intensive investigation of alternative sampling strategies such as full-frame sampling strategy, and (d) further developments in regard to a simulated approach for assessing the performance of the overall designed sampling and aggregation system.

17. Key Words (Suggested by Author(s))

Weighted aggregation procedure, simulation model, sampling frame, sampling unit size, within-stratum variance estimates, multivear models, rotation sample designs, full-frame sampling strategy, software conversion

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	81	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

FC-LO-00612
JSC-16819

FISCAL YEAR 1980-81 IMPLEMENTATION PLAN IN SUPPORT OF TECHNICAL DEVELOPMENT AND INTEGRATION OF SAMPLING AND AGGREGATION PROCEDURES

Job Order 72-420

This report describes Sampling and Aggregation activities of the Foreign Commodity Production Forecasting project of the AgRISTARS program.

PREPARED BY

Crop Applications Department

APPROVED BY

M. D. Pore, Supervisor
Statistical Analysis Section

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Resources Applications Division

Space and Life Sciences Directorate

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

March 1981

LEMSCO-15168

PRECEDING PAGE BLANK NOT FILMED

PREFACE

The Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing is a 6-year program of research, development, evaluation, and application of aerospace remote sensing for agricultural resources, which began in fiscal year 1980. This program is a cooperative effort of the National Aeronautics and Space Administration, the U.S. Agency for International Development, and the U.S. Departments of Agriculture, Commerce and the Interior.

The work which is the subject of this document was performed within the Earth Resources (Research/Applications) Division, Space and Life Sciences Directorate, at the Lyndon B. Johnson Space Center, National Aeronautics and Space Administration. Under Contract NAS 9-15800, personnel of Lockheed Engineering and Management Services Company, Inc., performed the tasks which contributed to the completion of this research.

PRECEDING PAGE BLANK NOT FILMED

PRECEDING PAGE BLANK NOT FILMED

CONTENTS

Section		Page
1.	INTRODUCTION	1-1
1.1	BACKGROUND	1-1
1.2	OBJECTIVES, APPROACHES, AND SCOPE	1-2
1.2.1	OVERALL SAMPLING AND AGGREGATION OBJECTIVES AND APPROACH IN AgRISTARS	1-2
1.2.2	SPECIFIC FY1980-81 OBJECTIVES AND APPROACH	1-2
1.3	SCOPE	1-3
2.	MULTICROP AGGREGATION PROCEDURES	2-1
2.1	THE GROUPED OPTIMAL AGGREGATION TECHNIQUE	2-1
2.1.1	CROP ACREAGE ESTIMATION	2-1
2.1.2	CROP PRODUCTION ESTIMATION	2-5
2.2	ESTIMATION OF MATRIX H	2-5
2.2.1	BACKGROUND	2-5
2.2.2	TASKS 1 AND 2	2-6
2.3	FURTHER TESTING OF STRATA GROUPING APPROACH	2-7
2.3.1	BACKGROUND	2-7
2.3.2	TASKS 3 AND 4	2-8
2.4	SIMULATION OF AgRISTARS CROP PRODUCTION	2-9
2.4.1	OBJECTIVES	2-9
2.4.2	TASKS 5, 6, 7, AND 8	2-9
2.4.3	SIMULATION MODEL FOR SEGMENT CROP PROPORTION ESTIMATION	2-11
2.4.4	DATA INPUT AND OUTPUT	2-13
3.	SAMPLING FRAME DEVELOPMENT	3-1
3.1	OBJECTIVES AND APPROACH	3-1

Section Page

3.2 SAMPLING UNIT SIZE STUDY 3-1

3.2.1 BACKGROUND 3-1

3.2.2 TASKS 9 AND 10 3-2

4. MULTICROP SAMPLE ALLOCATION 4-1

4.1 OBJECTIVE 4-1

4.2 INITIAL WITHIN-STRATUM VARIANCE ESTIMATES 4-1

4.2.1 TASK 11 4-1

4.2.2 TASK 12 4-1

4.3 ALLOCATION TO SUPPORT EXPLORATORY STUDIES 4-2

5. SEGMENT SELECTION AND LOCATION 5-1

6. STUDY OF ALTERNATIVE SAMPLING STRATEGIES 6-1

6.1 HARTLEY MULTIYEAR MODELS 6-1

6.1.1 HARTLEY MIXED ANALYSIS OF VARIANCE MODEL 6-1

6.1.1.1 Objectives and Approaches 6-1

6.1.1.2 Description of Hartley Mixed Analysis of Variance Model 6-1

6.1.1.3 Tasks 13 and 14 6-4

6.1.2 ROTATION SAMPLE DESIGNS 6-5

6.1.2.1 Objective and Approach 6-5

6.1.2.2 Description of Rotation Sample Designs 6-6

6.1.2.3 Tasks 15 and 16 6-9

6.2 FULL-FRAME SAMPLING STRATEGY 6-10

6.2.1 OBJECTIVES 6-10

6.2.2 TASKS 17 AND 18 6-10

6.2.2.1 Task 17 6-10

6.2.2.2 Task 18 6-10

Section | Page

7. SOFTWARE CONVERSION 7-1

7.1 TASK 19 7-1

7.2 TASK 20 7-1

7.3 TASK 21 7-1

8. ASSEMBLAGE OF DATA BASES 8-1

8.1 TASK 22 8-1

8.2 TASK 23 8-1

9. DOCUMENTATION 9-1

9.1 TASK 24 9-1

9.2 TASK 25 9-1

10. REFERENCES 10-1

Appendix

A. DEVELOPMENTAL TASKS: SUMMARIES OF BACKGROUND, ACTIONS, TIME, AND MANPOWER A-1

B. APPLIED TASKS: SUMMARIES OF CROP TYPES, BACKGROUND, TIME, AND MANPOWER IN VARIOUS COUNTIES B-1

C. SCHEDULES OF TASKS ACCORDING TO PRIORITY, TIME ALLOCATION, AND MANPOWER NEEDS C-1

D. DATA REQUIREMENTS D-1

PRECEDING PAGE BLANK NOT FILMED

ACRONYMS

AA	Accuracy Assessment
AgRISTARS	Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing
AI	analyst-interpreter
BLUE	best linear unbiased estimator
CAMS	Classification and Mensuration Subsystem
CAS	Crop Assessment Subsystem
EODLS	Earth Observations Division Laboratory System
ERSYS	Earth Resources System
FCPF	Foreign Commodity Production Forecasting
FY	fiscal year
JSC	Lyndon B. Johnson Space Center
LACIE	Large Area Crop Inventory Experiment Laboratory
LARS	Laboratory for Applications of Remote Sensing
MSE	mean square error
NASA	National Aeronautics and Space Administration
PDP	Programmed Data Processor
SAS	Statistical Analysis System
TAMU	Texas A&M University
UCB	University of California at Berkeley
USDA	U.S. Department of Agriculture
USDA/ESS	USDA/Economics and Statistics Service
USGP	U.S. Great Plains

1. INTRODUCTION

This implementation plan provides for the development of sampling and estimation technology supported by the Foreign Commodity Production Forecasting (FCPF) project of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program.

The purpose of this plan is to specify task objectives, descriptions, scope, data and resources, software requirements, and schedules for the fiscal year (FY) 1980-81 AgRISTARS-supported activities in the areas of multicrop sampling frame development, allocation, segment selection, aggregation, and variance estimation which are needed to support both the short term (FY1980-81) and long term objectives of the FCPF project of the AgRISTARS program. Those tasks required to support the integration of the overall procedures into a viable sampling and aggregation methodology are also included.

1.1 BACKGROUND

The FCPF project of the AgRISTARS program is designed to develop the capability to process large data sets in conjunction with objective information systems for performing multicrop inventorying using remote sensing technology. A rigorous development of a sampling and estimation technology is a critical component of the project.

In supporting the AgRISTARS/FCPF goals, the general sampling and aggregation approach is to continuously improve upon the existing methodology in order to achieve an efficient sampling strategy and reliable crop area and production forecasts.

Substantial effort went into developing an efficient sampling and aggregation strategy for crop acreage and production estimation for a single crop in the Large Area Crop Inventory Experiment (LACIE). The extension of the methodology to two or more crops occurred during the period before AgRISTARS. The tasks described in this sampling plan consist of the tasks to be addressed

primarily in FY1980-81 in support of the AgRISTARS objectives listed in section 1.2. These tasks (1 through 25) are more fully described in appendix A.

1.2 OBJECTIVES, APPROACHES, AND SCOPE

1.2.1 OVERALL SAMPLING AND AGGREGATION OBJECTIVES AND APPROACH IN AgRISTARS

For AgRISTARS, the overall objective of the sampling and aggregation procedures development program is to advance the multicrop sampling and aggregation approach to a level ready for inclusion in a future operational commodity production forecasting system. Specifically, it will support objective, timely, and reliable crop production forecasts at selected periods during the growing season for a range of crops in various countries.

The general approach requires that each crop region be partitioned into relatively homogenous subregions (called strata) within which samples of Landsat data are selected and machine-processed to identify and measure the areal extent by crop type of the crops of interest.

Further candidate improvements and integrations to be investigated in FY1980-81 for achieving a more efficient sampling and aggregation approach are detailed in this plan.

1.2.2 SPECIFIC FY1980-81 OBJECTIVES AND APPROACH

The specific objectives of the FY1980-81 tasks are to continue developments and refinements of various multicrop sampling and aggregation activities initiated during FY1978-79. Emphasis for the FY1980-81 tasks will be on further refinements to the grouped optimal aggregation technique, improved approaches for estimating within-stratum variance (with more reliance possibly upon machine-derived segment-level proportion estimates as opposed to reliance on poor historical estimates and/or analyst estimates), more intensive investigation of alternative sampling strategies such as a full-frame sampling strategy, and further developments in regard to a simulated approach for assessing the performance of the overall-designed sampling and aggregation system.

Specific tasks to be addressed include:

a. Further investigation of the optimal sampling unit size with emphasis on determining its relationship to the classification error variance [Primary emphasis will be the integration of findings obtained at the University of California at Berkeley (UCB) with further in-house investigations to arrive at a procedure for determining the appropriate sampling unit size.]

b. Additional developments and refinements to the grouped optimal aggregation technique with emphasis on improving the grouping logic and the approach to estimate the various input parameters such as the variance and covariance matrices of the historical acreage

c. Improvements to a priori estimates of within-stratum variances and to procedures for development of sampling frames (The utility of an automated procedure will be investigated for providing improvements in each of these areas. The capability to create crop proportion estimates at the sampling unit level will be investigated. This capability needs to be rapid but sufficiently accurate so as to improve upon the reliance of historical data for use in the within-stratum variance estimation.)

d. Investigations of a full-frame sampling strategy and of a multiyear estimation for improvements in multicrop inventorying (This activity will require interfacing with other support contractors who provide support to integration of such techniques into an overall-designed system.)

e. Determination of the feasibility of refining existing stratification procedures via more automated (hence, more rapid) approaches for analyzing Landsat data

1.3 SCOPE

The general scope of the FY1980-87 AgRISTARS sampling and aggregation activities involves many different crops including wheat, barley, rice, corn, soybeans, cotton, sorghum, and sunflowers over selected regions within the United States, Canada, India, Australia, U.S.S.R., Argentina, and Brazil. The scope for FY1980-81 entails primarily wheat, barley, corn, and soybeans over

portions of the yardstick [the U.S. Great Plains (USGP) and the major corn and soybean producing states in the United States] in support of development, test, and evaluation of a sampling and aggregation methodology for application in foreign areas. These applied tasks (26 through 82) are described in appendix B. The initial exploratory and pilot testing are to be completed in the yardstick region of the United States during the FY1980-81 time frame with some initial exploratory investigations into the foreign crop and regions mentioned above. Also, the sampling and aggregation software requirements in support of the U.S. corn and soybeans and the U.S. and Canada spring small grains pilot tests will be generated for implementation on the Earth Resources System (ERSYS).

Projected schedules of developmental and applied tasks according to priority, time allocation, and manpower needs are given in appendix C.

2. MULTICROP AGGREGATION PROCEDURES

2.1 THE GROUPED OPTIMAL AGGREGATION TECHNIQUE

A modified form of a stratified random sample design is the basis for the sampling strategy for these FY's of the AgRISTARS program (FY1980-87). Unfortunately, large errors in crop acreage and production estimation can be caused because of missing data. In fact, it is possible that all data for a stratum may be lost; consequently, a direct acreage estimate of the stratum may not exist. To address the problem, a grouped optimal aggregation technique has been developed by Dr. A. H. Feiveson (ref. 1). Primary objectives of the FY1980-81 sampling and aggregation strategy development are to continue to improve (by further refinement, testing, and evaluation) and to implement the weighted aggregation procedure. A brief description of this proposed procedure follows.

2.1.1 CROP ACREAGE ESTIMATION

Let $\underset{\sim}{a}$, $L \times 1$, be the vector of the unknown current year crop acreages over a group of L strata and $\underset{\sim}{h}$ a vector of corresponding historical acreages. Suppose

$$E[\underset{\sim}{a}|\underset{\sim}{h}] = \gamma \underset{\sim}{h} \tag{1}$$

$$V(\underset{\sim}{a}|\underset{\sim}{h}) = H \tag{2}$$

where γ is an unknown scalar constant of proportionality.

Suppose m of L strata has direct estimates $\hat{A}_i$ from the sample segment data, $i = 1, \cdots, m$; hence, there are $(L - m)$ strata having no direct estimates.

Let $\underset{\sim}{d}_{m\times 1}$ be the vector of direct estimates made from satellite data

$$\underset{\sim}{d} = \begin{bmatrix} \hat{A}_1 \\ \hat{A}_2 \\ \vdots \\ \hat{A}_m \end{bmatrix} \tag{3}$$

ORIGINAL PAGE IS
OF POOR QUALITY

where A_i is the true acreage of stratum i and is estimated as

$$\hat{A}_i = \frac{N_i}{n_i} \alpha_i \sum_{j=1}^{n_i} \hat{P}_{ij} \tag{4}$$

$$i = 1, 2, \cdots, m$$

and

$\hat{A}_i$ = unbiased crop acreage estimate of stratum i (i.e., $E[\hat{A}_i | \alpha_i] = \alpha_i$ with the assumptions $Var(\hat{A}_i | \alpha_i) = \sigma_i^2$ and $Cov(\hat{A}_i, \hat{A}_{i'} | \alpha_i, \alpha_{i'}) = 0$ for $i \neq i'$

N_i = number of area segments in stratum i

n_i = number of segments sampled from stratum i

α_i = agricultural area of a sampling unit (assumed constant) in the ith stratum

$\hat{P}_{ij}$ = estimate of the crop acreage proportion for sample segment j in stratum i

Suppose $\underset{\sim}{\alpha}$, $\underset{\sim}{h}$, and H be partitioned in the following way.

$$\underset{\sim}{\alpha} = \begin{bmatrix} \underset{\sim}{\alpha}_{1\,m\times 1} \\ \hline \underset{\sim}{\alpha}_{2\,(L-m)\times 1} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \\ \hline A_{m+1} \\ \vdots \\ A_L \end{bmatrix} \tag{5}$$

$$\underset{\sim}{h} = \begin{bmatrix} \underset{\sim}{h}_{1\,m\times 1} \\ \hline \underset{\sim}{h}_{2\,(L-m)\times 1} \end{bmatrix} \tag{6}$$

ORIGINAL PAGE IS
OF POOR QUALITY

$$H = \begin{bmatrix} \underset{m \times m}{H_1} & \underset{m \times (L - m)}{H_3} \\ \underset{(L - m) \times (L - m)}{H_3'} & H_4 \end{bmatrix} \tag{7}$$

Now, the current year crop acreage is supposed to be

$$A_{\cdot} = \sum_{i=1}^{L} A_i = \underset{\sim}{e}' \cdot \underset{\sim}{a} \tag{8}$$

where

$$\underset{\sim}{e} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \underset{m \times 1}{\underset{\sim}{e}_1} \\ \hdashline \underset{(L - m) \times 1}{\underset{\sim}{e}_2} \end{bmatrix} \tag{9}$$

a column vector of 1's and $A_{\cdot}$ = the total current year crop acreage in the area of interest (L strata). Then, a current year weighted crop acreage estimator has the form

$$\hat{A}_{\cdot} = \underset{\sim}{X}' \cdot \underset{\sim}{d} \tag{10}$$

where

$\underset{\sim}{X}$ = the vector obtained according to the following two conditions

$$\underset{\sim}{X}' \underset{\sim}{h}_1 = \underset{\sim}{e}' \cdot \underset{\sim}{h}$$

and (11)

$$E(\hat{A}_{\cdot} - A_{\cdot})^2 \text{ is minimum}$$

The first two moments of the estimate $\hat{A}_{\cdot}$ are

$$E[\hat{A}_{\cdot}] = \gamma h_{\cdot} = \gamma \underset{\sim}{e}' \cdot \underset{\sim}{h} \tag{12}$$

and

$$V(\hat{A}_{\cdot}) = \underset{\sim}{X}' \, \underset{\sim}{\Sigma} \, \underset{\sim}{X} \tag{13}$$

where

$$\Sigma = V(\underset{\sim}{d}|\underset{\sim}{a}) \tag{14}$$

with

$$\underset{\sim}{e}_1 = E[\underset{\sim}{d}|\underset{\sim}{a}] \tag{15}$$

and

$$\underset{\sim}{X} = (\Sigma + H_1)^{-1} (\lambda \underset{\sim}{h}_1 + H_3 \underset{\sim}{e}_2 + H_1 \underset{\sim}{e}_1) \tag{16}$$

where

$$\lambda = \left[\underset{\sim}{e}'\underset{\sim}{h}_1 - \underset{\sim}{h}_1'(\Sigma + H_1)^{-1} (H_3 \underset{\sim}{e}_2 + H_1 \underset{\sim}{e}_1)\right] \Big/ \underset{\sim}{h}_1'(\Sigma + H_1)^{-1} \underset{\sim}{h}_1 \tag{17}$$

Equations (13 and 14) are the results of the minimization, with respect to $\underset{\sim}{X}$, of the mean square error (MSE) = $E[(X'\underset{\sim}{d} - \underset{\sim}{e}'\underset{\sim}{a})^2]$. In summary, the estimate for the weighted aggregation procedure current year crop acreage is equation (11).

$$\hat{A}_. = \underset{\sim}{X}'\underset{\sim}{d}$$

with its expectation

$$E[\hat{A}_.] = \gamma h_. = \gamma \underset{\sim}{e}'\underset{\sim}{h}$$

and its variance

$$V(\hat{A}_.) = \underset{\sim}{X}' \Sigma \underset{\sim}{X}$$

where $\underset{\sim}{x}$, $\underset{\sim}{d}$, $\underset{\sim}{h}$, $\underset{\sim}{e}$, Σ, and $\underset{\sim}{X}$ are defined in equations 1, 3, 6, 9, 14, and 16, respectively.

The approach for estimating the H matrix and for grouping strata to conform with equation (11) is discussed in section 2.2 and 2.3. The states and crops involved in this investigation are given in Task 2.

2.1.2 CROP PRODUCTION ESTIMATION

Let $\underset{\sim}{\eta}$ be the vector of strata crop yields and $\hat{\underset{\sim}{\eta}}$ be its estimate. Then, the crop production, P, and its estimate, $\hat{P}$, are as follows.

$$P = \underset{\sim}{\eta}'\underset{\sim}{\alpha} \tag{18}$$

and

$$\hat{P} = \hat{\underset{\sim}{\eta}}'X'\underset{\sim}{d} \tag{19}$$

where

X is a matrix of weights obtained so that $E[\hat{P} - P]^2$ is minimum

and

$$\underset{\sim}{X}'\underset{\sim}{h}_1 = \underset{\sim}{h} \tag{20}$$

For further details on crop production estimation, refer to reference 1.

2.2 ESTIMATION OF MATRIX H

2.2.1 BACKGROUND

The grouped optimal aggregation technique requires input for the matrix H which is unknown and needs to be estimated. The current estimation approach is outlined below.

The symbols $\underset{\sim}{\alpha}$ and $\underset{\sim}{h}$ in equation (1) have been used to represent realizations of a random vector of crop acreages during 2 different years. Let a_t be the $n \times 1$ vector of crop acreages in year t ($t = 1, 2, \cdots$). Then, the proportionality model in equation (1) can be generalized to

$$a_t = \gamma_t a_{t-1} + \varepsilon_t \tag{21}$$

where the γ_t are scalar constants of proportionality and the ε_t are independent random vectors satisfying

$$E[\varepsilon_t] = 0$$

and (22)

$$E[\varepsilon_t \varepsilon_t^T] = H_t$$

which are conditional on a_{t-1}. In particular, one desires to estimate H_T where T is the current year.

In order to facilitate the estimation of H_T with the present meager historical data supply, a simplified parameteric model is used for H_t, specifically

$$H_t = \theta \mathrm{Diag}(a_{t-1,1}, \cdots, a_{t-1,n}) \tag{23}$$

where $a_{t-1,i}$ is the ith element of a_{t-1} and θ is a scalar parameter to be estimated. Using equation (23), θ can be estimated by

$$\hat{\theta} = \frac{1}{(N-1)(n-1)} \sum_{t=2}^{N} (a_t - \hat{\gamma}_t a_{t-1})^T D_{t-1}^{-1} (a_t - \hat{\gamma}_t a_{t-1}) \tag{24}$$

where

$$\hat{\gamma}_t = \frac{e^T a_t}{e^T a_{t-1}} \tag{25}$$

$$D_{t-1} = \mathrm{Diag}(a_{t-1,1}, \cdots, a_{t-1,n}) \tag{26}$$

and

$$e = (1, 1, \cdots, 1)^T \tag{27}$$

Using equation (24), the estimate of H_T is constructed by

$$\hat{H}_T = \hat{\theta}\ \mathrm{Diag}(h_{T-1,1}, \cdots, h_{T-1,n}) \tag{28}$$

2.2.2 TASKS 1 AND 2

The problem follows. If $h \equiv a_{T-1}$, then equation (28) may be a feasible way of estimating $H \equiv H_T$; however, if $h = a_{T-k}$ for some $k > 1$ (i.e., the previous $k - 1$ year of historical data are missing), there needs to be a method of adjusting the estimate in equation (28) to account for the degraded accuracy of the method in equation (1).

The tasks for the estimation of matrix H follow.

a. Task 1

The task is to develop an approach for adjusting the estimate in equation (28). One approach would be to use the model of equation (1) recursively to write $a_T(\equiv\alpha)$ in terms of $\hat{a}_{T-k}$ $(\equiv h)$; $\varepsilon_{T-k+1}, \cdots, \varepsilon_T$; and $\gamma_{T-k+1}, \cdots, \gamma_T$. Then estimate the variance of the error term of the new model.

b. Task 2

The new procedure of Task 1 will be applied to obtain equation (28) by using the 1972-76 historical crop statistics for the states of Illinois, Indiana, and Iowa; the estimation will be determined for both corn and soybeans. An approach will be determined for those areas having perhaps only 1 year of historical data available. (This will include the possibility of using an automated approach for processing appropriate current and archived sampling units to support estimation of H_T.)

This same analysis will be conducted for wheat and barley over North Dakota and Minnesota.

2.3 FURTHER TESTING OF STRATA GROUPING APPROACH

2.3.1 BACKGROUND

For the grouped optimal aggregation technique to be most effective, the strata comprising a large region must be grouped so that (a) the model given by equation (1) approximately holds within each group and (b) the elements of the matrices H and Σ are about the same magnitude. This grouping approach, which affects the final estimate $\hat{A}$, is an important element of this approach. Its testing is another FY1980-81 task. This grouping approach, in its current form, is explained in the following steps.

a. Step 1 — Within L strata of a crop region where each stratum is named an "active group", one selects an active group with the smallest (historical) acreage, say i^{th}, where the variance of its acreage estimate is V(i).

b. Step 2 — For each other group, say j^{th}, which is eligible to be joined with i (e.g., adjacent) and has the variance V(j), compute a statistic

$$\rho_j = \frac{V(i) + V(j)}{V(i,j)} \tag{29}$$

where V(i,j) is the variance of the combined "active groups" i and j. Then let

$$\rho \equiv \max_j \rho_j \tag{30}$$

and suppose group j_0 gives that $\max_j \rho_j = \rho$.

c. Step 3 — If $\rho < 1$, make the ith group inactive and choose the second smallest (historical) acreage active group to restart step 1.

d. Step 4 — If $\rho > 1$, replace groups i and j_0 by a new group which is the combination of i and j_0. Then, reactivate all remaining groups which are inactive. This process is carried out until the final grouping is obtained.

2.3.2 TASKS 3 AND 4

a. Task 3

Since APU's ostensibly consist of political subdivisions possessing homogeneous agricultural and meteorological characteristics, it has been suggested that the grouped optimal aggregation technique will produce superior aggregated estimates if the groups used for ratioing are restricted to lie within APU's. The task is to compare aggregated estimates using restricted grouping with those computed using unrestricted grouping.

b. Task 4

The grouped optimal aggregation technique will be programmed, flowcharted, and documented. This modular, structured program will replace the existing software written by A. H. Feiveson (ref. 1). The new software will be an existing subroutine and may be installed on the Earth

Observations Division Laboratory System (EODLS) as a Statistical Analysis System (SAS) procedure. The partial response model to be delivered by Texas A&M University (TAMU) will be one component of the aggregation system to be installed on the EODLS. The task is to integrate this model into the system.

2.4 SIMULATION OF AgRISTARS CROP PRODUCTION

In view of the limited availability of data, a full scale testing of the grouped optimal aggregation technique with real survey data is not feasible. Consequently, a simulation study will be conducted to evaluate the aggregation procedure for its bias and precision.

2.4.1 OBJECTIVES

The objectives of the multicrop simulation study are:

a. To test whether or not the grouped optimal aggregation procedure is unbiased and provides efficient estimates

b. To evaluate the sampling and classification error variance components

c. To study bias due to nonacquisition of segments

2.4.2 TASKS 5, 6, 7, AND 8

a. Task 5

The simulation process could involve the following four factors.

1. Crop type — They are corn, soybeans, wheat, and barley.
2. Error simulation model — This would be exercised at a stratum level and a county level.
3. Error types — Those to be considered are no errors, sampling error only, and sampling and classification errors.
4. Acquisition rates — This would be in percentages of 100, 90, 80, 70, and 60.

Factors 3 and 4 would form 15 configurations as shown in table 2-1.

TABLE 2-1.- CONFIGURATIONS FOR FACTORS 3 AND 4

Error type	Acquisition rate, percent				
	100	90	80	70	60
No error					
Sampling error					
Sampling and classification error					

Factors 4, 3, and 2 lead to 30 configurations for each crop type. In fact, the input to the simulation study would be obtained from the 1978 corn and soybean acreage and yield data for Illinois, Indiana, and Iowa and from the wheat and barley for North Dakota and Minnesota.

Results for the simulation study will be based on 100 aggregation runs in each case. For this study, results include (a) average acreage and production estimates for each state as well as for the entire area of interest, (b) estimates of the bias and variance of an acreage or production estimate (variance estimate for the MSE of an estimate), and (c) the accuracy goal achieved.

The Monte Carlo technique will be used to generate the inputs for obtaining corn/soybean and wheat/barley acreage and production estimates. Based on these aggregation runs, the following will be determined: (a) the bias and MSE of an estimate for each state as well as for all three or five states, (b) contributions of the sampling and classification error components, (c) the effect of nonacquisition of segments on the estimate, and (d) other such factors.

b. Task 6

An integrated simulation system will be developed and utilized by combining the following pieces of software: segment crop proportion estimation simulation, a nonresponse simulation, and a yield estimate simulation.

c. Task 7

New simulation software will be developed to facilitate the simulation of segment-level proportion estimates, yield numbers, and nonresponse because of cloud cover. The software will be three independent subroutines and may be used in any combination to fit many applications.

d. Task 8

Research will be conducted for the purpose of developing more realistic nonresponse models (to simulate cloud cover, etc.) than are now available. Development of these models will help to answer the question: Is significant bias from loss of data due to cloud cover introduced in aggregated acreage and production estimates?

2.4.3 SIMULATION MODEL FOR SEGMENT CROP PROPORTION ESTIMATION

The two models considered to simulate segment crop proportion estimates are simulation models I and II. The simulation model I is

$$\hat{P}_{ij} = P_i + \delta_{ij} + \varepsilon_{ij} \tag{31}$$

where

$\hat{P}_{ij}$ = crop proportion estimate for j^{th} segment in stratum i

P_i = actual crop proportion for stratum i

δ_{ij} = sampling error component

ε_{ij} = classification error component

with the following assumptions

$$\begin{aligned} E[\delta_{ij}] &= 0 \quad , \quad V(\delta_{ij}) = \sigma^2_{ie} \\ E[\varepsilon_{ij}] &= 0 \quad , \quad V(\varepsilon_{ij}) = \sigma^2_{ic} \end{aligned} \tag{32}$$

and

$$Cov(\delta_{ij}, \varepsilon_{ij}) = \rho_i \sigma_{ie} \sigma_{ic}$$

Here, ρ_i is the correlation coefficient of δ_{ij} and ε_{ij}; therefore,

$$E[\hat{P}_{ij}] = P_i \tag{33}$$

and

$$V(\hat{P}_{ij}) = \sigma_{ie}^2 + \sigma_{ic}^2 + 2\rho_i \sigma_{ie} \sigma_{ic} \equiv \tau_i^2$$

The simulation model II is

$$\hat{P}_{ikj} = P_{ik} + \delta_{ikj} + \varepsilon_{ij} \tag{34}$$

where

$\hat{P}_{ikj}$ = crop proportion estimate for segment j in county k of stratum i

$k = 1, 2, \cdots, K$

P_{ik} = actual crop proportion for county k of stratum i

δ_{ikj} = sampling error component at the county level

ε_{ij} = classification error component at the stratum level

with the following assumptions

$$\begin{aligned} E[\delta_{ikj}] &= 0 \quad , \quad V(\delta_{ikj}) = \sigma_{i \cdot e}^2 \\ E[\varepsilon_{ij}] &= 0 \quad , \quad V(\varepsilon_{ij}) = \sigma_{i \cdot c}^2 \end{aligned} \tag{35}$$

and

$$Cov(\delta_{ikj}, \varepsilon_{ij}) = \rho_i \sigma_{i \cdot e} \sigma_{i \cdot c}$$

Here, ρ_i is the correlation coefficient of δ_{ikj} and c_{ij}; therefore,

$$\sigma^2_{i\cdot e} = |\sigma^2_{ie} - \sigma^2_{ib}|$$

$$\sigma^2_{i\cdot c} = \sigma^2_{ic} \tag{36}$$

$$\sigma^2_{ib} = \frac{1}{K-1} \sum_{k=1}^{K} (P_{ik} - P_i)^2$$

so

$$E[\hat{P}_{ikj}] = P_{ik}$$

and

$$V(\hat{P}_{ikj}) \equiv \tau^2_{ik} = \sigma^2_{i\cdot e} + \sigma^2_{i\cdot c} + 2\rho_i \sigma_{i\cdot e} \sigma_{i\cdot c} \tag{37}$$

2.4.4 DATA INPUT AND OUTPUT

The simulation input will be based on the 1978 historical crop acreages and FY78 segment data as follows:

a. The 1978 crop acreage estimates by the U.S. Department of Agriculture (USDA)/Economics and Statistics Service (ESS) are to be used for determination of the P_i or P_{ik} (i.e., for the simulation model to generate the crop proportion for the sample segments).

b. The initial within-stratum variance estimates are to be used for determination of sampling error σ^2_{ie} or $\sigma^2_{i\cdot e}$.

c. Segment crop proportion estimates from the 1978 crop year will be compared to ground-truth segment proportions, and the estimation errors will be used to compute $\sigma^2_{i\cdot c}$ and $\sigma^2_{i\cdot e}$.

d. Computations of the above parameters will be carried out at the refined stratum level.

e. The historical 1972-76 crop statistics will be used for computation of matrix H. (This procedure is repeatable in a foreign country provided

that at least 3 years of historical crop acreage statistics are available; where this is not the case, considerations as indicated in Task 2 will be investigated.)

f. The strata and higher level aggregated estimates of crop acreage and production will be obtained using the simulated inputs. The associated variance estimates from the grouped optimal aggregation technique will be computed. The bias and repeatability of the proposed estimates will be determined by the Monte Carlo technique.

3. SAMPLING FRAME DEVELOPMENT

3.1 OBJECTIVES AND APPROACH

The objectives of the sampling frame development are (a) to integrate the approach for creating sampling frames as developed by the USDA/ESS into the overall sampling and aggregation methodology and (b) to develop and implement an automated procedure in support of creating crop-specific strata.

The approach consists of (a) the implementation of the USDA/ESCS software and sampling frame data base into the EODLS at the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center (NASA/JSC) and (b) the implementation of the USDA/ESS-developed procedure for the overlaying of varying sized sampling units as well as the module for selection of the sampling units.

In addition, an automated procedure for making crop proportion estimates at the sampling unit level will be utilized in developing an automated approach for creating crop-specific strata.

3.2 SAMPLING UNIT SIZE STUDY

3.2.1 BACKGROUND

The LACIE sampling unit size of 5 by 6 nautical miles was considered to be large enough for the Classification and Mensuration Subsystem (CAMS) analysts to obtain wheat acreage estimates and small enough not to tax the computer and manpower resources. However, it is time to reexamine the sampling unit size for the purpose of obtaining sampling unit sizes of the stratum level that are more adequately suited to meeting project requirements. Other significant facts which influence the considered change in the sampling unit size are:

a. The analysts are more experienced.

b. The crops of interest in AgRISTARS are not only wheat or small grains but also soybeans, corn, rice, cotton, sorghum, and sunflowers.

c. The areas of interest include not only Canada, U.S.S.R., and the USGP which is the major wheat and barley producing area in the U.S. but also Australia, India, Argentina, and Brazil.

d. The sampling efficiency needs to be improved by reducing the sampling unit size, if necessary, since a proper size for a sampling unit can result in more homogeneous sampling units and, hence, may lead to a significantly smaller number of sampling units requiring processing.

Criteria to be considered in deciding upon a sampling unit size are:

1. The accuracy of labeling and classification of segments should be considered as a function of size. There is a need to examine the extent to which a smaller segment size leads to larger labeling and classification errors; it is well known that the interpretation of the Landsat imagery, the estimation of crop signatures, and the registration procedure require an adequate segment size.

2. The overall processing cost with smaller segment sizes is expected to increase in comparison to that of larger segment sizes because a larger number of segments will need to be processed to reduce the sampling error.

3.2.2 TASKS 9 AND 10

A cost analysis study is being conducted by the UCB to investigate the effect of segment size on the analyst labeling accuracy and on the overall segment processing cost. Subsequently, another sampling unit size study is envisioned for Robeson County, North Carolina, where wall-to-wall ground data on crop types are expected to be available from the USDA.

a. Task 9

This task will integrate the results from the two studies (UCB and Robeson County) into a procedure for determining the optimal sampling unit size for multicrop estimation.

The purpose of this task is to develop and to evaluate the approach for the integration of the two studies. An evaluation will also be made of the cost analysis study conducted at UCB. The necessary inputs will be provided on the experiment design for the Robeson County study (conducted by USDA/ESS) of the sampling unit size.

b. Task 10

This task requires the designing and implementing of a sampling frame to be delivered by USDA/ESS.

The USDA/ESS is required to develop a semiautomatic digitization and enumeration software in support of sampling frame development for transfer to NASA/JSC. Upon receipt of the sampling frame software from USDA/ESS, an in-house version will be designed, implemented, and tested. The design will provide the capability to permit complete automation of the task of creating and overlaying a sampling unit grid allowing for differing sampling unit sizes, plus the actual selection of the sampling units.

It is envisioned that the implementation of the sampling frame will have an easy-to-use interface with the SAS. A capability for "collapsing" or "splitting" the elemental units which make up the sampling frame will be developed, tested, and implemented (probably as a SAS procedure or, at least, be interfaced with SAS).

4. MULTICROP SAMPLE ALLOCATION

4.1 OBJECTIVE

The objective of the multicrop (stratified) sample allocation task is to determine the sample size n_h ($h = 1, \cdots, L$; L being the number of strata) so that the variances of the individual crop production estimates satisfy some specified precision tolerances with the cost minimized.

4.2 INITIAL WITHIN-STRATUM VARIANCE ESTIMATES

Before determining n_h, one needs to estimate the within-stratum crop variances. A procedure for estimating the within-stratum variances (ref. 2) has been developed; however, it needs to be tested further for different crops and different areas of interest.

4.2.1 TASK 11

This task requires further testing of the proposed within-stratum variance estimation procedure for corn and soybeans in Iowa, Indiana, and Illinois, for the data set in Robeson County, North Carolina and for wheat and barley in North Dakota and Minnesota.

Stratum variances will be obtained from the crop proportion estimates for the segments. These variance estimates will then be compared with those determined by using the proposed method. The performance of the method will be judged based on this comparison and a statistical test for equality.

4.2.2 TASK 12

Task 10 deals with a machine classification procedure. This procedure shows potential for improving initial variance estimation; therefore, it needs to be tested by evaluating its classification accuracy and investigating the feasibility of its implementation beginning with the automatic segment strip-off from the high density tapes to output of within-stratum variances. It may not be feasible to perform these tasks this year.

4.3 ALLOCATION TO SUPPORT EXPLORATORY STUDIES

Sample allocation procedures will be determined for corn/soybeans in Brazil and Argentina and for spring wheat/barley in the U.S.S.R. (These tasks will be performed as applied tasks.)

5. SEGMENT SELECTION AND LOCATION

With the sample allocations completed for the FY1981 exploratory studies, the tasks of selecting sample segments and locating them on the sampling frame will be undertaken. Sampling-frame software to be developed by USDA/ESS is not yet available to select and locate the sample segments in (a) Brazil for the corn/soybean exploratory study, (b) U.S.S.R. for the spring wheat/barley exploratory study as planned for FY81, and (c) Argentina for the corn/soybeans exploratory study. Therefore, these selections will be made manually.

6. STUDY OF ALTERNATIVE SAMPLING STRATEGIES

Investigation of some alternative sampling strategies is warranted. In particular, the most promising of these investigations are the multiyear model and the full-frame sampling.

The multiyear models for crop acreage estimation are being developed, tested, and evaluated at TAMU and the Environmental Research Institute of Michigan (ERIM), whereas consideration for development of a full-frame sampling strategy approach has received attention at the Laboratory for Applications of Remote Sensing (LARS), UCB, and ERIM.

6.1 HARTLEY MULTIYEAR MODELS

6.1.1 HARTLEY MIXED ANALYSIS OF VARIANCE MODEL

6.1.1.1 Objectives and Approaches

The objective is to improve the conventional current year crop acreage proportion estimates. The multiyear models will utilize not only the current year sample segments but also those of many successive years in order (a) to explore the properties of consistency in segment crop acreages from year to year and (b) to deal with the systematic differences between early-season and late season estimates.

A mixed analysis of the variance model will be considered wherein the current year direct estimates (from Landsat data) and the previous year estimates will be investigated for providing a more precise current year estimate.

6.1.1.2 Description of Hartley Mixed Analysis of Variance Model

The basic model (ref. 3) is

$$y(P_{ts\ell}) = \alpha_t + b_s + \delta_\ell + e_{ts\ell} \tag{38}$$

where

$s = 1, 2, \cdots$, segments studied

$\ell = 1, 2, 3$ for early season, midseason, and late season

$t = 1, 2, \cdots$, years under study

$P_{ts\ell}$ = the analyst-interpreter (AI) estimate of the wheat proportion of segment s of stratum h in the crop calendar period ℓ of year t, say $\tilde{P}_{ts\ell}$

$y(P_{ts\ell})$ = a mathematical variate transform of $P_{ts\ell}$ (say its logarithm or logit transform); $y(P_{ts\ell})$ will be abbreviated as $y_{ts\ell}$

α_t = an effect constant of year t (representing for example, the favorable or unfavorable economic outlook for wheat in year t at planting time or, at least, the average transform)

b_s = an effect variable for segment s consistent through all years of the data bank (e.g., the soil type effect); it assumes $b_s \sim N(0,\sigma_b^2)$

δ_ℓ = the consistent (systematic) difference between the late-season estimate (which corresponds to $\delta_3 = 0$) and the estimate made at crop calendar period ℓ ($\ell = 1$) for early season, $\ell = 2$ for midseason)

$e_{ts\ell}$ = the aggregate of sampling and classification errors in the transformed AgRISTARS records; it assumed $e_{ts\ell} \sim N(0,\sigma_e^2)$

The Aitken weighted least squares estimator of $\underset{\sim}{\theta} = (\alpha_1, \cdots, \alpha_t, \cdots, \alpha_T, \cdots, \delta_1, \delta_2, \delta_3)'$ is

$$\hat{\underset{\sim}{\theta}} = (X'H^{-1}X)^{-1} X'H^{-1}\underset{\sim}{y} \tag{39}$$

where

$H = I + \gamma UU'$

$\gamma = \sigma_b^2 | \sigma_e^2$

X and U are design matrices and $\underset{\sim}{y}$ is vector of $y_{ts\ell}$'s

(Note: Each $y_{ts\ell}$ will be utilized only for one value of ℓ, and this will be the crop-calendar period ℓ for which an acquisition was utilized that was not available at the previous forecast $\ell - 1$.)

Equation model (38) can be utilized under three forms of mathematical variate transforms $y(P_{ts\ell})$. They are:

a. The Identity Transformation

For the identity transformation $y(P_{ts\ell}) \equiv P_{ts\ell}$, the model will be

$$P_{ts\ell} = P_t + b_s + \delta_\ell + e_{ts\ell} \tag{40}$$

where

$$\alpha_T = P_T \quad ; \quad \text{that is: } \hat{P}_T = \hat{\alpha}_T$$

This model will help to analyze the P_t records directly; hence, aggregations of wheat acreage estimates $a_{\hat{P}_t}$ are needed for strata, strata combinations, and countries. This model will give unbiased (or near unbiased) estimates $\hat{P}_t$ if the additive equation (38) is satisfied for the $P_{ts\ell}$.

b. The Logarithmic Transformation

For the logarithmic transformation $y_{ts\ell} \equiv \log P_{ts\ell}$, the estimate of the crop proportion in crop calendar period $\ell = L$ and current year $t = T$ is

$$\hat{P}_{TsL} = C \operatorname{Exp}\{\hat{\alpha}_T + \hat{\delta}_L\} \tag{41}$$

where

$$C = \operatorname{Exp}\left\{\frac{1}{2}\left(\sigma_b^2 + \frac{\sigma^2}{W_{ts\ell}}\right) - \frac{1}{2}\operatorname{Var}(\hat{\alpha}_t + \hat{\delta}_\ell)\right\}$$

$$W_{ts\ell} = \tilde{\pi}_{ts\ell} | (1 - \tilde{\pi}_{ts\ell}) \tag{42}$$

$$\tilde{\pi}_{ts\ell} = \exp(\tilde{\alpha}_t + \tilde{b}_s + \tilde{\delta}_\ell) \tag{43}$$

$\tilde{\alpha}_t$, $\tilde{b}_s$, and $\tilde{\delta}_\ell$ are unweighted least squares estimates of equation (38) with all factors being fixed.

c. The Logistic Transformation

For the logistic transformation

$$y_{ts\ell} = \frac{1}{2} \log \frac{P_{ts\ell}}{1 - P_{ts\ell}} \tag{44}$$

$$\hat{P}_{TsL} = C'\exp[2(\hat{\alpha}_T + \hat{\delta}_L)]/\{1 + \exp[2(\hat{\alpha}_T + \hat{\delta}_L)]\} \tag{45}$$

where

$$C' = \frac{F\left(\mu_{ts1},\ \sigma_b^2 + \frac{\sigma^2}{W_{ts1}}\right)}{F\left[\mu_{ts\ell},\ V(\hat{y}_{ts\ell})\right]} \tag{46}$$

with

$$V(\hat{y}_{ts\ell}) = V(\hat{\alpha}_t) + V(\hat{\delta}_t) + 2\ \mathrm{cov}(\hat{\alpha}_t,\ \hat{\delta}_\ell) \tag{47}$$

and

$$E[\hat{P}_{ts\ell}] = E\{\exp(2\hat{y}_{ts\ell})\big/[1 + \exp(2\hat{y}_{ts\ell})]\} = F[\mu_{ts\ell}|V(\hat{y}_{ts\ell})] \tag{48}$$

$$E[\hat{y}_{ts\ell}] = \mu_{ts\ell} \tag{49}$$

Note: The efficiency is defined by the variance reduction ratio

$$R = \frac{\mathrm{Var}(\hat{\alpha}_T)}{\mathrm{Var}(\bar{y}_T)} \tag{50}$$

where $\bar{y}_T$ is the corresponding estimate of $\hat{\alpha}_T$ in the current year T based only on current year acquisitions and $\hat{\alpha}_T$ is the estimate from the proposed multiyear model.

6.1.1.3 Tasks 13 and 14

The following tasks will be implemented using previous segment estimates for wheat and small grains over a 4-year period for North Dakota.

a. Task 13

The requirement is to examine the feasibility of utilizing the estimate given by the multiyear models in the weighted aggregation procedure by replacing the $\hat{P}_{ij}$'s in equation (5) with $\hat{P}_{ts\ell}$'s from the multiyear model. The combined procedure will be called the Multiyear Weighted Aggregation Procedure (MYWAP).

b. Task 14

The requirement is to investigate the utilities of the multiyear models for (a) crop estimation with missing data because of cloud cover or poor quality imageries and (b) early season, midseason, and late season estimation by estimating R in each case. (R is the efficiency parameter defined above.)

6.1.2 ROTATION SAMPLE DESIGNS

The original (and simplest) model of the multiyear models is the rotation sample design (ref. 4).

6.1.2.1 Objective and Approach

The objective of this rotation sample design is to provide current year crop acreage estimates which are more efficient than the conventional current year crop acreage estimates by utilizing the knowledge that the variation of the crop acreage of a particular segment from year to year is usually less than the variation of the crop acreage of different segments within a particular year.

The approach follows. The current year direct segment average estimates from the satellite data and the previous year estimated acreages are utilized to provide a more precise current year estimate.

6.1.2.2 Description of Rotation Sample Designs

The basic model (ref. 5) is

$$A_{ts} = \alpha_t + b_s + e_{ts} \tag{51}$$

for $t = 1, 2, 3, 4$ years (the current year is $T \equiv 4$)

$s = 1, 2, 3, 4$ (s is the segment or sampling unit number)

where

α_t = average true wheat acreage per segment in year t (α_t's are fixed year constants.)

b_s = true segment variables applicable to all years with the assumption $b_s \sim N(0,\sigma_b^2)$

e_{ts} = composite segment error variable of segment s in year t with the assumption $e_{ts} \sim N(0,\sigma_e^2)$

A_{ts} = current year directly estimated (from satellite data) or previously estimated wheat acreage of segment s of stratum h in year t

This segment s is observed in stratum h. Within the population of stratum h, there are allocated $n_h = S$ segments for each year t with a condition of $S = 1$, 2, 3, or 4 segments; i.e., the model is applied to nominal Group I crop regions. The allocation of sample segments is followed by a rotation pattern. The authors in reference 5 investigated many different rotation patterns in using equation (51). The optimal pattern consists of $S = 2$ segments per year which are retained 1 segments (past year to current year). One observed segment will return to the sample after a 2-year absence. This rotation pattern is presented in figure 6-1 which utilizes four segments in 4 years. Figure 6-2 shows the pattern ($S = 2$, $r = 2$).

Year Number

Segment Number (s) \ t	1	2	3	4
1	X			X
2	X	X		
3		X	X	
4			X	X

Figure 6-1.- (S = 2, r = 3) rotation pattern [2 retained 1; T = 4].

Year Number

Segment Number (s) \ t	1	2	3
1	X		X
2	X	X	
3		X	X

Figure 6-2.- (S = 2, r = 2) rotation pattern.

ORIGINAL PAGE IS
OF POOR QUALITY

In matrix form, the basic equation (51) of the optimal rotation pattern can be written as

$$\underset{\sim}{a} = X\underset{\sim}{\alpha} + U\underset{\sim}{b} + I\underset{\sim}{e} \tag{52}$$

where

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{53}$$

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{54}$$

and

$\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$

α_4 is to be estimated as $\hat{\alpha}_4$

$\underset{\sim}{b} = (b_1, b_2, b_3, b_4)' \sim NI_4(\underset{\sim}{0}, I\sigma_b^2)$

$\underset{\sim}{e} = (e_{11}, e_{41}, e_{12}, e_{22}, e_{23}, e_{33}, e_{34}, e_{44}) \sim NI_8(\underset{\sim}{0}, I\sigma_e^2)$

$\underset{\sim}{a} = (A_{11}, A_{41}, A_{12}, A_{22}, A_{23}, A_{33}, A_{34}, A_{44})$

Then, it is proved that

$$\hat{\underset{\sim}{\alpha}} = (X'H^{-1}X)^{-1}\, X'H^{-1}\underset{\sim}{a} \tag{55}$$

which is the best linear unbiased estimator (BLUE) of $\underset{\sim}{\alpha}$

where

$$H = I + \hat{\gamma} U U' \tag{56}$$

$$\hat{\gamma} = \hat{\sigma}_b^2 / \hat{\sigma}_e^2 \tag{57}$$

with $\hat{\alpha}_4$ being deduced from $\hat{\underset{\sim}{\alpha}}$, and the estimate of the variance of $\hat{\alpha}_4$ being

$$Var(\hat{\alpha}_4) = \frac{1 + 2\hat{\gamma}}{8} \left(1 + \frac{1}{1 + 2\hat{\gamma}} - \frac{2}{1 + \hat{\gamma}}\right) \hat{\sigma}_e^2 \tag{58}$$

therefore, the estimate of the stratum h crop acreage at the year T ≡ 4 will be

$$\hat{A}_{4h} = N_h \hat{\alpha}_4 \tag{59}$$

with its variance estimate

$$Var(\hat{A}_{4h}) = N_h^2 \, Var(\hat{\alpha}_4) \tag{60}$$

where N_h is the population stratum h size and the current year estimate of the total crop production in the crop region of interest will be

$$Prod_4 = \sum_{h=1}^{L} (\hat{A}_{4h} \hat{\bar{Y}}_{4h}) \tag{61}$$

where $\hat{\bar{Y}}_{4h}$ is the current year stratum h average wheat yield estimate.

6.1.2.3 Tasks 15 and 16

These tasks will be implemented using previous estimates for wheat and small grains in North Dakota for a 4-year period.

a. Task 15

The requirement is to examine the efficiency of the rotation sample design on the wheat data available in North Dakota.

b. Task 16

The requirement is to develop an aggregation procedure tailored to the rotation sample designs. The combined sampling and aggregation procedures will be called "Rotation-Prestratified Sampling and Aggregation."

6.2 FULL-FRAME SAMPLING STRATEGY

6.2.1 OBJECTIVES

The objective of a full-frame sampling strategy is to provide an alternate, yet more efficient, sampling and estimation approach based on a sampling unit size which is as small as one pixel. LARS, ERIM, and UCB will investigate the feasibility of a full-frame sampling strategy.

6.2.2 TASKS 17 AND 18

6.2.2.1 Task 17

This task will consist of the integration of the recommendations resulting from the investigations of a full-frame sampling strategy into the overall sampling and aggregation methodology. A primary part of task 17 is the implementation of the adapted procedures on ERSYS.

6.2.2.2 Task 18

The requirement is to develop an alternate full-frame sample design.

7. SOFTWARE CONVERSION

Currently, most of the sampling and aggregation software and data bases developed in the past reside on the Programmed Data Processor, Model 11/45 (PDP-11/45) computer system. Three tasks will be implemented to consolidate the sampling and aggregation capabilities on the LARS system at Purdue University and on EODLS Analysis Subsystem (EAS) when it becomes available. Results of the three tasks are to be maintained on the LARS system until the in-house EAS becomes available. Then similar conversions will be required for implementing all software and data bases in-house.

7.1 TASK 19

The requirement is to transfer the LACIE and TY sampling and aggregation software from the PDP 11/45 computer system to the LARS system with control configurations specified and procedure requirements documented.

7.2 TASK 20

The requirement is to transfer from the PDP 11/45 computer system to the LARS system all data bases consisting of agricultural statistics and CAMS segment estimates for the supporting USGP and foreign countries. A record of all data files thus created on the LARS system will be maintained.

7.3 TASK 21

The requirement is to conduct verification tests for the transferred softwares.

8. ASSEMBLAGE OF DATA BASES

8.1 TASK 22

The requirement is to prepare a data base containing CAMS segment estimates for early season, midseason, and late season as well as the blind-site true crop proportions from LACIE Phases II and III.

8.2 TASK 23

The requirement is to assemble and file all Crop Assessment Subsystem (CAS) reports for LACIE Phases I, II, and III and TY.

9. DOCUMENTATION

9.1 TASK 24

The procedure requirements for the sampling and aggregation software will be specified and documented.

9.2 TASK 25

The preparation of this implementation plan is Task 25.

10. REFERENCES

1. Feiveson, A. H.: Weighted Ratio Estimation for AgRISTARS. NASA-JSC, (to be published).

2. Chhikara, R. S. and Perry, C. R.: Estimation of Within-Stratum Variance for Sample Allocation. JSC-16343, LEMSCO-14067, July 1980.

3. Hartley, H. O.: A Survey of Multiyear Estimation Procedures. Duke University (Durham, N. C.), Report DS1, August 1980.

4. Hartley, H. O. and Lycthuan-Lee, T. G.: Rotation Sample Designs for Estimation of Crop Acreages in Satellite Agriculture Surveys. Presented at the American Statistical Association annual meeting, Houston, Texas, August 1980.

APPENDIX A

DEVELOPMENTAL TASKS: SUMMARIES OF BACKGROUND, ACTIONS, TIME, AND MANPOWER

APPENDIX A

DEVELOPMENTAL TASKS: SUMMARIES OF BACKGROUND, ACTIONS, TIME, AND MANPOWER

Appendix A consists of the developmental tasks in outline form. These tasks were generated for this Implementation Plan.

The equations in this appendix are taken from the text of this document (LEMSCO-15168) and retain the equation number given in the text.

PRECEDING PAGE BLANK NOT FILMED

TASKS 1 AND 2: ESTIMATION OF MATRIX H

BACKGROUND: The crop acreages at the year t which are to be estimated should be in the model

$$a_t = \gamma_t a_{t-1} + \varepsilon_t \ , \quad t = 1,2,\cdots,T \tag{21}$$

where

$$E\left[\varepsilon_t | a_{t-1}\right] = 0 \tag{22}$$

$$E\left[\varepsilon_t \varepsilon_t^T | a_{t-1}\right] = H_t$$

$$\hat{H}_t = \hat{\theta}\ \mathrm{Diag}\left(h_{T-1,1}, \cdots, h_{T-1,n}\right) \tag{28}$$

ACTIONS IN TASK 1: Develop a procedure to adjust the estimate in equation (28)

TIME: 13 weeks

MANPOWER: 135 man-weeks

ACTIONS IN TASK 2: Actually estimate $\hat{H}_T$ using the procedure in Task 1 and the 1972-76 historical crop statistics for the states of Illinois, Indiana, and Iowa. This estimation will be done based on corn and soybean data.

TIME: 4 weeks

MANPOWER: 45 man-weeks (If the soybean data base is in raw form, two (2) more man-weeks will be needed to prepare soybean data.)

TASKS 3 AND 4:	TESTING OF STRATA GROUPING APPROACH
BACKGROUND:	Since APU's ostensibly consist of political subdivisions possessing homogeneous agricultural and meteorological characteristics, it has been suggested that the weighted aggregation procedure will produce superior aggregated estimates if the groups used for ratioing are restricted to lie within APU's.
ACTIONS IN TASK 3:	Compare aggregated estimates using restricted grouping with those computed using unrestricted grouping.
TIME:	16 weeks
MANPOWER:	8 man-weeks
ACTIONS IN TASK 4:	The grouped optimal aggregation technology will be programmed, flowcharted, and documented. This modular, structured program will replace the existing software as written by A. H. Feiveson. The new software will be an existing subroutine and may be installed on the EODLS as a SAS procedure.
TIME:	April 1, 1981 (date due)
MANPOWER:	Contact person is George Clouette.

TASKS 5, 6, 7, AND 8

SIMULATION OF AgRISTARS CROP PRODUCTION

BACKGROUND:

Because of the limited data available, a full-scale testing of the grouped optimal aggregation technique is not feasible. Therefore, some simulation processes are considered for generating the estimates of the segment proportion to be used in testing.

For P_i or P_{ik} which represent the actual crop proportion for stratum i or for county k of stratum i, the crop proportion estimate (simulated) for j^{th} segment in (county k) stratum i will be $\hat{P}_{ij}$ or $\hat{P}_{ikj}$.
It will be $\hat{P}_{ij}$ where

$$E[\hat{P}_{ij}] = P_i$$

and (33)

$$V(\hat{P}_{ij}) = \sigma^2_{ie} + \sigma^2_{ic} + 2\rho_i \sigma_{ie} \sigma_{ic}$$

and

σ^2_{ie} is the variance of sampling error component δ_{ij}
σ^2_{ic} is the variance of classification error component ε_{ij}
and ρ_i is the correlation coefficient of δ_{ij} and ε_{ij}
It will be $\hat{P}_{ikj}$ where

$$E[\hat{P}_{ikj}] = P_{ik}$$

$$V(\hat{P}_{ikj}) = \sigma^2_{i\cdot e} + \sigma^2_{i\cdot c} + 2\rho_i \sigma_{i\cdot e} \sigma_{i\cdot c} \qquad (37)$$

and $\sigma^2_{i.c} \equiv \sigma^2_{ic}$

$$\sigma^2_{i\cdot e} = |\sigma^2_{ie} - \sigma^2_{ib}|$$

$$\sigma^2_{ib} = \frac{1}{k-1} \sum_{k=1}^{K} (P_{ik} - P_i)^2$$

ACTIONS IN TASK 5: Use 1978 crop acreages estimated by USDA/ESS to determine the P_i's or P_{ik}'s. Use the initial within-stratum variance estimates to determine the sampling error σ^2_{ie} or $\sigma^2_{i \cdot e}$. Segment crop proportion estimates will be compared to ground-truth segment proportions and the estimation errors will be used to compute σ^2_{ic} and $\sigma^2_{i \cdot c}$.

NOTE: Computations of the parameters will be calculated at the refined stratum level. The input to the simulation study will be obtained from the 1978 corn and soybean acreage and yield for Illinois, Indiana, and Iowa.

The strata and higher level aggregated estimates of crop acreage and production will be obtained using the simulated inputs. The associated variance estimates from the grouped optimal aggregation technique will be computed. The bias and repeatability of the proposed estimates will be determined by the Monte Carlo process.

TIME: Current year

MANPOWER: 35 man-weeks

ACTIONS IN TASK 6: Develop and utilize an integrated simulation system by combining the following pieces of software.

a. Segment crop proportion estimation simulation
b. Yield estimate simulation.
c. Nonresponse simulation

TIME: Current year

MANPOWER: 35 man-weeks

ACTIONS IN TASK 7:	Develop new simulation software to facilitate the simulation of segment-level proportion estimates, yield numbers, and nonresponse because of cloud cover. The software will be three independent subroutines and may be used in any combination to fit many applications.
TIME:	April 1, 1981 (due date)
MANPOWER:	Contact person is George Clouette.
ACTIONS IN TASK 8	Develop more realistic nonresponse (cloud cover, etc.) models than are now available. (Development of these models will help to answer the question: Is significant bias from loss of data due to cloud cover introduced in aggregated acreage and production estimates?)
TIME:	32 weeks
MANPOWER:	28 man-weeks

TASKS 9 AND 10: DESIGN AND IMPLEMENTATION OF SAMPLING FRAME

BACKGROUND: An optimal sampling unit size different from 5 by 6 nautical miles needs to be determined. The University of California at Berkeley is investigating the effect of segment size on the analyst labeling accuracy and on the overall segment processing cost. The USDA/ESS is studying another sampling unit size which is envisioned for use in Robeson County, North Carolina.

ACTIONS IN TASK 9: Integrate the two studies (University of California at Berkeley study and the USDA study in Robeson County) and help in determining an optimal sampling unit size for the multicrop estimation for the areas of application. Then evaluate the cost analysis study of the University of California at Berkeley and provide the necessary input on the experiment design for the Robeson County study of sampling unit size.

TIME: 26 weeks

MANPOWER: 0.5 man-equivalents for 6 months for coordination and integration

ACTIONS IN TASK 10: The USDA/ESS is required to develop a semiautomatic digitization and enumeration software in support of sampling frame development for transfer to NASA/JSC. Upon receipt of the sampling frame software from USDA/ESCS, an in-house system will be designed, implemented, and tested. The design will permit complete automation; e.g., the task of creating and overlaying a sampling unit grid allowing for differing sampling unit sizes plus the actual selection of the sampling units. The implementation of the sampling frame will have an easy-to-use interface with the Statistical Analysis System (SAS). A capability for collapsing or splitting the element units making up the sampling frame will be developed, tested,

PRECEDING PAGE BLANK NOT FILMED

and implemented (probably as a SAS procedure or, at least, be interfaced with SAS).

TIME:

MANPOWER: If USDA/ESS delivers as a package and not in bits and pieces, then the following will be required.

a. 13 man-weeks for software requirements
b. Programming needs will be:
   Design specifications — 8 man-weeks
   Coding — 2 man-equivalents for 9 weeks
   Documentation — 9 man-weeks
c. Acceptance: 2 man-weeks

TASKS 11 AND 12: MULTICROP SAMPLE ALLOCATION

BACKGROUND: The objective is to determine the sample size n_h of each stratum h, h = 1, 2, •••, L, so that the variances of the individual crop production estimates satisfy some specified premium tolerances with cost minimized. Before determining n_h, the within-stratum crop variances needs to be estimated.

ACTIONS IN TASK 11: Initial within-stratum variance estimates — A procedure for estimating the within-stratum variances was proposed by Chhikara and Perry (ref. 2). Testing of the proposed within-stratum-variance estimation procedure will be implemented for corn and soybeans in Iowa, Indiana, and Illinois. The exploratory study (1979) segments will be used for this testing. Strata variances will be obtained from the crop proportion estimates for the segments. These variance estimates will then be compared with those which were determined by using the proposed method. The performance of the method will be judged based on this comparison and the statistical test developed for testing their equality.

TIME: 6 weeks

MANPOWER: 6 man-weeks

ACTIONS IN TASK 12: The machine classification procedure will be tested by evaluating its accuracy and by investigating the feasibility of its implementation, beginning with the automatic segment strip-off from the high density tapes to output of the within-stratum variances.

TIME: 8 weeks

MANPOWER: 12 man-weeks

TASKS 13 AND 14: HARTLEY MULTIYEAR (LOG) MODEL

BACKGROUND: For the three forms of the base model

$$y(P_{ts\ell}) = \alpha_t + b_s + \delta_\ell + e_{ts\ell} \tag{38}$$

the logarithmic model is favored to give the estimate of the crop proportion for segments in crop calendar period $\ell = L$ at current year $t = T$ which is

$$P_{TsL} = C \operatorname{Exp}\{\alpha_T + \delta_L\} \tag{41}$$

where

$$C = \operatorname{Exp} \frac{1}{2} \left(\sigma_b^2 + \frac{\sigma^2}{W_{ts\ell}}\right) - \frac{1}{2} \operatorname{Var}(\alpha_t + \delta_\ell) \tag{42}$$

and other notation definitions given in section 6.1.1.2.

ACTIONS IN TASK 13: Examine the feasibility of utilizing the estimate given by the multiyear models in the weighted aggregation procedure by replacing $\hat{P}_{ij}$'s in equation (5) of the weighted aggregation procedure with $\hat{P}_{ts\ell}$'s from the multiyear model and by developing an aggregation using the proportion estimates for wheat and small grain over 4 years for North Dakota (adaption of weighted aggretation procedure to MYWAP).

TIME: 13 weeks

MANPOWER: 13 man-weeks

ACTIONS IN TASK 14: Investigate the utilities of the multiyear model for (a) crop estimation with missing data due to cloud cover or poor quality images and (b) early-season, midseason, and late-season estimations derived by estimating R (the efficiency parameter defined above) in each case.

TIME:	13 weeks
MANPOWER:	13 man-weeks

ORIGINAL PAGE IS
OF POOR QUALITY

TASKS 15 AND 16: ROTATION SAMPLE DESIGNS

BACKGROUND: The multiyear model (38) in its simplest form is

$$A_{ts} = \alpha_t + b_s + e_{ts} \tag{51}$$

where A_{ts}, α_t, b_s, and e_{ts} are defined in section 6.1.2.2.

The results given by this rotation sample design are the estimates of the average wheat acreage per segment in year t = T where

$$\underset{\sim}{\alpha} = (X\ H^{-1}X)^{-1}\ X\ H^{-1}\underset{\sim}{a} \tag{55}$$

$$\text{with } Var(\alpha_4) = \frac{1 + 2\gamma}{8}\left(1 + \frac{1}{1 + 2\gamma} - \frac{2}{1 + \gamma}\right)\sigma_e^2 \tag{58}$$

ACTIONS IN TASK 15: Use postsegment acreage estimates for wheat and small grains over 4 years for North Dakota to examine the rotation sample design (fig. 6-1 in section 6).

TIME: 10 weeks

MANPOWER: 10 man-weeks

ACTIONS IN TASK 16: Develop an aggregation procedure tailored to the rotation sample designs. The combined sampling and aggregation procedure will be called "Rotation-Prestratified Sampling and Aggregation." (Use the 4 years of available data.)

TIME: 10 weeks

MANPOWER: 10 man-weeks

TASKS 17 AND 18:	FULL-FRAME SAMPLING
BACKGROUND:	The LARS at Purdue University is investigating the feasibility of a full-frame sampling strategy to provide an alternate, yet more efficient estimator, based on a sampling unit size which is as small as one pixel.
ACTIONS IN TASK 17:	Evaluate all procedures recommended by the LARS study. Appraise and recommend future development and implementation of such strategies. (This is a task-defining task.)
TIME:	8 weeks
MANPOWER:	12 man-weeks
ACTIONS IN TASK 18:	Develop an alternative full-frame sample design.
TIME:	18 weeks
MANPOWER:	18 man-weeks

PRECEDING PAGE BLANK NOT FILMED

TASKS 19, 20, AND 21:	SOFTWARE CONVERSION
BACKGROUND:	Currently, most of the sampling and aggregation software and data bases developed are on the PDP 11/45 computer system. The following tasks will be completed to consolidate the sampling and aggregation capabilities on the LARS system at Purdue University and on the ERSYS when it becomes available.
ACTIONS IN TASKS 19, 20, AND 21:	Transfer the LACIE and TY sampling and aggregation softwares from the PDP 11/45 to the LARS system with control configurations specified and procedure requirements documented. Transfer the PDP 11/45 data to the LARS systems for all supporting USGP and foreign countries data bases consisting of agricultural statistics and CAMS segment estimates. Maintain a record of all data files thus created on the LARS system. Conduct verification tests for the transferred softwares. These tests are to be maintained on the LARS system until the in-house ERSYS becomes available. Then similar conversion will be required for getting all software and data bases implemented in-house.
TIME:	4 weeks for data 5 weeks for aggregation 13 weeks for development 4 weeks for multicrop allocation 4 weeks for within-stratum variance
MANPOWER:	4 man-weeks for data 5 man-weeks for aggregation TY 13 man-weeks for the development of simulation of AgRISTARS crop production 4 man-weeks for multicrop allocation 4 man-weeks for within-stratum variance

TASKS 22 AND 23:	ASSEMBLAGE OF DATA BASES
ACTIONS IN TASK 22:	Prepare a data base containing CAMS segment estimates for early season, midseason, and late season as well as the blind-site true crop proportions from LACIE Phases II and III.
TIME:	6 weeks
MANPOWER:	6 man-weeks
ACTIONS IN TASK 23:	Assemble and file all CAS reports for LACIE Phases I, II, and III and TY.
TIME:	4 weeks
MANPOWER:	2 man-weeks

TASK 24:	DOCUMENTATION
ACTION IN TASK 24:	The procedure requirements for the sampling and aggregation software will be specified and documented (1978-80).
TIME:	12 weeks
MANPOWER:	18 man-weeks

APPENDIX B

APPLIED TASKS: SUMMARIES OF CROP TYPES, BACKGROUND, TIME, AND MANPOWER IN VARIOUS COUNTRIES

TASK TITLE: U.S. CORN/SOYBEAN PILOT AGGREGATIONS FOR AREA AND PRODUCTION

TASK 26:	Prepare aggregation data base
SCHEDULE AND TIME:	December 15, 1980 - March 1, 1981; 10 weeks
MANPOWER:	10 man-weeks
TASK 27:	Prepare for operational implementation
SCHEDULE AND TIME:	May 1 - June 15, 1981; 6 weeks
MANPOWER:	12 man-weeks
TASK 28:	Perform pilot segment allocation
SCHEDULE AND TIME:	November 1, 1979 - February 1, 1980; 13 weeks
MANPOWER:	15.6 man-weeks
TASK 29:	Aggregate area and production
SCHEDULE AND TIME:	June 15 - August 1, 1981; 6 weeks
MANPOWER:	6 man-weeks
TASK 30:	Prepare report for accuracy assessment
SCHEDULE AND TIME:	August 1 - October 1, 1981; 8 weeks
MANPOWER:	12 man-weeks
TASK 31:	Refine baseline procedures
SCHEDULE AND TIME:	July 1 - October 1, 1980; 13 weeks
MANPOWER:	13 man-weeks
TASK 32:	Refine design procedures
SCHEDULE AND TIME:	October 15, 1980 - January 1, 1981; 10 weeks
MANPOWER:	4 man-weeks
TASK 33:	Demonstrate multiyear model
SCHEDULE AND TIME:	January 1 - April 1, 1981; 13 weeks
MANPOWER:	13 man-weeks

TASK 34:	Demonstrate partial response model
SCHEDULE AND TIME:	February 1 - May 1, 1981; 13 weeks
MANPOWER:	13 man-weeks

TASK 35:	Demonstrate other procedures
SCHEDULE AND TIME:	February 1 - May 1, 1981; 13 weeks
MANPOWER:	6.5 man-weeks

BACKGROUND: These tasks will be implemented after the improvement is made to the multicrop sampling and aggregation procedures for U.S. corn and soybeans. The development, testing, and evaluation procedures were improved by using actual and simulated data from selected areas in the United States that approximate the corn and soybean data for Brazil and Argentina.

TASK TITLE: U.S./CANADA WHEAT AND BARLEY PILOT AGGREGATIONS FOR AREA AND PRODUCTION

TASK 37:	Refine baseline procedures
SCHEDULE AND TIME:	July 1 - October 1, 1980; 13 weeks
MANPOWER:	13 man-weeks
TASK 38:	Refine design procedures
SCHEDULE AND TIME:	November 1 - December 15, 1980; 6 weeks
MANPOWER:	3 man-weeks
TASK 39:	Demonstrate multiyear model
SCHEDULE AND TIME:	January 1 - April 1, 1981; 13 weeks
MANPOWER:	6.5 man-weeks
TASK 40:	Demonstrate partial response model
SCHEDULE AND TIME:	February 1 - April 15, 1981; 10 weeks
MANPOWER:	6 man-weeks
TASK 41:	Demonstrate other procedures
SCHEDULE AND TIME:	January 1 - April 15, 1981; 15 weeks
MANPOWER:	7.5 man-weeks
TASK 42:	Prepare aggregation data base
SCHEDULE AND TIME:	September 15 - November 1, 1980, and December 1, 1980 - January 15, 1981; 12 weeks
MANPOWER:	12 man-weeks
TASK 43:	Prepare for operational implementation
SCHEDULE AND TIME:	April 15 - June 1, 1981; 6 weeks
MANPOWER:	12 man-weeks

TASK 44: Check and adjust allocation
SCHEDULE AND TIME: January 1 - March 1, 1980; 8 weeks
MANPOWER: 8 man-weeks

TASK 45: Aggregate area and production
SCHEDULE AND TIME: June 1 - July 15, 1981; 6 weeks
MANPOWER: 6 man-weeks

TASK 46: Prepare report for accuracy assessment
SCHEDULE AND TIME: July 15 - September 15, 1981; 8 weeks
MANPOWER: 12 man-weeks

BACKGROUND: These tasks will be implemented after the overall designed system is exercised by combining the component technologies and the improved baseline multicrop sampling and aggregation procedures. These procedures were applied with actual and simulated data from selected areas in the United States that approximate the data for U.S./Canada spring wheat and barley.

TASK TITLE: U.S.S.R. BARLEY SAMPLING SUPPORT TO EXPLORATORY EXPERIMENT

TASK 48:	Design adaptation
SCHEDULE AND TIME:	December 15, 1980 - February 1, 1981; 6 weeks
MANPOWER:	3 man-weeks
TASK 49:	Test data base
SCHEDULE AND TIME:	March 15 - May 1, 1981; 6 weeks
MANPOWER:	6 man-weeks
TASK 50:	Design a sampling scheme
SCHEDULE AND TIME:	March 1 - April 1, 1980; 4 weeks
MANPOWER:	2 man-weeks
TASK 51:	Demonstrate multiyear model
SCHEDULE AND TIME:	January 1 - April 1, 1981; 13 weeks
MANPOWER:	6.5 man-weeks
TASK 52:	Demonstrate partial response model
SCHEDULE AND TIME:	February 1 - May 1, 1981; 13 weeks
MANPOWER:	6.5 man-weeks
TASK 53:	Demonstrate grouping logic
SCHEDULE AND TIME:	January 1 - May 1, 1981; 17 weeks
MANPOWER:	8.5 man-weeks
TASK 54:	Demonstrate other procedures
SCHEDULE AND TIME:	April 1 - August 1, 1981; 17 weeks
MANPOWER:	17 man-weeks

TASK 55:	Select segments
SCHEDULE AND TIME:	April 1 - June 1, 1980; 8 weeks
MANPOWER:	4 man-weeks

TASK 56:	Define Foreign Similarity Region (FSR) segments
SCHEDULE AND TIME:	October 15 - December 1, 1980; 6 weeks
MANPOWER:	3 man-weeks

TASK 57:	Prepare report
SCHEDULE AND TIME:	August 1 - October 1, 1981; 8 weeks
MANPOWER:	12 man-weeks

BACKGROUND: These tasks require employing the improved sampling scheme.

TASK TITLE: U.S.S.R. BARLEY PILOT EXPERIMENT AGGREGATIONS FOR AREA AND PRODUCTION

TASK 59:	Refine design procedures
SCHEDULE AND TIME:	May 15 - August 1, 1981; 6 weeks
MANPOWER:	3 man-weeks
TASK 60:	Refine baseline procedures
SCHEDULE AND TIME:	September 1, 1981 - January 1, 1982; 17 weeks
MANPOWER:	17 man-weeks
TASK 61:	Prepare aggregation data base
SCHEDULE AND TIME:	September 15 - November 1, 1981; 6 weeks
MANPOWER:	6 man-weeks
TASK 62:	Perform pilot segment allocation
SCHEDULE AND TIME:	December 15, 1980 - March 15, 1981; 13 weeks
MANPOWER:	15.6 man-weeks

BACKGROUND: These tasks will be implemented after the improvements are made to the multicrop sampling and aggregation procedures for the U.S.S.R. barley. The development, testing, and evaluation procedures were improved by using actual and simulated data from selected areas in the United States that approximate the wheat and barley data for the U.S.S.R.

TASK TITLE: BRAZIL CORN AND SOYBEAN SAMPLING SUPPORT TO EXPLORATORY EXPERIMENT

TASK 64:	Study sample frame
SCHEDULE AND TIME:	December 1, 1980 - August 1, 1981; 26 weeks
MANPOWER:	20 man-weeks
TASK 65:	Design a sampling scheme
SCHEDULE AND TIME:	March 1 - March 22, 1980; 3 weeks
MANPOWER:	0.6 man-week
TASK 66:	Select segments
SCHEDULE AND TIME:	March 22 - April 22, 1980; 4 weeks
MANPOWER:	0.8 man-week
TASK 67:	Define FSR segments
SCHEDULE AND TIME:	October 15 - December 1, 1980; 6 weeks
MANPOWER:	3 man-weeks
TASK 68:	Locate segments
SCHEDULE AND TIME:	April 22 - June 1, 1980; 6 weeks
MANPOWER:	1.2 man-weeks

BACKGROUND: These tasks will implement the improved sampling scheme.

TASK TITLE: ARGENTINA CORN/SOYBEAN EXPLORATORY EXPERIMENT AGGREGATIONS FOR AREA AND PRODUCTION

TASK 70:	Refine baseline procedures; design adaptation
SCHEDULE AND TIME:	March 15 - May 1, 1981; 6 weeks
MANPOWER:	3 man-weeks
TASK 71:	Procedures adaptation
SCHEDULE AND TIME:	August 1 - November 1, 1981; 13 weeks
MANPOWER:	13 man-weeks
TASK 72:	Test data base
SCHEDULE AND TIME:	June 15 - August 1, 1981; 6 weeks
MANPOWER:	3 man-weeks
TASK 73:	Define FSR segments
SCHEDULE AND TIME:	October 15 - December 1, 1980; 6 weeks
MANPOWER:	3 man-weeks

BACKGROUND: These tasks will be implemented after improvements are made to the multicrop sampling and aggregation procedures for the Argentina corn and soybeans. The development, testing, and evaluation procedures were improved by using actual and simulated data from selected areas in the United States that approximate corn and soybean data for Argentina.

TASK TITLE: ARGENTINA WHEAT SAMPLING SUPPORT TO EXPLORATORY EXPERIMENT

TASK 75:	Design adaptation
SCHEDULE AND TIME:	March 15 - May 1, 1981; 6 weeks
MANPOWER:	3 man-weeks

TASK 76:	Define FSR segments
SCHEDULE AND TIME:	October 15 - December 1, 1980; 6 weeks
MANPOWER:	3 man-weeks

BACKGROUND: These tasks will result in the implementation of the improved sampling scheme.

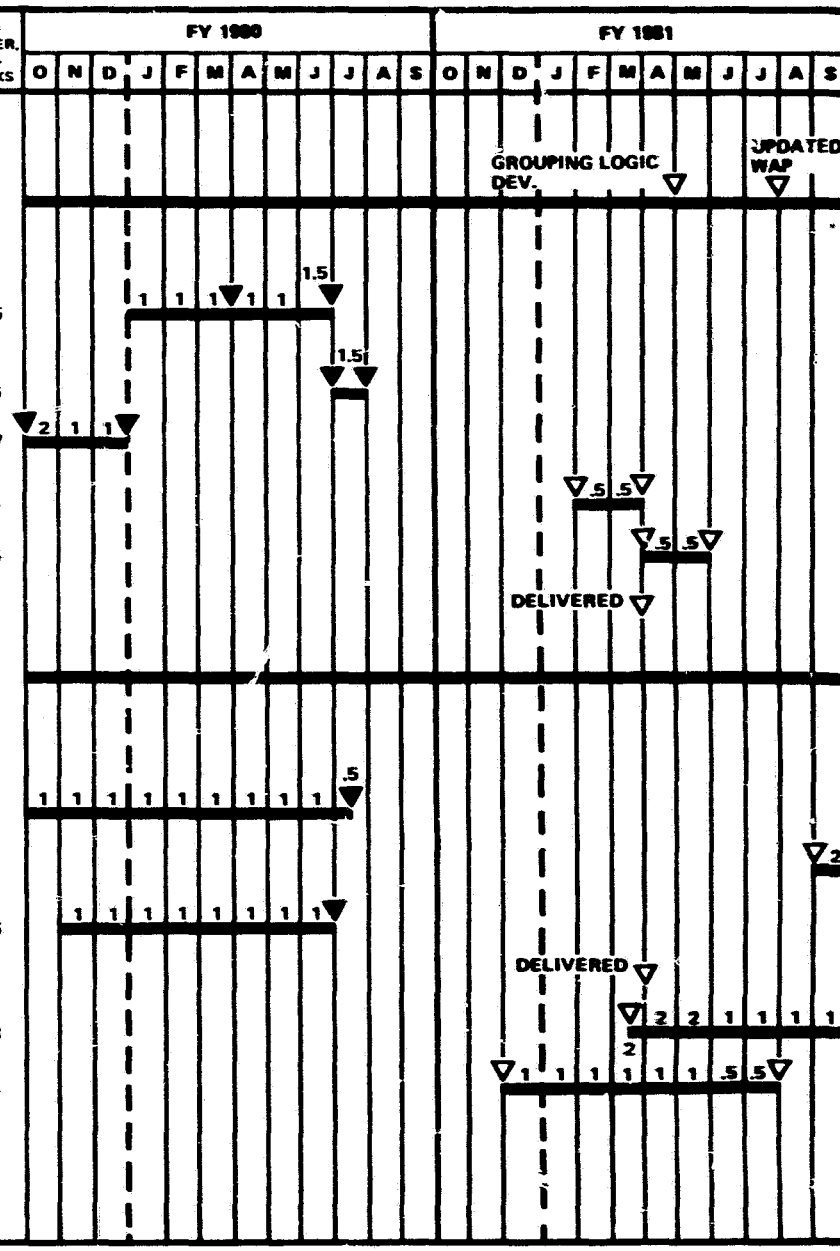
TASK TITLE: AUSTRALIA WHEAT EXPLORATORY EXPERIMENT AGGREGATIONS FOR AREA AND PRODUCTION

TASK 78:	Refine baseline procedures; design adaptation
SCHEDULE AND TIME:	March 15 - May 1, 1981; 6 weeks
MANPOWER:	3 man-weeks
TASK 79:	Procedures adaptation
SCHEDULE AND TIME:	September 1 - December 1, 1981; 13 weeks
MANPOWER:	13 man-weeks
TASK 80:	Test data base
SCHEDULE AND TIME:	August 1 - September 15, 1981; 6 weeks
MANPOWER:	6 man-weeks
TASK 81:	Define indicator region segments
SCHEDULE AND TIME:	October 1 - November 1, 1980; 4 weeks
MANPOWER:	4.8 man-weeks
TASK 82:	Define FSR segments
SCHEDULE AND TIME:	October 15 - December 1, 1980; 6 weeks
MANPOWER:	6 man-weeks

APPENDIX C

PROJECTED SCHEDULES OF TASKS ACCORDING TO PRIORITY, TIME ALLOCATION, AND MANPOWER NEEDS

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN-POWER, MAN-WEEKS
DEVELOPMENTAL TASKS			
◉ GROUPED OPTIMAL AGGREGATION TECHNIQUE IMPROVEMENTS	1		
• ESTIMATION OF MATRIX H_T			
1. Develop an estimation procedure	1	13	15
2. Estimate $\hat{H}_T$ using crop statistics of Illinois, Indiana, and Iowa corn/soybeans	1	4	6
• TESTING OF STRATA GROUPING APPROACH	1	13	17
3a. Restricted and unrestricted groupings	1	8	4
3b. Large area historical data procedure development	1	8	4
4*. Software development	1	•	•
◉ SIMULATION MODEL IMPROVEMENTS	1		
• SIMULATION OF AgRISTARS CROP PRODUCTION			
5a. Segment crop proportion and segment acquisition simulation (corn/soybeans)	1	35	35
5b. Multiyear data simulation	1		
6. Software development (documentation)	1	35	35
7a.* New simulation software development	1	•	•
7b. Update simulation model	1	24	48
8. New cloud cover models development	1	26	24

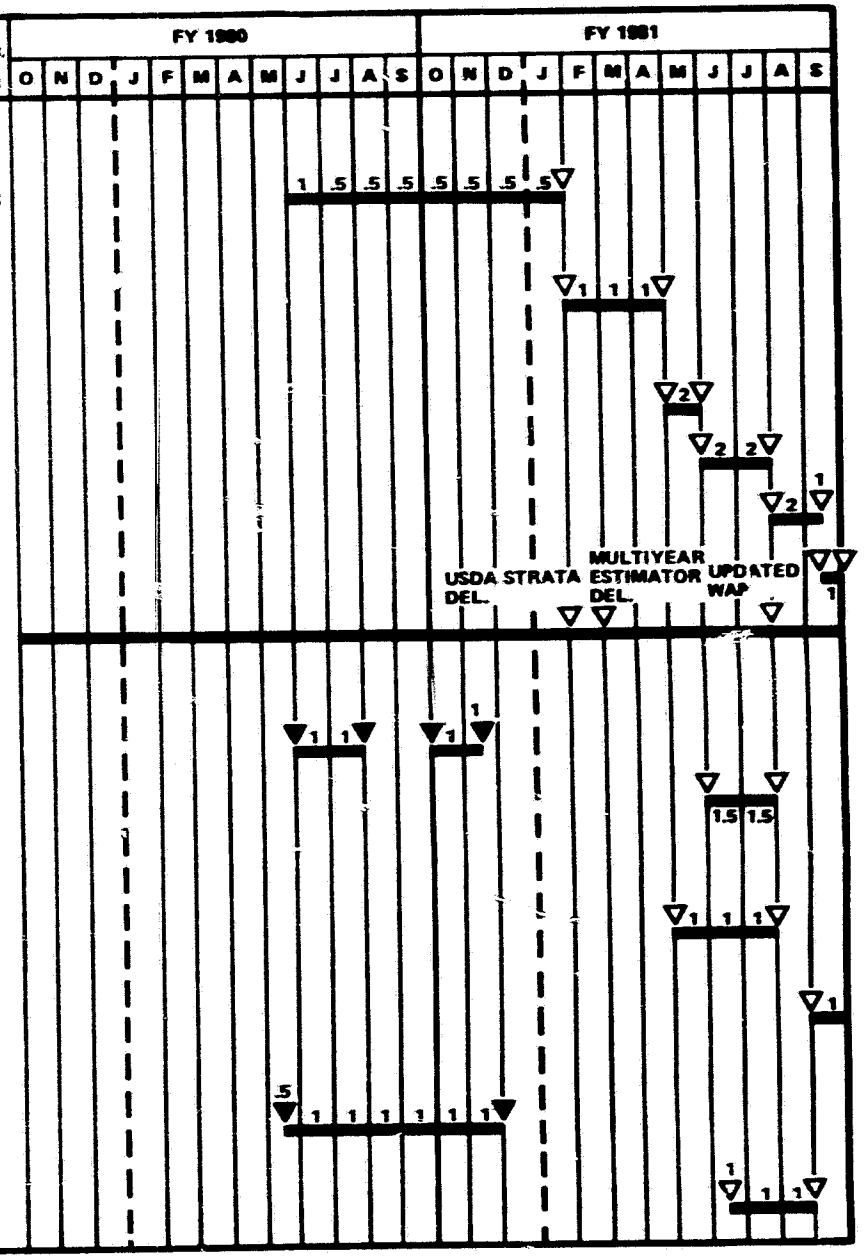


*Indicates the task does not belong to the Sampling and Aggregation group.

ORIGINAL PAGE IS OF POOR QUALITY

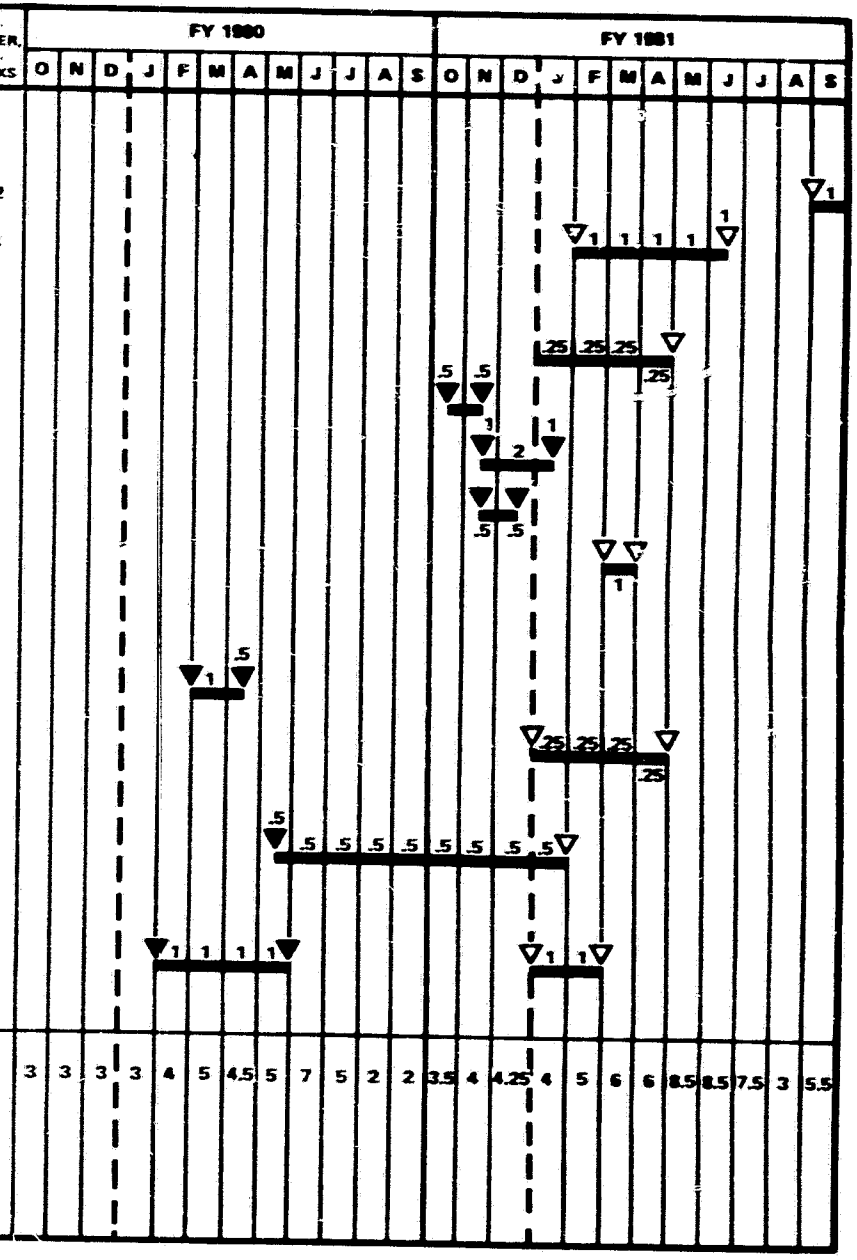
ORIGINAL PAGE IS OF POOR QUALITY

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN-POWER, MAN-WEEKS
◉ SAMPLING FRAME DEVELOPMENT			
9. *Sampling unit size study	1	35	18.5
10. *Design and implementation of sampling frame			
a. Software requirements	3	13	13
b. Programming			
Design specifications	3	4	8
Coding	3	9	18
Documentation	3	4	9
c. Acceptance	3	2	2
◉ ALTERNATIVE STRATIFICATION APPROACHES	1		
• MULTICROP SAMPLE ALLOCATION			
11. Test the within-stratum variance estimation procedure	2	14	14
12. Alternative methods of estimating within stratum variance	2	8	12
• HARTLEY MULTIYEAR MODEL			
13. Apply multiyear estimators to weighted aggregation procedure (Adaptation of WAP to MY WAP)	2	13	13
14. Examine the sensitivity of the multiyear models to nonresponse	2	13	13
• ROTATION SAMPLE DESIGNS			
15. Investigate the efficiency of the designs	2	26	26
16. Develop Rotation - Prestratified sampling and aggregation procedure	2	10	10



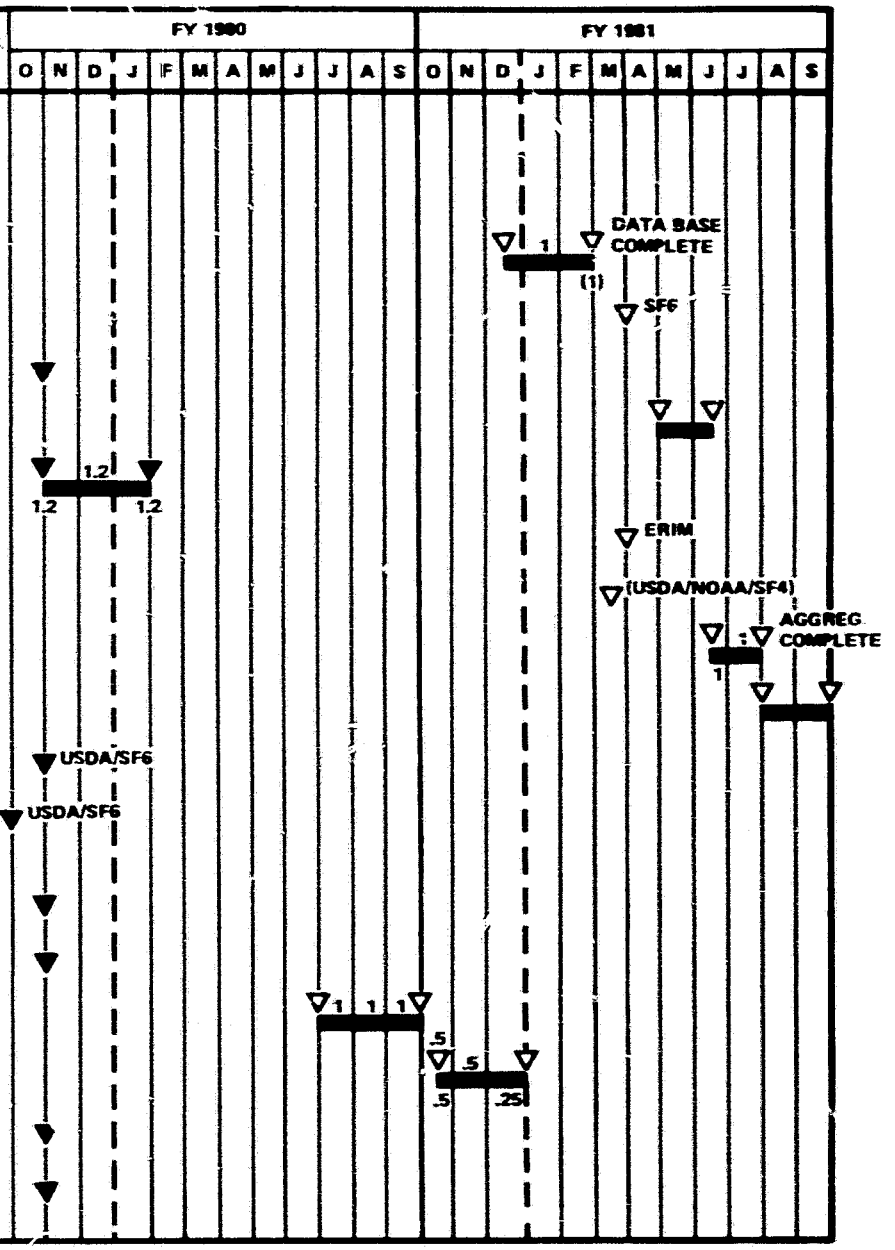
*Indicates the task does not belong to the Sampling and Aggregation group.

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MANPOWER, MAN-WEEKS
• FULL-FRAME SAMPLING			
17. Evaluate the proposed sampling strategies	1	8	12
18. Develop an alternative full-frame sample design	1	18	18
◉ CONVERSION OF PDP SOFTWARE/DATA BASE TO LARS			
19. Data	3	4	4
20. Aggregation TY	3	5	5
21a. Development of simulation of AgRISTARS crop production	3	6	13
21b. Multicrop Allocation	3	4	4
21c. Within-stratum variance	3	4	4
◉ LACIE DATA BASE			
22. Prepare a data base of LACIE Phases II and III early-season, midseason, and late-season estimates	1	6	6
23. File CAS reports for LACIE Phases I, II, and III and for TY	1	4	2
◉ DOCUMENTATION			
24. ERSYS software requirement documents	1	36	18
◉ IMPLEMENTATION PLAN			
25. Document	1	17	17
TOTAL DEVELOPMENTAL TASKS			



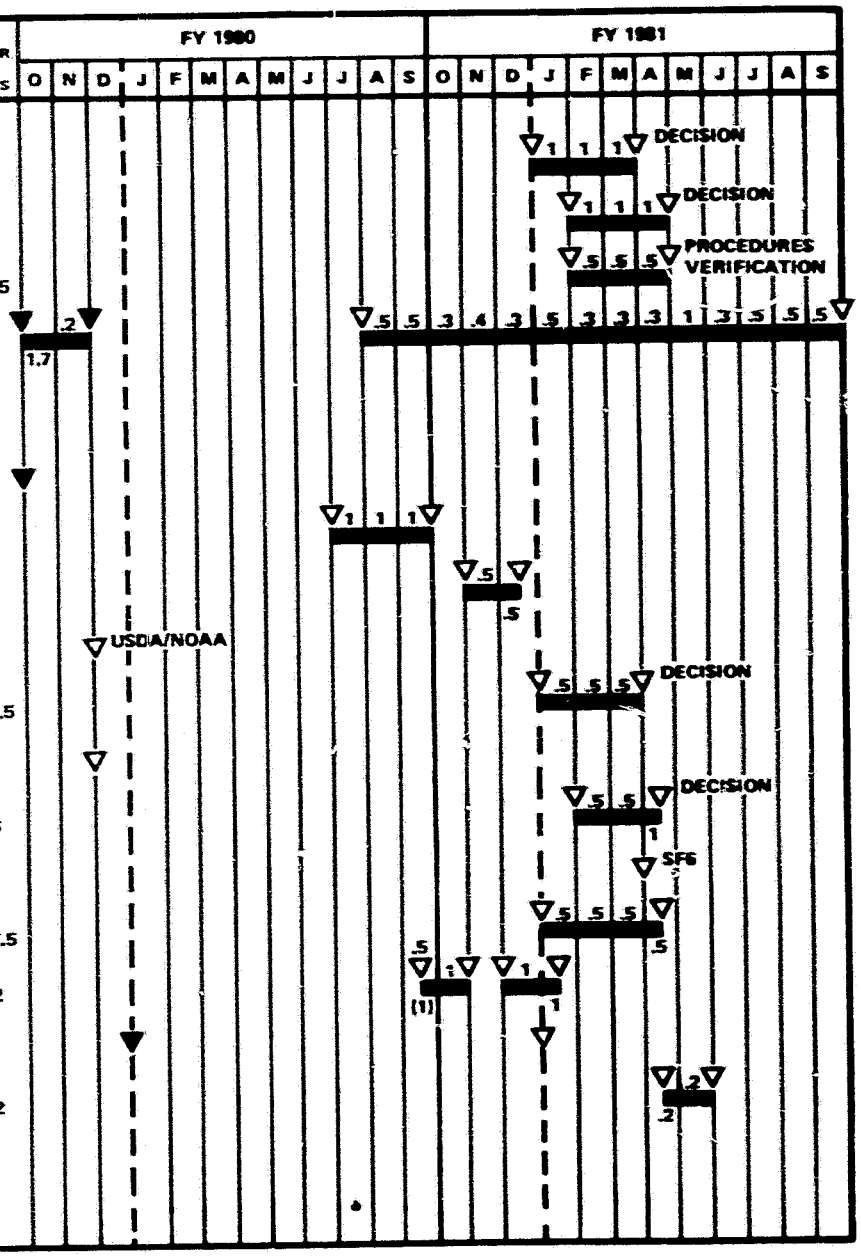
*Indicates the task does not belong to the Sampling and Aggregation group.

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN POWER MAN WEEKS
APPLIED TASKS			
U.S. CORN/SOYBEAN PILOT AGGREGATIONS FOR AREA AND PRODUCTION			
26. Prepare aggregation data base	1	10	10
(SF6) Quality assured software	1		
(SF6) Prepare allocation data base (USDA)	1		
27. Prepare for operational implementation	1	6	12
28. Perform pilot segment allocation	1	13	15.6
*Segment proportions complete (early-, mid-, and late-season)	1		
*Yield estimate and associated variance received	1		
29. Aggregate area and production	1	6	6
30. Prepare report for accuracy assessment	1	8	12
*Receipt of accuracy goals	1		
*Create agronomic data base	1		
*Determine U.S. and Foreign similarity regions and target area (experiment design)	1		
*Create pilot target area sampling frame	1		
31. Refine baseline procedures	1	13	13
32. Refine design procedures	1	10	4
*Define yield strata	1		
*Define pilot classification approach	1		



*Indicates the task does not belong to the Sampling and Aggregation group.

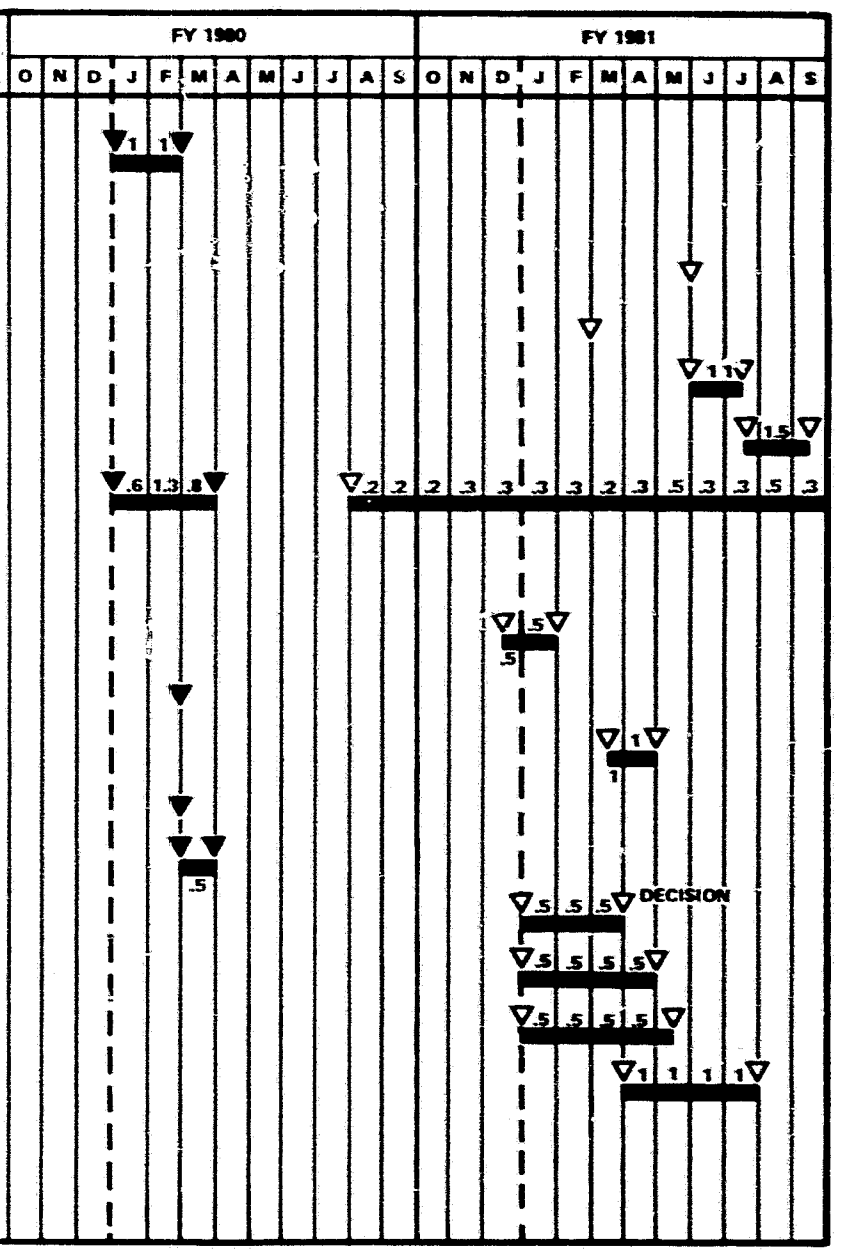
TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN POWER MAN WEEKS
33. Demonstrate multiyear model	1	10	10
34. Demonstrate partial response model	1	13	13
35. Demonstrate other procedures	1	13	6.5
36. Other task-related actions			
U.S./CANADA SPRING SMALL GRAINS PILOT AGGREGATIONS FOR AREA AND PRODUCTION			
*Receipt of accuracy goal	1		
37. Refine baseline procedures	1	13	13
38. Refine design procedures	1	6	3
*Define yield strata	1		
39. Demonstrate multiyear model	1	13	6.5
*Define pilot classification approach	1		
40. Demonstrate partial response model	1	10	6
*Quality assured software to support pilot test delivered	1		
41. Demonstrate other procedures	1	15	7.5
42. Prepare aggregation data base	1	12	12
*Prepare allocation data base	1		
43. Prepare for operational implementation	1	6	12



*Indicates the task does not belong to the Sampling and Aggregation group.

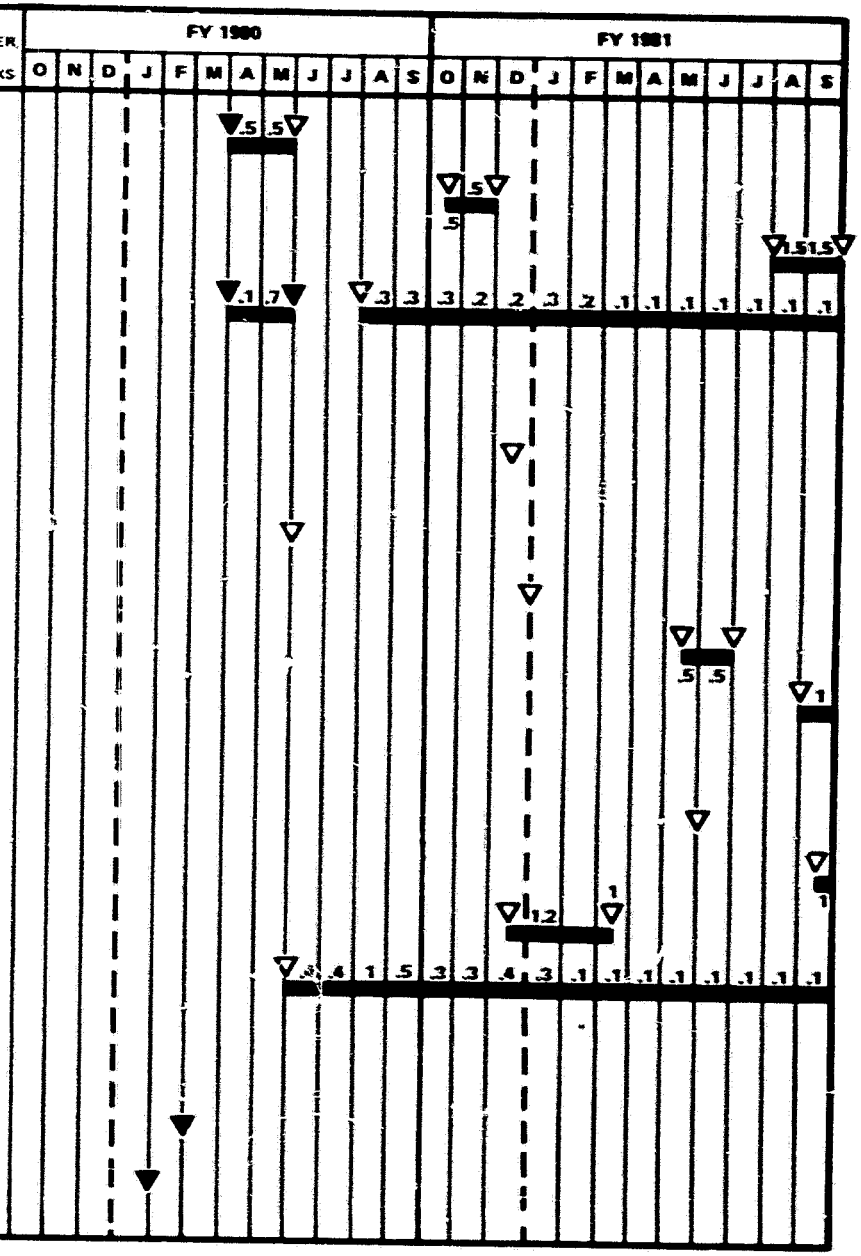
ORIGINAL PAGE IS OF POOR QUALITY

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN POWER, MAN WEEKS
44. Check and adjust allocation	1	8	8
*Perform pilot aggregation:	1		
Segment proportions (early-, mid-, and late-season)	1		
Complete yield estimate and associated variance received	1		
45. Aggregate area and production	1	6	6
46. Prepare report for accuracy assessment	1	8	12
47. Other task-related actions			
U.S.S.R. BARLEY SAMPLING SUPPORT TO EXPLORATORY EXPERIMENT			
48. Design adaptation	1	6	3
*Receipt of exploratory Landsat data requirements	1		
49. Test data base	1	6	6
*Select exploratory target areas (Experiment design)	1		
50. Design a sampling scheme	1	4	2
51. Demonstrate multiyear model	1	10	5
52. Demonstrate partial response model	1	10	5
53. Demonstrate grouping logic	1	17	8.5
54. Demonstrate other procedures	1	17	17



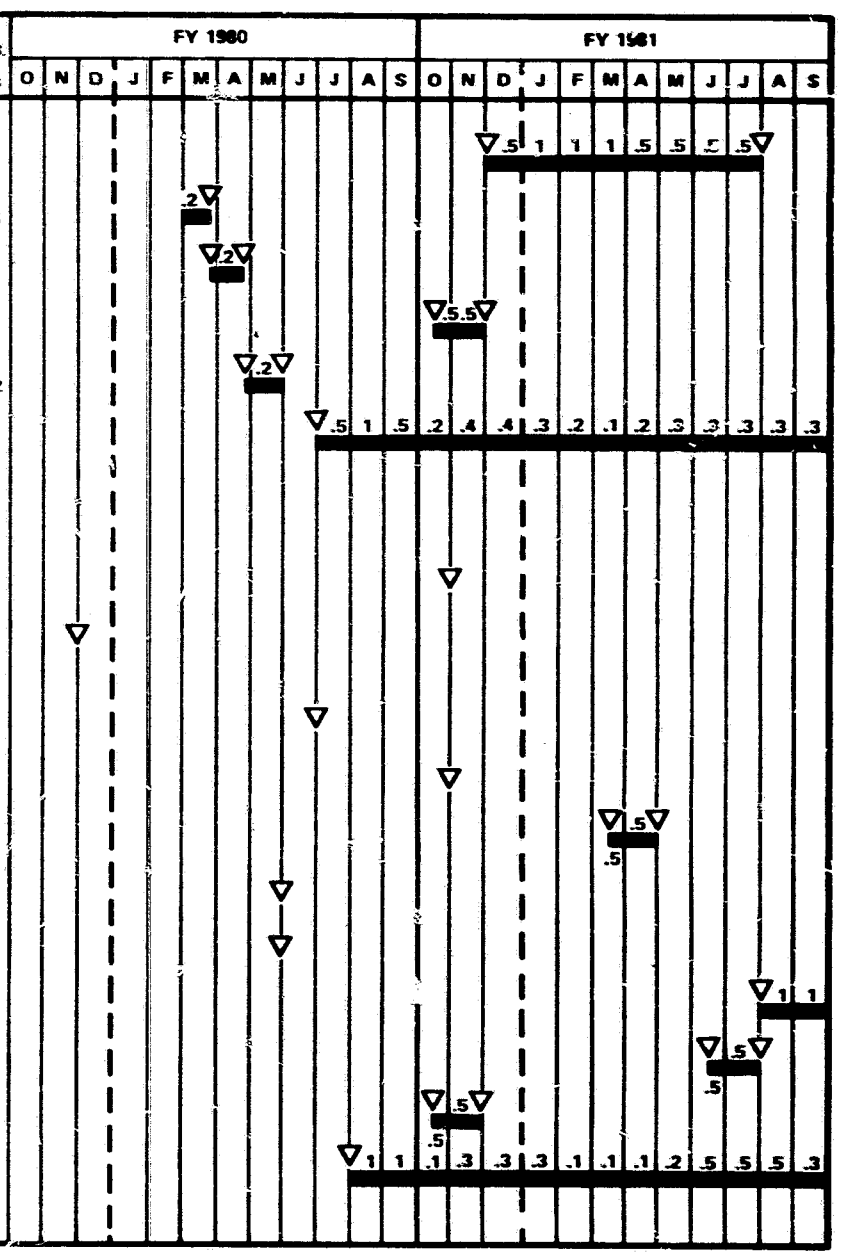
*Indicates the task does not belong to the Sampling and Aggregation group.

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN POWER, MAN WEEKS
55. Select segments	1	8	4
56. Define FSR segments	1	6	3
57. Prepare report	1	8	12
58. Other task-related actions			
U.S.S.R. BARLEY PILOT EXPERIMENT AGGREGATIONS FOR AREA AND PRODUCTION			
*Receipt of accuracy goals; create agronomic data base	1		
*Determine U.S. and Foreign similarity regions and target area (experiment design)	1		
*Create pilot target area sampling frame (USDA)	1		
59. Refine design procedures	1	6	3
60. Refine baseline procedures	1	17	12
*Define yield strata	1		
*Define pilot classification approach	1		
61. Prepare aggregation data base	1	6	6
62. Perform pilot segment allocation	1	13	15.6
63. Other task-related actions			
BRAZIL CORN/SOYBEAN SAMPLING SUPPORT TO EXPLORATORY EXPERIMENT			
*Receipt of exploratory Landsat data base (experiment design)	1		
*Select exploratory target areas (experiment design)	1		



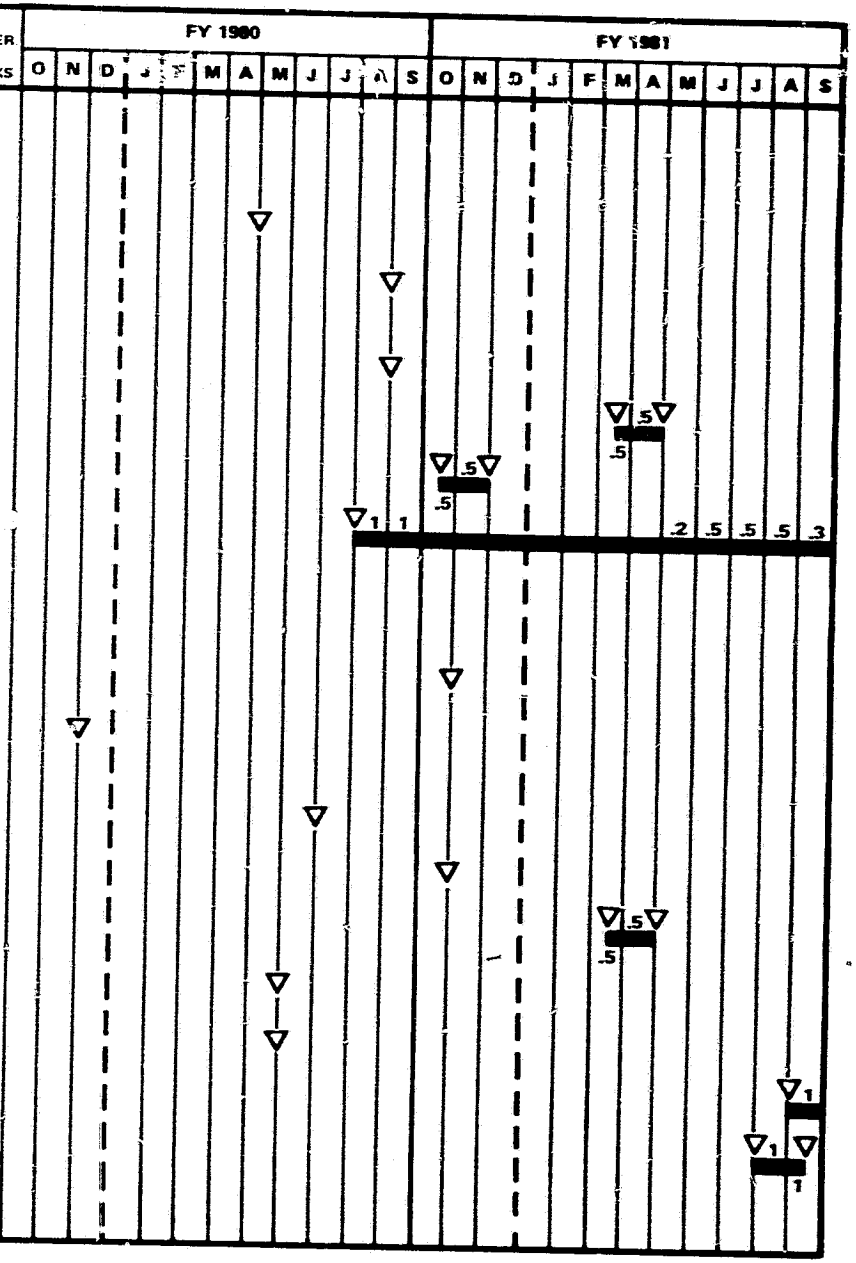
*Indicates the task does not belong to the Sampling and Aggregation group.

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN POWER, MAN WEEKS
64. Study sample frame	1	26	20
65. Design a sampling scheme	1	3	.6
66. Select segments	1	4	.8
67. Define FSR segments	1	6	3
68. Locate segments	1	6	1.2
69. Other task-related actions			
ARGENTINA CORN/SOYBEAN EXPLORATORY EXPERIMENT AGGREGATIONS FOR AREA AND PRODUCTION			
*Receipt of accuracy goal (experiment design)	1		
*Create agronomic data base (experiment design)	1		
*Determine U.S. and Foreign similarity regions and target areas (experiment design)	1		
*Create pilot target area sampling frame (USDA)	1		
70. Refine baseline procedures; design adaptation	1	6	3
*Define yield strata	1		
*Define pilot classification approach	1		
71. Procedures adaptation	1	13	13
72. Test data base	1	6	3
73. Define FSR segments	1	6	3
74. Other task-related actions			



*Indicates the task does not belong to the Sampling and Aggregation group.

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN POWER PLAN WEEKS
ARGENTINA WHEAT SAMPLING SUPPORT TO EXPLORATORY EXPERIMENT			
*Create agronomic data base	1		
*Receipt of exploratory Landsat data requirements	1		
*Select exploratory target areas and U.S. and Foreign similarity regions	1		
75. Design adaptation	1	6	3
76. Define FSR segments	1		3
77. Other task-related actions			
AUSTRALIA WHEAT EXPLORATORY EXPERIMENT AGGREGATIONS FOR AREA AND PRODUCTION			
*Receipt of accuracy goals (experiment design)	1		
*Create agronomic data base (experiment design)	1		
*Determine U.S. and Foreign similarity regions and target areas (experiment design)	1		
*Create pilot target area sampling frame (USDA)	1		
78. Refine baseline procedures; design adaptation	1	6	3
*Define yield strata	1		
*Define pilot classification approach	1		
79. Procedures adaptation	1	13	13
80. Test data base	1	6	6



*Indicates the task does not belong to the Sampling and Aggregation group.

TASK TITLES AND DESCRIPTIONS	PRIORITY	TIME, WEEKS	MAN-POWER, MAN WEEKS	FY 1980												FY 1981											
				O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S
81. Define indicator region segments	1	4														1.2											
82. Define FSR segments	1	6	6														1										
83. Other task-related actions	1	17												1.4	2.2	2.3	2.3	2.8	.8	.1	.1	.1	.1	.1	.1	.1	.1
TOTAL APPLIED TASKS				1.7	1.4	1.2	2.8	2.3	1.5	.8	1.4	.4	3.4	8.4	8.7	13.6	9.4	8.85	10.3	10	8.6	9.5	8.3	7.0	6.7	9.1	11
TOTAL DEVELOPMENTAL TASKS				3	3	3	3	4	5	4.5	5	7	5	2	2	3.5	4	4.25	4	5	6	6	8.5	7	7	4	6.5
SUPERVISION				.3	.3	.3	.3	.3	.3	.3	.3	.3	.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
TOTAL				5.0	4.7	4.5	6.1	6.6	6.8	5.6	6.7	7.7	8.7	11.7	12	15.4	14.7	13.6	15.6	16.3	15.9	16.8	18.1	15	15	14.4	18.9

AVERAGE: 7.10

AVERAGE: 15.83

*Indicates the task does not belong to the Sampling and Aggregation group.

ORIGINAL PAGE IS OF POOR QUALITY

APPENDIX D

DATA REQUIREMENTS

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE D-1.- DATA REQUIREMENTS

TASK	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1												X		
2	X													
3												X		
4	X													
5	X	X		X						X			X	
6													X	
7												X		
8									X					
9			X											
10					X	X								
11												X		
12		X					X					X		
13		X										X		
14		X					X							
15		X												
16												X	X	
17												X		
18							X	X				X		
19							X	X						
20							X	X						
21							X	X	X					
22							X	X	X					X
23								X	X					X
24														X
25														X

Symbol definitions:

A - The 1972-76 historical crop states: Illinois, Indiana, and Iowa

B - Wheat and barley in North Dakota and Minnesota

C - Corn and soybeans in Illinois, Indiana, and Iowa

D - The 1978 crop acreage and proportion estimates (USDA/ESCS)

E - Corn and soybeans in Brazil

F - Wheat and barley in U.S.S.R.

G - The CAMS estimates

H - LACIE and TY software, PDP-11/45

I - Software design and implementation

J - Initial within-stratum variance estimates

K - Sampling frame, USDA/ESCS

L - Theoretical research

M - Monte Carlo technique

N - Documentation and reports

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

☆ U.S. GOVERNMENT PRINTING OFFICE: 1981 — 771-014/1028