CR-171645
C.1

AgRISTARS

MU-E2-04226
NAS9-15476

E83-10350

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing
January 1982

Multiple

FINAL REPORT

RESEARCH AND DEVELOPMENT OF LANDSAT-BASED CROP INVENTORY TECHNIQUES

R. Horvath, R. C. Cicone, W. A. Malila

(E83-10350) RESEARCH AND DEVELOPMENT OF LANDSAT-BASED CROP INVENTORY TECHNIQUES Final Report, 16 Nov. 1980 - 31 Dec. 1981 (Environmental Research Inst. of Michigan) 300 p HC A13/MF A01 CSCL 02C G3/43 N83-27325 Unclas 00350

ERIM
ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN
ANN ARBOR, MICHIGAN

UCB
SPACE SCIENCES LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

NASA

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
MU-E2-04226		

4. Title and Subtitle

Research and Development of Landsat-Based Crop Inventory Techniques

5. Report Date

January 1982

6. Performing Organization Code

7. Author(s) Robert Horvath, Richard C. Cicone and William A. Malila

8. Performing Organization Report No.

152400-21-F

9. Performing Organization Name and Address

Environmental Research Institute of Michigan
P.O. Box 8618
Ann Arbor, Michigan 48107

10. Work Unit No.

11. Contract or Grant No.

NAS9-15476

13. Type of Report and Period Covered

Final Report
16 November 1980 through
31 December 1981

12. Sponsoring Agency Name and Address

NASA/Johnson Space Center
Houston, Texas 77058

14. Sponsoring Agency Code

15. Supplementary Notes

The NASA Technical Coordinators for the activities reported herein were Dr. G. Houston/SG3 and M. C. Trichel/SH3.

16. Abstract

This final report summarizes the research activities conducted by the Environmental Research Institute of Michigan (ERIM) for NASA/JSC in 1981 under two projects of the AgRISTARS (Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing) Program. These are the Supporting Research (SR) and the Inventory Technology Development (ITD) projects. The reported work was part of a larger effort conducted by a consortium composed of ERIM and the Space Sciences Laboratory of the University of California at Berkeley (UCB).

A wide spectrum of technology pertaining to the inventory of crops using Landsat without in situ training data is addressed. Methods considered include Bayesian based through-the-season methods, estimation technology based on analytical profile fitting methods, and expert-based computer aided methods. Though the research was conducted using U.S. data, consideration was given to the adaptation of the technology to the Southern Hemisphere, especially Argentina.

17. Key Words	18. Distribution Statement
AgRISTARS, Landsat, Crop Inventory, Through-The-Season Estimation, Multi-Segment Estimation, Crop Signatures, Spectral/Spatial Simulation, Software Systems, Argentina Agriculture	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	xxii + 291	

MU-E2-04226
NAS9-15476

FINAL REPORT

RESEARCH AND DEVELOPMENT OF LANDSAT-BASED CROP INVENTORY TECHNIQUES

By

Robert Horvath
Richard C. Cicone
William A. Malila

Environmental Research Institute of Michigan
P.O. Box 8618
Ann Arbor, Michigan 48107

January 1982

PREFACE

The Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing Program, AgRISTARS, is a multi-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. Departments of Agriculture, Commerce, and the Interior, and the U.S. Agency for International Development.

The work reported herein was sponsored by the Supporting Research (SR) Project and Inventory Technology Development (ITD) Project under the auspices of the National Aeronautics and Space Administration, NASA. Mr. Robert B. MacDonald, NASA Johnson Space Center, is the NASA Manager of the SR Project and Dr. Glen Houston was the Technical Coordinator for the reported SR effort. Dr. Jon Erickson is the NASA Manager of the ITD Project and Mr. Mickey Trichel was the Technical Coordinator for the reported ITD effort.

The Environmental Research Institute of Michigan and the Space Sciences Laboratory of the University of California at Berkeley comprised a consortium having responsibility for development of corn/soybeans area estimation procedures for foreign applications. This report focuses primarily on the ERIM efforts in detail, while only summarizing UCB efforts.

This reported research, which addresses a broad spectrum of technical issues related to Landsat-aided crop inventory technology, was performed within the Environmental Research Institute of Michigan's Infrared and Optics Division, then headed by Mr. Richard R. Legault, a Vice-President of ERIM. Mr. Robert Horvath acted as overall Program Manager. Dr. William Malila was Technical Manager of the SR effort, while Mr. Richard Cicone was Technical Manager of the ITD effort.

A number of ERIM personnel share in authorship of this document. In addition to Mr. Horvath, Dr. Malila and Mr. Cicone, contributions were made by (alphabetically): Eric Crist, David Hicks, Karen Johnson, Michael Metzler, Christian Pestre, Frank Pont, Daniel Rice, Albert Sellman, and Brian Thelen. Capable secretarial support was provided by Darlene Dickerson and Patricia Wessling.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Concluded)

LIST OF ILLUSTRATIONS

PRECEDING PAGE BLANK NOT FILMED

LIST OF ILLUSTRATIONS (Continued)

LIST OF ILLUSTRATIONS (Concluded)

LIST OF TABLES

PRECEDING PAGE BLANK NOT FILMED

LIST OF TABLES (Concluded)

LIST OF MAPS

1
INTRODUCTION

This report summarizes the research activities conducted by the Environmental Research Institute of Michigan (ERIM) for NASA under two projects of the AgRISTARS (Agriculture and Resources Inventory Surveys through Aerospace Remote Sensing) Program. These are the Supporting Research (SR) Project and the Inventory Technology Development (ITD) Project (formerly Foreign Commodity Production Forecasting (FCPF) Project). The reported work was part of a larger effort conducted from 15 November 1980 - 31 December 1981 by a consortium composed of ERIM and the Space Sciences Laboratory of the University of California at Berkeley (UCB), for which ERIM had the overall technical lead.

The objective of this report is to give a concise technical description of the research activities conducted, the results achieved, and the technical insights gained. Several of the research topics are supplemented by separate technical reports or papers giving additional details about the research. These supplemental documents are referenced within the main body of the text.

1.1 POTENTIAL CONTRIBUTIONS OF AEROSPACE REMOTE SENSING TO AGRICULTURAL INVENTORY AND ASSESSMENT

Aerial photography has gained a place in operational inventory and assessment activities of the U.S. Department of Agriculture and other state and local government agencies. Aerospace remote sensing technology potentially can make additional contributions. Exploration of this potential is the major objective of the AgRISTARS Program.

A summary of types of information that are potentially extractable from aerospace remote sensing data is presented in Table 1.1. The first is crop identification which has received a majority of the attention in agricultural studies to date, especially in conjunction with crop area estimation. Next are indications of crop development stage and crop

TABLE 1.1 POTENTIAL CONTRIBUTIONS OF AEROSPACE REMOTE SENSING TO AGRICULTURAL INVENTORY AND ASSESSMENT

- Crop Identification
  - Crop Group
  - Crop Type
- Crop Development Stage
  - Planting and Harvesting Progress
  - Key Growth/Development Stages
  - Management Practices (e.g., crop rotations)
- Crop Conditions
  - Vigor, Stress
  - Ground Cover, LAI
  - Management Practices (e.g., irrigation and double cropping)
  - Homogeneity
  - Episodal Events
- Inputs to Yield Models
  - Spectral-based Deductions of Development, Condition, and Management Practices
  - Meteorological
  - Combined Spectral and Meteorological
- Soil Characteristics
- Crop Area
  - Total Area Planted, Emerged, and/or Harvested
  - Area by Crop Group and Crop Type
  - Area by Condition Class
- Crop Production

condition which could provide important inputs to yield models. Soils are a topic that have an important effect on yield and productivity. Together, estimates of crop area and crop yield permit estimates of overall crop production, the "bottom line" of agricultural crop inventories.

Investigations conducted prior to AgRISTARS, such as the Large Area Crop Inventory Experiment (LACIE), have demonstrated the practical feasibility and effectiveness of the sample survey approach for satellite-based estimation of crop area and production. Elements of this approach which were developed and tested were the sample-frame design, the sampling design (allocation and location of sampling units), area estimation or measurement at a segment level, area and production estimation at stratum and large-area levels, and analysis of errors and error sources.

However, the scope and needs of AgRISTARS require technological capabilities beyond those demonstrated previously, necessitating continued research and development activities. For example, the single-crop focus of sampling and measurement procedures needs expansion to mutliple crops, including corn and soybeans. Aggregation procedures should more accurately handle different levels of accuracy in segment-level estimates, including non-response. Also measurement procedures should be more objective and accurate.

Very important are the facts that current Landsat-based crop area estimation technology is not efficient in terms of expert labor, computer, and time resource requirements,is not geared to crop inventory estimates throughout the growing season, and has not been applied to all major crops and world production regions. Improvements are being made during the course of AgRISTARS in sensors (e.g., Thematic Mapper and meteorological satellites), information extraction techniques, inventory system technology, and in joint use of meteorological and spectral data. Also, as a result of AgRISTARS, this technology will be adapted to, and evaluated in, additional geographic regions and for additional crops.

1.2 GENERAL OBJECTIVES OF THE CONTRACT

The contract research was directed at supporting requirements of the two separate AgRISTARS projects. The project activities have both distinct objectives and mutually supportive aspects.

1.2.1 OBJECTIVES UNDER THE SUPPORTING RESEARCH PROJECT

The direction of our Supporting Research Project activities evolved toward support of two broad long-range objectives. The first long-range objective was to develop advanced techniques for timely, efficient, and cost-effective estimation of crop areas using remotely sensed data from Landsat together with cellateral data. These techniques should be capable of generating estimates at any time throughout the growing season, since a capability to produce early estimates is highly desirable for AgRISTARS. They should utilize multiple segments to facilitate efficient and effective crop inventories over large areas. Where a crop/region focus was needed for the research, corn, soybeans, and their confusion crops were to be emphasized, keeping in mind an eventual application in South America.

The second long-range objective was to understand and capitalize on the information content of Landsat MSS and Thematic Mapper data and their relationships to agronomic and biophysical phenomena. A subobjective was to develop simulation and modeling capabilities that will enhance this type of research.

1.2.2 OBJECTIVES UNDER THE INVENTORY TECHNOLOGY DEVELOPMENT PROJECT

The overall objective of the ITD program at ERIM was to research and develop, integrate, implement, test and evaluate technology which uses remote sensing to assist in assessing the status of crops without ground derived observations. The primary focus of this technology is the inventory of the corn and soybeans production in Argentina and Brazil, two countries that are major producers of agricultural commodities and therefore influential in the overall economic and nutritional picture of world food balance.

The specific objective of the work reported in this document was to formulate a base of component technology that, through evaluation in a U.S. scenario, shows promise in being adaptable to agricultural conditions of Argentina and Brazil. Both end-to-end area estimation procedures and component techniques using Landsat MSS would be developed, implemented and objectively tested.

1.3 GENERAL APPROACH

The research activities were divided into two groups of tasks addressing objectives of the SR Project and ITD Project, respectively. SR Project tasks were:

(1) Sampling and Estimation Technology Research

(2) Measurements Technology Research

ITD Project tasks were:

(1) Experiments

(2) Technology Development, Evaluation and Integration

1.3.1 GUIDELINES FOR TECHNOLOGY DEVELOPMENT

The eventual application of research under both AgRISTARS projects is for crop inventories in foreign areas, with emphasis for ERIM/UCB on corn and soybeans area estimation in Argentina and Brazil. This and other sponsor guidelines established general constraints on the types of technology that were to be utilized and developed. These include:

(1) No dependence on direct ground identifications for procedure performance: use permitted only for develooment and evaluation purposes.

(2) Use of Landsat as the prime sensor--MSS now and TM added later.

(3) Initial dependence on segment-based technology, e.g., the 5x6-mile segments utilized in LACIE.

(4) Implementation of selected technology in a formal configuration controlled environment on NASA/JSC AS-3000 computing system.

1.3.2 THE ERIM/UCB CONSORTIUM

A consortium was established to promote a unified attack on the development of corn and soybeans area estimation technology. It was composed of the Environmental Research Institute of Michigan (ERIM) and the Space Sciences Laboratory of the University of California at Berkeley (UCB). Both contractors have had extensive experience in the development of remote sensing technology for agricultural applications, including participation in the LACIE project, and in other applications. They brought complementary capabilities in addition to common understandings and capabilities, forming an effective research team. A majority of the program described in this report was pursued in a joint manner by ERIM and UCB, with ERIM having the overall technical lead. ERIM and UCB efforts are reported separately.

2
SUPPORTING RESEARCH TECHNICAL PROGRESS AND RESULTS

A broad spectrum of research activities was conducted in pursuit of the Supporting Research objectives. They are reported here by research topic at the level of subtasks. Substantial progress was made in several areas.

2.1 GENERAL APPROACH AND TASK STRUCTURE

Two major long-range objectives for our Supporting Research Project activities were identified in Section 1.2.1. A compatible task structure was established, with two major tasks covering research areas in sampling and estimation technology research and in measurement technology research, respectively. The subtasks under those two headings are listed in Table 2.1. This table also identifies the fact that UCB conducted complementary research under the first task whereas only ERIM addressed the second. The nature of UCB research is mentioned where appropriate in this report but is being reported separately [1].

2.1.1 SAMPLING AND ESTIMATION TECHNOLOGY RESEARCH

The identified needs for efficiency, accuracy, and timeliness in estimation impact all aspects of area-estimation procedure research and development: sampling, measurement, aggregation and estimation. The approaches taken in the various subtasks considered these criteria.

Efficiency requirements suggest a high degree of automation throughout a procedure. An ability to process multiple segments without retaining or examining each in detail is highly desirable. Flexibility is another attribute which can enhance efficiency. If elements of the procedure can adapt to local conditions (e.g., degree of complexity) and variable accuracy requirements, overall efficiency gains can be made.

TABLE 2.1. ERIM/UCB SUPPORTING RESEARCH TASK STRUCTURE

Task	Title	Participation ERIM	UCB
1.0	Sampling and Estimation Technology Research		
1.1	Multisegment Estimation Research	X	X
1.2	Through-the-Season Estimation Research	X	X
1.3	Argentina/Brazil Agronomic Understanding	X	X
2.0	Measurement Technology Research		
2.1	Seed-to-Satellite Model Development and Analysis	X	
2.2	Information Extraction Technology Research	X	
(2.3)*	Small-Grains Labeling Techniques	X	

*Not a full subtask; it represents completion of R&D efforts initiated during the preceding year.

Accuracy requirements also impact all elements of a procedure, particularly the measurement element (e.g., information extraction or labeling) which is addressed more fully under the second SR task (Section 2.1.2). One must not lose sight of the interaction between elements; for example, the size of sampling units can affect measurement accuracy.

Timeliness is important both in terms of speed of response, once a particular set of data becomes available, and in terms of being able to produce estimates at any given time throughout the growing season. The latter requires a good understanding of the increasing information content of Landsat data as the season progresses and its use with other forms of information to produce the best possible estimate for each situation.

Lack of ground "truth" observations, especially in foreign regions, hampered LACIE research and development activities. Information from foreign regions is essential for an understanding of differences from U.S. test areas so that developed techniques can be general and extendable or adaptable to those regions. In AgRISTARS, a major regional focus for corn and soybeans inventory technology is South America (Argentina and Brazil). Agencies in Argentina and Brazil have given evidence of being amenable to cooperative ground-truth data collection efforts. Initial data collection efforts were successfully carried out in Argentina early in the contract year, with a minimal amount of time for planning due to the timing of their growing season. Plans were made for new field activities in 1982, though not carried through due to political instability in Argentina.

2.1.2 MEASUREMENT TECHNOLOGY RESEARCH

Measurement technology, which extracts agrophysically meaingful features (including assignment of cover class labels to observations), is a critical element in area estimation procedures that use Landsat, especially those that cannot rely on ground "truth" observations in their operational context. The measurement component of an advanced area estimation procedure must support goals of accuracy, efficiency, timeliness, and information content for advanced procedures that employ multisegment concepts and/or new sensor systems, such as the Thematic Mapper. This requirement defines both the general characteristics of an advanced measurement component and guidelines for research under this task.

The key to extraction and use of meaningful and accurate information from remotely sensed data is the ability to consistently relate observed patterns in the remotely sensed data to agronomic and biophysical characteristics of the various crop and cover classes in the scene. The need has been identified for techniques which are more automatic and objectively perform these functions, especially on spatially registered multidate data sets over large areas.

However, more research and development effort is required to produce techniques and procedures that can attain the full potential of information extraction from remotely sensed and collateral data. In particular, additional research into the relationships between crop phenology and morphology and remote sensing observables is required. Substantial progress was made through study of agronomic literature and analysis of field measurement data.

Use of simulation can help provide the understanding necessary to develop effective information extraction and measurement techniques. Existing simulation models can be useful but need to be improved since they do not adequately represent the full range and character of factors that affect remotely sensed data from agricultural scenes. Three advancements in simulation capability were made during the year.

The crop emphasis of our research was directed to be on corn and soybeans and their confusion crops. During the first part of the contract year, however, we did complete work previously begun on small grains labeling techniques.

2.2 THROUGH-THE-SEASON ESTIMATION RESEARCH

The research emphasis of the AgRISTARS SR and ITD (FCPF) Projects has been broadened from techniques for producing estimates near the end of the growing season to include techniques for producing estimates early in the season. To guide our research, we generalized the problem to one of being able to produce estimates at any given time throughout the growing season, making full use of available information from all sources. Emphasis, consequently, was placed on identifying and extracting the agronomically related information available from Landsat and developing a framework and ways of using it. This emphasis was promoted by our establishment of a context and perspective for viewing the through-the-season (TTS) estimation process and the potential contributions of Landsat, within the general context of AgRISTARS area estimation using stratified estimation approaches with no use of current-year ground observations. The focus was narrowed to estimation of corn and soybean acreages, but the general approach and principles should be applicable to other crops as well. Comments also are made where appropriate to yield and production estimation. Finally, although the data available for study were from the U.S. Corn Belt and, to a lesser extent, the south and southeast United States, portions of the analysis should apply to crops in other countries, such as Argentina.

Landsat is used throughout this section to identify the remote sensing system. In most instances the ideas and concepts would apply to Thematic Mapper data as well, and it should provide additional capability when available.

2.2.1 THROUGH-THE-SEASON ESTIMATION CONCEPTS AND CONTEXT

Crop production assessment can be viewed as a combination of prediction and observation (e.g., direct measurement) processes for crop acreages, yields, and resultant production. Throughout the season, the relative importance of these two processes gradually change. Prediction dominates pre-season forecasts of both farmer's planting intentions and their expected successes. However, as the season progresses, information accumulates and opportunities increase for direct measurement of the actual situations and realizations. Thus, estimates can be updated and refined, based on those measurements.

Figure 2.1 pictorially illustrates the time-varying importance of prediction and observation/measurement in the assessment process. It also indicates the situation for early season estimates and later season estimates in AgRISTARS.

Information for use in crop assessment can come from a variety of sources. Table 2.2 lists conventional sources for predictive and observational variables. It also indicates that remote sensing has the potential for providing both types of information, a reflection of spaceborne sensor's capabilities to survey large areas and to make site-specific and even field-specific identifications of crop type and condition.

Figure 2.2 highlights the general way in which predictive and observational variables would enter the TTS estimation process. The uncertainty in predicting or deducing planting decisions is reduced as the number and/or quality of predictive variables is increased. On the other hand, observational variables provide an increasingly better basis for induction or measurement of the crops actually planted as the season progresses and the number of observations increases. Ideally, one would make use of both types of decision processes to produce the best possible estimates using all available information at the time the estimate is required.

Predictive variables can come from crop identifications and are estimates made for the proceeding year(s) using Landsat data. For instance, they could include crop rotation histories on individual fields which, with knowledge of rotational practices, could establish prior probabilities for specific crops in these fields, before they are

ORIGINAL PAGE IS
OF POOR QUALITY

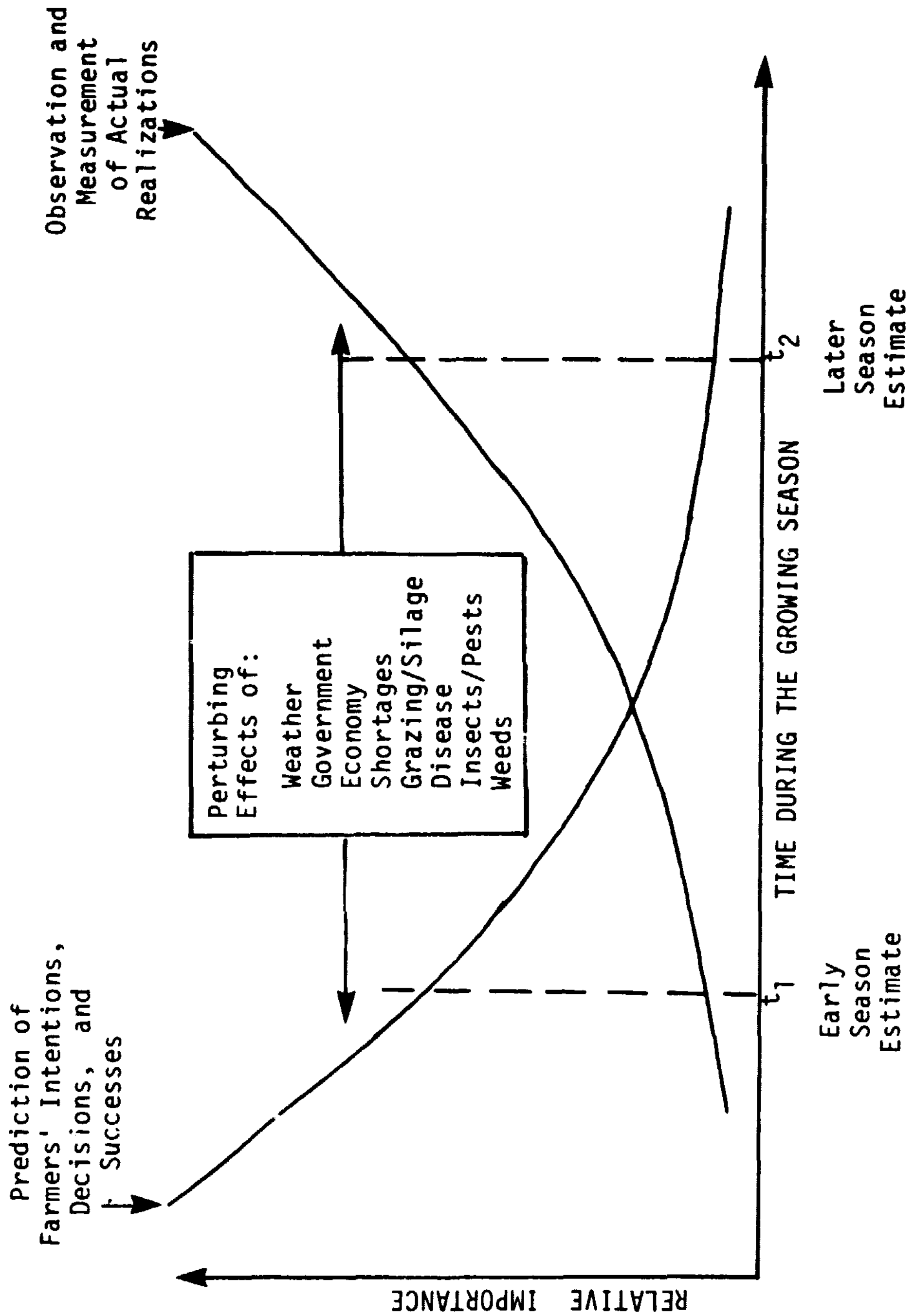


FIGURE 2.1. THE ESSENCE OF CROP PRODUCTION ASSESSMENT

TABLE 2.2. INFORMATION SOURCES FOR CROP ASSESSMENT

Prediction Sources	Observational Sources
Market variables	Enumerative surveys
Government policy	Ag expert observations
Prior weather	Ag exper judgments
Prior experience	Market surveys
Trends in technology	Production reports
Mail surveys	Current-year weather
Remotely sensed data	Remotely sensed data

ORIGINAL PAGE IS
OF POOR QUALITY

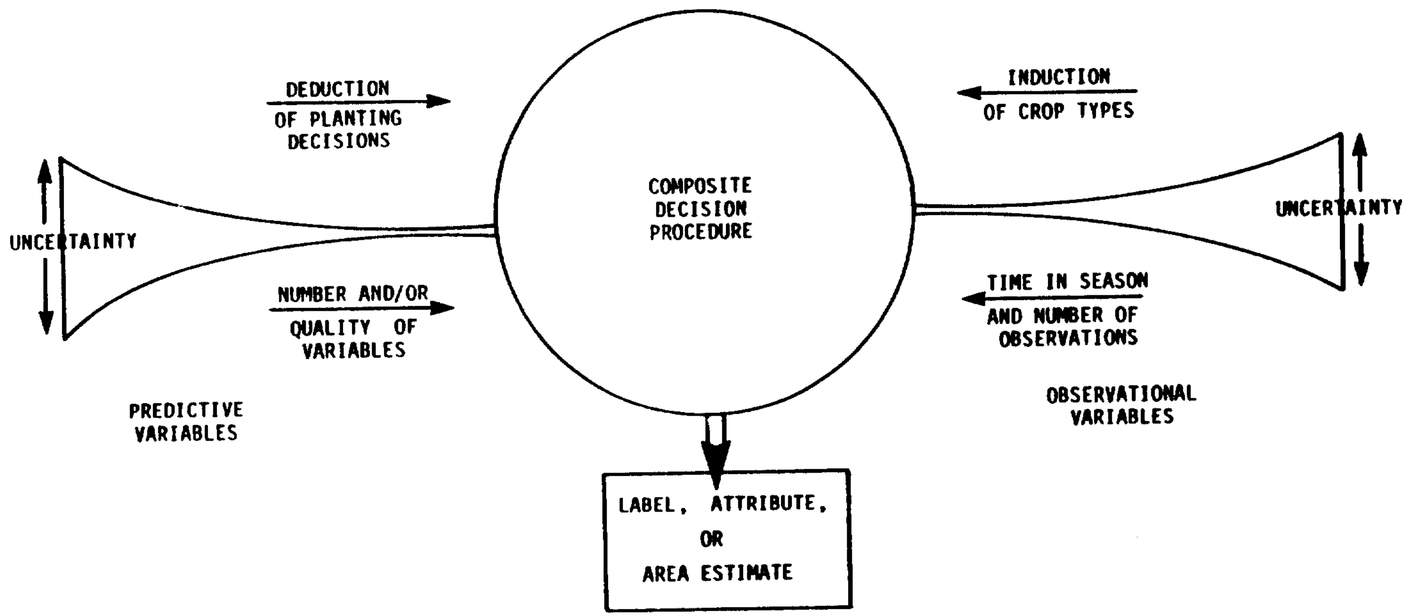


FIGURE 2.2. A MODEL OF THE TTS ESTIMATION PROCESS

observed in the current year. Observations with predictive uses also can be found early in the season when tilled soil is observed rather than plants or when emerged crops are not yet differentiable by Landsat. These could lead to identification of a crop group (e.g., summer crop) before the field can be identified as being corn, soybeans, or sorghum. We have developed a new approach that incorporates this type of information directly into conventional, econometric, crop acreage response models.

As far as direct measurement is concerned, we note that Landsat observes only what is present on the ground at the times of its overpasses. In order to select appropriate features and maximize the amount of agronomic information that is extracted from the Landsat data, one should have a thorough understanding of the practice and history of agriculture in the region(s) being surveyed. In addition, full use should be made of collateral information sources. To facilitate the realization of these needs and provide a perspective for Landsat observables, we suggest that observational data be utilized and analyzed with predictive models of Landsat responses fromthe crops of interest and the relevant scene classes. Note that these models predict the appearances of crops at the times of overpass rather than the crop acreages in segments, strata, or regions, as estimated by the previously mentioned crop acreage response models and related models. They could get down to the detail of how specific crops would appear in specific fields at the specified times.

A final comment is that overall needs for agronomic understanding, effective use of collateral data, and integration of data from multiple segments all are intensified early in the season and also in situations where Landsat coverage is frequently precluded by cloud cover.

The next four sections present in greater detail the concept of continuously merging prediction and direct observation/measurement in TTS estimation, and describe some specific procedures we have developed.

Reference also is made there to publications which have additional detail.

The first of these sections (2.2.2) discusses which agricultural phenomena might be observable by Landsat, what one might deduce about agricultural practices from these observations, and how that knowledge can enhance TTS interpretation of Landsat data.

The second section (2.2.3) develops a new approach for using early season Landsat crop-group area estimates to augment conventional crop acreage response models that predict on the basis of prior yields, prices, acreages, and government policy. An exploratory study is presented which produced encouraging results.

The third section (2.2.4) develops an approach for merging prediction variables and Landsat observational variables in a segment area estimation procedure that has the capability to incorporate multiyear data. A Bayesian classification approach applied to quasi-field targets was chosen as an alternative to direct estimation approaches being pursued at NASA/JSC. Prior probabilities are based on the predictive variables discussed in preceding sections.

The fourth section (2.2.5) introduces the longer-range possibility of building the required capability around the concepts of knowledge engineering, artificial intelligence, or expert systems.

A final section (2.2.6) summarizes the major concepts developed and conclusions drawn from the TTS estimation research.

2.2.2 USING KNOWLEDGE OF AGRICULTURAL PRACTICES TO ENHANCE TTS INTERPRETATION OF LANDSAT DATA

Much of the material summarized in this section is to be described in greater detail in a separate technical report [2].

2.2.2.1 Review of Seasonal Practices and Decisions in Agriculture

The practice of agriculture is, of course, carried out by real farmers, in real fields, under real market conditions, and in real weather. A host of decisions and practices take place which are based on past, present and expected conditions and the personal preferences of the farmer. An understanding of these can improve the process of estimating their results.

Planning. In a farmer's planning for the approaching crop season, expectations of profit and market conditions, previous cropping practices (such as rotation and fallowing), existing soil conditions and weather, etc., all play a role in his decisions. They affect decisions regarding the specific use of each field, as well as the amount of each crop to plant, the varieties to order and plant, the balance to maintain between crops and livestock, and the timing of preparations. Consideration also is given by farmers to the policies of various governments and governmental bodies and to the availability and cost of fuel, fertilizer, and equipment.

Preparation. Based on this planning, fields are prepared by plowing, disking, incorporating fertilizer and/or by fallowing or pasturing. These preparations may take place in the previous growing season, at its end, or early in the current season. More elaborate preparations might include ditching, tiling, leveling or diking for irrigation, as well as drilling of wells and preparation of irrigation equipment.

Planting. Planting will normally follow the planned schedule and prevalent practices, but can depart from them. The weather may be too wet, too dry, or too cold. A late season may force use of another cultivar or another crop. Early planting may fail and require abandonment or replanting. The market may undergo a significant change, forcing a change in crop selection.

Crop Management. During the growing season, decisions to spray, to cultivate, to fertilize, to apply herbicides, or to irrigate will be affected by weather and other factors such as degree of infestations and costs of materials. Catastrophic conditions can cause defoliation, severe lodging, crop failure, and a decisionto abandon or replant a crop. Harvest may be affected in various ways by weather, available storage, market conditions, or need for grazing or silage.

2.2.2.2 Agricultural Features and Events Observable by Landsat

The agricultural practices, features, and events briefly described above may be observed in, or inferred from, Landsat data in some cases. This discussion sets forth a brief introduction to these potentialities. These features have various spatial associations, applying to different strata such as pixels, fields, districts, soil groups, regions and even countries.

Pre-Season Conditions and Planting. Observations continued over several years can be used to determine cropping practices for the individual fields and regions. Crop rotations, for example, can be tabulated and sequences learned to establish prior probabilities for specific crops. Fallowing or green manuring sequences can be observed. Patterns can be found for planting time sequences based on local soil conditions and topography (wetness, contours, etc.) and crop types. In general, an extensive history of each field may be obtained and related to factors affecting subsequent use.

Pre-Planting and Planting. Pre-season preparations may be observed in one or more acquisitions enabling mapping of stubble, plowed ground, wet soils, and predecessor crops. Irrigation preparations or practice may be observed. Flooding and abnormal conditions can be seen. Abandoned crops, unplowed ground, and indications of changed usage can be observed. Possibly various stages of preparation may be distinguished for various crop types.

Early Growth and Growing Season. Emergence may be detected and used to infer planting dates. By continued observation during the season, rates of greening may be determined for all fields. Estimates of percent cover or leaf-area-index and time of peak greenness may be calculated. Declines in greenness and occurrence of reproductive events, such as tasseling and heading may be observed or inferred. Effects of grazing, hail, lodging, disease, flooding, and crop loss also may be observed or inferred. In addition,the duration and general timing of plant cycles may be observed and crop development stages estimated. All of the above are subject to having an adequate acquisition history.

Harvest. Time of harvest and progression of harvest may be monitored. Unusual timing can be noted when crops are cut early for silage or are left unharvested for long periods. Fields may be determined to be abandoned or unharvestable after sufficient time. Beginning of late-season cultural practice may be monitored. Winter cropping practices may be observed and monitored for later mapping.

2.2.2.3 Using Agronomic Understanding to Enhance the Predictive Value of Landsat Data

The predictive aspect of crop assessment in essence attempts to understand the farmers' situations and anticipate both their decisions and the eventual results of those decisions. Landsat's potential to contribute varies as a function of time, as the various agricultural features and events discussed earlier become observable and detectable.

Landsat is usually thought of as providing agricultural information only by direct measurement of crop acreages (etc.) during the current growing season. However, Landsat does have potential for improving predictive capability as well, including both prior-year and current-year aspects.

End-of-season crop area estimates from prior years give a basis for relating sample-segment estimates to aggregated values and values from other sources for larger regions to which they belong, revealing tendencies to be higher or lower. They also provide information on year-to-year variance for individual segments and within-year variance among segments. Over time, the Landsat-based data base may come to rival other sources of information, at least in developing countries.

Existing crop acreage prediction models do not generally include current-year inputs, let alone inputs from Landsat. We gave consideration to ways in which the frequent looks for Landsat might be capitalized upon for predictive purposes. We identified several uses.

One major use of current-year Landsat data we identified and investigated was for augmenting conventional crop acreage response models (CARM's). This study is described in detail in Section 2.2.3, as applied to predicting acreages for summer crops like corn, soybeans, and sorghum. The main idea is that early in the season, before summer crops are differentiable, Landsat still can identify acreages of predecessor crops, like wheat, and identify the total acreage that has been prepared for (and, later, planted to) summer crops.

Use of these current-year quantities should improve acreage predictions for the individual summer crops because they give partial information on what the farmer's decisions have been. This, together with historical information and conventional predictor variables should improve predictions. A simulation of Landsat-augmented CARM's, based on USDA statistics over 18 years for the state of Missouri, showed substantial decreases in unexplained variance with the Landsat augmentation, as described later in Section 2.2.3.

Another predictive use of current-season Landsat data takes advantage of the fact that individual fields can be detected and their emergence dates and growth patterns monitored for yield-related information. One clear example is that of double-cropped sobyeans which are planted later than single-cropped soybeans and generally have lower yields. Their acreages should be tabulated and aggregated separately in the estimation process. Other elements of AgRISTARS are investigating the use of Landsat inputs to yield models; these could include derived measures of leaf area or percent cover, condition, and development stage (vs. time and weather) based on peak greenness, rates of greenup and decline, duration, etc. These uses suggest research into questions of preparation and planting practices as distinguishable events in Landsat (and Thematic Mapper) data as well as the use of Landsat to estimate soil type and condition which can affect planting choices and yield-related acreage estimates.

Another use of Landsat would be with models that predict farmers' decisions leading to switches to alternative crops or cultivars as a function of factors such as weather-caused planting delays. For example, in the U.S. Corn Belt, there are dates beyond which each day's delay in planting corn decreases its expected yield substantially. Up to a point shorter-season cultivars of corn could be used. Beyond that point in time it would become prudent to switch from corn to soybeans which are more tolerant of the reduced length of growing season. Landsat could confirm the delayed emergence of crops and predictions could be improved.

2.2.2.4 Using Agronomic Understanding to Enhance the Measurement Value of Landsat Data

The biggest problem in using Landsat data for crop identification and acreage measurement is that of determining the spectral characteristics of the crops of interest and detecting differences from their confusion classes. This is especially difficult under the given constraint that precludes use of local ground truth information. Additionally, early season requirements add more difficulty since Landsat acquisitions are fewer and crops are not all fully developed. Therefore, any way that agronomic understanding of conditions at the local level can be used to improve spectral definition and expectations will be beneficial. This has two aspects, one geographic and one temporal.

Geographically, one observes spectral differences within the crops of interest and in the mix of classes present, as a function of soil type, topography, climate, and other regionally and locally varying agrophysical factors. An objective of any Landsat-based measurement system should be to adapt or "tune" its relevant parameters to local agrophysical conditions at both the segment level and the individual-field level.

Temporally, we have the dominant influence of weather, which can cause substantial differences from year to year in the timing of planting and subsequent operations and in the overall vigor and appearance of the crops throughout the season at a fixed location. Again, adaption to the local, this time weather-related, conditions is highly important. Of course, other factors and episodal events, such as insects, disease, and floods, should similarly be accounted for when they are important.

Another key, longer term temporal factor is the pattern of crop rotations which can be used to establish prior probabilities for crops in individual fields for use in crop identification and classification.

Just as for prediction, we can divide discussion of enhancing the measurement value of Landsat into consideration of previous years' data and of current year's data.

Previous years' data provide a basis for local expectations of spectral signatures, spectral classes, and crop calendars for the various crops, as functions of the conditions encountered during those years. In addition to providing spectral expectations, they might show where flooding is likely to occur and areas where planting operations are more likely to be advanced or retarded from the average due to drainage, topographical influences, or other factors. Year-end crop identifications from previous years can be used to determine crop rotations on a field-by-field basis and used to establish prior probabilities, as previously mentioned. They also would provide information on previously non-cropped areas and fields, which can be excluded from further consideration after appropriate confirmation of no change from past usage. The field pattern from previous years should be a good starting point for use in early season analysis of current-year data.

To investigate early-season uses of multiyear data, we conducted a study of crop rotation patterns in several U.S. Corn Belt segments and found that soybeans seldom followed soybeans in rotation. Agricultural extension agents indicated that this was due both to increased erosion effects with continuously cropped soybeans, where land is not flat, and to increased incidence of certain root diseases. By using last year's field patterns and crop identifications, we found that we could identify crop strata of high crop purity to get an early sample of crop spectral signatures for use in identification and classification. For example, any field that was soybeans the preceding year was very likely to be corn the following year, if it remained a summer crop. Furthermore, it would be a relatively unbiased sample of corn, including both early and late planted fields. This has an advantage over other approaches we examined which used only current-year data and used the fact that corn is usually planted earlier than soybeans, so that the earliest emerging summer crop fields are primarily corn and the latest primarily soybeans. These latter samples are biased.

The preceding is one example to illustrate the use of knowledge of local agricultural practices to improve Landsat measurement accuracy. Other geographic regions would require their own approaches. For example, several weeks separate the usual planting dates for corn and soybeans in Argentina, so simple temporal discrimination between them would be more powerful there than in the U.S. Corn Belt but different confusion classes would exist. Double-cropping with soybeans following wheat is another practice that leads to substantial within-crop diversity and can lead to confusion if not recognized in the segment.

The high-purity crop strata from the multiyear example above also provide an opportunity to gain a good estimate of the local crop calendar for the segment. They can be used to adjust calendars computed with local weather data which have planting date prediction as their greatest source of uncertainty. Even without benefit of the previous year's data, one should be able to use Landsat observations with general knowledge of local cropping practices to improve crop calendar estimates. Another general use of Landsat data would be to search for and flag anamolous conditions in comparison with data from nearby segments or prior years.

For identification and classification with current-year Landsat data, two classes of variables can benefit from agricultural understanding. One is the prior probability of a given crop, which can be based on general information for the region, but more desirably would be field-specific, given prior year data and past rotation history. The second class is the expected temporal-spectral signature of each crop, which includes effects mentioned above, such as crop calendar, crop vigor, weather, and local agrophysical factors. We suggest that, in the long term, a systematic approach for incorporating this type of information would be the joint use of predictive models and Landsat-based measurements for estimation. One would develop a predictive model for each crop signature based on local weather data with perturbation factors to account for field-by-field deviations due to site and seasonal effects. These signature models would be used for classification and identification in the absence of other information and would be updated and refined as more and more spectral observations are obtained in the current year and as a multi-year data base is assembled.

Obviously, management of the required amount and types of information could be very complex and a well defined framework would be required. These issues are addressed further in both Section 2.2.4, where a specific segment-level approach is discussed which could be implemented in a relatively short period of time, and Section 2.2.5, where a longer-lead-time approach involving knowledge engineering techniques is discussed.

2.2.3 LANDSAT AUGMENTATION OF CROP ACREAGE RESPONSE MODELS (CARM)

The research reported in this section has a different emphasis on the use of Landsat than is found in the rest of this report. Rather than being the primary source of data for crop acreage estimation, Landsat is here considered in a new role, one of providing supplemental current-year inputs to an econometric prediction model. This research effort is to be documented more fully in a separate technical report [3].

2.2.3.1 Introduction

Research indicates that a sequence of information, with respect to time, is obtainable from remote sensing, for corn and soybeans acreage estimation. At an early stage, it may be possible to estimate only acreages of gross crop groups, such as summer crops (which would include corn, soybeans, sorghum, and cotton), and at some later date it may be possible to estimate corn and soybeans acreages directly.

An important question arises as to the method of using the early stage, crop group estimates available from remote sensing. A natural candidate is to use these observed crop group acreage estimates as added, current-year inputs into an econometric crop acreage estimation scheme based on the predictive variables of historical and current prices, historical yields, government policy, and historical crop acreages.

This section documents a study of early season Landsat augmentation (via crop group estimates) of a crop acreage response model (CARM) for corn, soybeans and sorghum. The results of the study indicate that accuracy of crop acreage estimation could be significantly increased by Landsat augmentation of sufficient accuracy.

Eventual application in Argentina was of interest, but detailed data were available only for the United States. Therefore, we searched for a state that grows substantial acreages of corn, soybeans, and small grains, as they do in Argentina. It also was desirable that there had been substantial year-to-year changes in the acreages devoted to these crops. The state of Missouri met these criteria.

Crop acreages and other historical information on prices, yields, and government policy was available for the years 1962 through 1979. Since Landsat data were not available for those years, USDA estimates of crop group acreages were used as substitutes for inputs derivable from current-year Landsat data in the analysis.

2.2.3.2 Unique Aspects of This Study

This research was unique for two reasons. The first is that this is one of the first crop acreage estimation models we are aware of that merges the incomplete early season Landsat information with a conventional crop acreage prediction model. There was a similar effort conducted in LACIE in attempting to estimate winter and spring wheat acreages when the extractable information from Landsat was only for the winter and spring-small grains crop groups [4]. Their approach was to estimate the ratio of winter wheat to winter small grains (or spring wheat to spring small grains) using conventional predictive variables (historical and current prices, historical acreages, etc.). This ratio was then multiplied by the winter small grain acreage estimate from Landsat to give a final figure for winter wheat acreage. Our approach is new in that we estimate directly the target crop using both the Landsat estimates of crop group acreages and the normal predictive variables in a conventional type of crop acreage response model.

Secondly, there appears to be few models in the literature developed at a regional level. It is precisely in the regional setting that one can observe the true competitive nature of the crops of interest for acreage (and quantify it). At the national level the differing regional competitive relationships are aggregated and smeared. Thus, in our opinion, it is advantageous for this purpose to develop CARM's at the regional levels where they can model the true competitive relationships and where the Landsat augmentation would be most helpful.

2.2.3.3 Model Specification and Notation

The purpose of the study was to determine the importance of early season Landsat crop group information for crop acreage estimation. Thus two models were compared, one which was a conventional crop acreage response model and the other which was the same model augmented by Landsat inputs. The first model for crop acreage has the form of a regression equation, with a number of independent variables representing expected revenues, last year's acreages, and government policy effects. Both the crop of interest and a competitive or substitute crop are represented. A mathematical representation is as follows:

$$AP_{i,t} = f(C, ExREV_{i,t}, ExREV_{j,t}, AP_{i,t-1}, PV1_{i,t}, PV2_{i,t}) + \varepsilon_t \quad (1)$$

where

$AP_{i,t}$ is the acreas planted to commodity i in year t in thousand acres

C is a constant

$ExREV_{i,t}$ is the expected gross revenue per acre by U.S. farmers for commodity i in year t

$ExREV_{j,t}$ is the expected gross revenue per acre by U.S. farmers for commodity j (substitute commodity which farmers may choose to plant) in year t

$PV1_{i,t}$ is a government policy variable which encourages producers to plant commodity i in year t

$PV2_{i,t}$ is a government policy variable which encourages producers to plant commodity i in year t

ε_t is an error term

The variable $ExREV_{i,t}$ was computed by multiplying last year's price by the average yield per acre over the last three years for crop i.

This is a conventional specification that is used by agricultural economists to explain crop acreage. The origin of this specification and a full discussion of acreage estimation procedures is available (Houck, et al, 1976) [5].

The second specification which includes Landsat augmentation of summer crops and small grains (previously defined) is as follows:

$$AP_{i,t} = f(C, ExREV_{i,t}, ExREV_{j,t}, AP_{i,t-1}, PV1_{i,t}, PV2_{i,t}, APSC_t, APSG_t) + \varepsilon_t \qquad (2)$$

where

$APSC_t$ is acreage planted to summer crops in year t

$APSG_t$ is acreage planted to small grains in year t

It is envisioned that these latter, current-year acreages will be estimated via Landsat. But for the purpose of model development, current-year USDA estimates of summer crops and small grains were used, as was previously stated.

The approach taken for the analysis was as follows:

(a) Assume f is linear.

(b) Determine, by stepwise regression techniques, which explanatory variables to exclude, i.e., the models in (1) and (2) are over specified and certain variables that have insignificant explanatory power should be deleted.

(c) Determine if ε_t are serially correlated. If not, then use ordinary least squares; otherwise, modify the coefficient computation scheme.

(d) Use the coefficient of determination to measure the increase of explanatory power of Model 2 over Model 1.

(e) Determine the increase of prediction accuracy of Model 2 over Model 1.

(f) Determine the level of error which could be incurred on the Landsat estimates of summer crops and small grains before the prediction error of Model 2 degrades to prediction error of Model 1.

Section 2.2.3.4 documents the results of Steps (b)-(d) through normal regression type analysis. Using prediction analysis and simulation, the results of Steps (e) and (f) are documented in Section 2.2.3.5.

2.2.3.4 Regression Analysis and Results

In general, the R^2 value for Model 1 (conventional) were high, ranging from 0.87 to 0.94. The Landsat augmentation (Model 2), nevertheless, made substantial improvements, decreasing the unexplained variances by 17% to 49%.

Corn. After Step (b), the explanatory variables in Model 1 for corn were a constant, the expected revenue of corn, expected revenue of soybeans, and both policy variables. The results for Model 2 were the same, except for the addition of the explanatory Landsat variable of current-year summer-crop acreage. The unexplained variability is decreased by 17% from Model 1 (conventional) to Model 2 (Landsat augmented). The test for serial correlation was not significant at 0.95 level for either Model 1 or Model 2. The results for the regression analysis of corn, along with those for soybeans and sorghum, are listed in Table 2.3 for Steps (b), (c) and (d).

Soybeans. After Step (b), the explanatory variables in Model 1 for soybeans were a constant, the expected revenue of soybeans, expected revenue of corn, and the previous-year's planted acreage for soybeans. For Model 2, the same variables are included, with the addition of both summer-crop and small-grain acreages from Landsat. Unexplained variability is decreased by 49% from Model 1(conventional) to Model 2 (Landsat augmented). The test for serial correlation was not significant for either model. It should be noted that the test for serial correlation used here is a modified Durbin Watson Statistic since this is an autoregressive process [6].

Sorghum. After Step (b), the explanatory variables in Model 1 for sorghum are a constant, the expected revenue for sorghum, expected revenue for wheat, and both government policy variables. For Model 2, the same variables are included with the addition of summer-crop acreage from Landsat. Unexplained variability is decreased by 49% from Model 1 (conventional) to Model 2 (Landsat augmented). The test for serial correlation again was not significant for either model.

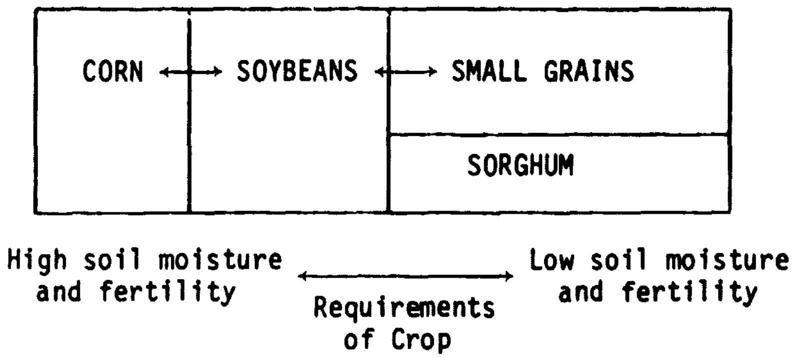
Discussion of the Results. The results of the regression analysis in general are consistent with our agricultural understanding of the crops and agriculture in Missouri. An example of this is evident when comparing the corn and soybeans Model 2 (Landsat-augmented) specifications. The different crops vary significantly in their soil moisture and fertility needs, with corn having the highest requirements followed by soybeans, and lastly wheat and sorghum (which can be combined because of their similar requirements). These varying crop requirements are depicted in the following figure:

TABLE 2.3. REGRESSION ANALYSIS RESULTS

	Models	Variables	R^2	Percentage Decrease In Unexplained Variance
Corn	1	C, $ExREV_i$, $ExREV_j$, PV1, PV2	.8742	17%
	2	C, $ExREV_i$, $ExREV_j$, PV1, PV2, APSC	.8958	
Soybeans	1	C, $ExREV_i$, $ExREV_j$, $AP_{i,t-1}$	.9382	49%
	2	C, $ExREV_i$, $ExREV_j$, $AP_{i,t-1}$, APSC, APSG	.9686	
Sorghum	1	C, $ExREV_i$, $ExREV_j$, PV1, PV2	.9006	49%
	2	C, $ExREV_i$, $ExREV_j$, PV1, PV2, APSC	.9492	

R^2 - Coefficient of variation

Note: Serial correlation of errors was not significant at the 0.95 level in every case.



The corn and soybeans Model 2 specifications both include summer crops acreage which one would expect since corn and soybeans comprise a major portion of the summer crops. But the soybeans Model 2 specifications includes small grain acreages also, which is consistent with the figure in that they compete for the same land. On the other hand, the figure depicts the fact that corn does not compete directly with small grains. Thus, it is appropriate that the soybean model include small grains as a variable and the corn model omit it. Furthermore, the signs of the coefficients of current-year summer-crop and small-grain acreages were consistent with the supportive and competitive nature of these interactions.

2.2.3.5 Prediction Analysis and Results

Prediction errors were analyzed and then a prediction scenario was simulated. The prediction analysis consisted of estimating prediction error via the normal type of analysis for least squares regression. The explanatory variables for prediction error estimation were obtained by averaging over data from 1974-1979. The estimated prediction errors decreased, from Model 1 to Model 2, by 5.1, 22.6, and 23.5 percent for corn, soybeans, and sorghum, respectively. Also included in prediction analysis was a determination of the affects of errors in the Landsat estimates of summer crops and small grains. Specifically, a determination was made of the magnitude of the coefficient of variation that is tolerable before Model 2 prediction would become more inaccurate than Model 1 predictions. The assumptions for the analysis were that normal USDA estimates have coefficient of variation of 0.04 and that Landsat area estimation errors would be independent of regression errors. The results of prediction analysis are given in Table 2.4. The tolerable errors are slightly larger than those assumed for the USDA estimates.

Prediction simulation consisted of simulating an actual prediction scenario, i.e., developing models on data up to year T and predicting for year T+1 given current-year acreage estimates of summer crops and small grains. This was done for the values of T = 1971 through 1978 for both Models 1 and 2. The results, given in Table 2.5, are that the error for the conventional and the Landsat-augmented CARM are about equal for sorghum while the Landsat-augmented CARM is significantly better for soybeans and corn. The results, however, also suggest instability of both Models 1 and 2 when developed over fewer years. Thus, the results also show that one needs a good data base to achieve acceptable accuracies using this regression approach.

2.2.3.6 Discussion of Extension to Argentina

As was stated earlier, Missouri's and Argentina's agricultures have similarities. Specifically they have similar crop mixes, similar meteorological conditions, and both have had recent expansions in soybeans and sorghum. The differences lie in government agricultural policy and agricultural technology. Based on previous work [7], it is believed that international prices and past acreages are the primary explanatory variables able to be incorporated into a conventional CARM specification for Argentina. This specification is the same as the conventional model for soybeans for Missouri for which the added Landsat inputs of current-year summer crops and small-grain acreages dramatically increased the model's explanatory power. It is our belief that this would also occur in Argentina.

TABLE 2.4. PREDICTION ANALYSIS RESULTS

	Models	Relative Error: Prediction Error / Crop Acreage	Absolute Error: Prediction Error* / Agricultural Acreage	Bounds for the Coefficient of Variation of Landsat Estimates of APSC and APSG**
Corn	1	.0482	.0098	---
	2	.0452	.0093	.0458
Soybeans	1	.0634	.0220	---
	2	.0497	.0171	.0442
Sorghum	1	.1385	.0077	---
	2	.1059	.0059	.0497

*This is prediction error relative to total agricultural acreage. Thus since sorghum has smaller share of agricultural acreage, its prediction error is small in relation to its relative error.

**This is bounds on relative error of Landsat estimation of APSG and APSG in order for Model 2 prediction error to be less than that of Model 1.

TABLE 2.5. PREDICTION SIMULATION SUMMARY

	Models	Average Relative Error	Standard Deviation of Relative Error	Maximum Relative Error	Minimum Relative Error
Corn	1	.06	.12	.12	-.26
	2	.02	.12	.19	-.15
Soybeans	1	-.07	.15	-.09	-.35
	2	-.02	.13	.21	-.18
Sorghum	1	.02	.36	.66	-.49
	2	.02	.41	.89	-.52

Another scenario in Argentina which we simulated is the following. We performed the regression analysis for soybeans using only the explanatory variables of last year's soybean acreage, summer-crop acreage, and small-grain acreage. The R^2 of this specification was 0.9473 which is significantly better than the R^2 of 0.9382 for soybeans Model 1 (non-Landsat model). This suggests that in a year in which government policy may be very strong and tending to dampen the effect of prices, it might be better to exclude the pricing variables and only use the three acreage variables for prediction. The results discussed above suggest that this is a possible successful estimation scheme in the face of adverse prediction conditions.

2.2.3.7 Summary

This feasibility study has shown that current-year Landsat estimates of gross crop groups could be of importance in augmenting conventional estimation of crop acreages with acreage response models. It also has shown potential advantages of CARM's developed at the regional level. The approach has been shown to have a fair robustness to errors in the Landsat estimates. We therefore recommend that additional research be directed at exploring the Landsat augmentation of conventional crop acreage response models, including a first look at its potential in a foreign country like Argentina.

2.2.4 THROUGH-THE-SEASON SEGMENT ESTIMATION APPROACH

2.2.4.1 Introduction

The objective of Through-the-Season (TTS) area estimation research is to provide the basis for a technology for estimating target crop acreages at any user-specified time. This technology should be automated, timely, and cost effective. It also should make use of Landsat data as well as pertinent ancillary information such as meteorological data and regional agronomic practices. In this section, we address the generation of segment-level estimates, although multisegment aspects will become important, especially early in the season and also where Landsat coverage is not complete.

Our research on this aspect to date has been limited to the development of an initial approach to segment-level estimation which is presented here, in the form of a flow diagram, along with first-cut details of specific approaches that might be taken at various points in the estimation process. It is consistent with the more general concepts presented elsewhere. The next step in a detailed approach would more fully address the merging of the new concepts with current techniques such as profile classification techniques.

Our concept of a TTS segment-level estimation system is illustrated in the flow diagram of Figure 2.3. Note that we have identified a classification approach in contrast to a direct estimation approach. This was done for the following reasons:

(1) We believe that an augmented classification approach is a viable candidate with several potential advantages:

(a) It more readily permits the incorporation of prior information from a variety of sources, including agronomic and econometric ones.

(b) It has growth potential since refinement of priors can improve a procedure's accuracy from year to year in a multiyear context.

(c) When spectral information is limited or uncertain, emphasis on priors can reduce the possibility of major errors in estimates.

(d) Previous studies of classification techniques with prior probabilities did not use as sophisticated a method for obtaining the priors as we envision and one should be able to reduce or control tendencies toward bias.

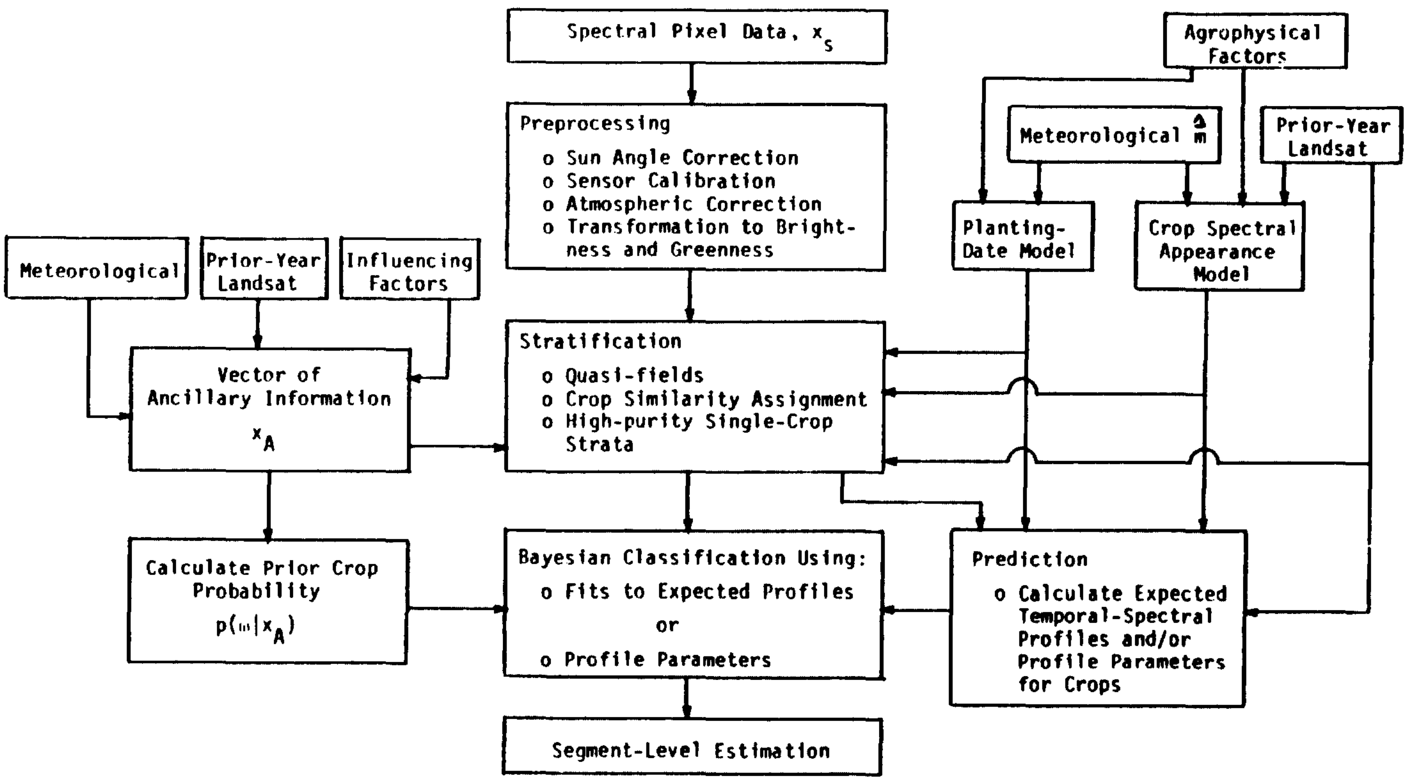


FIGURE 2.3. AN ALTERNATIVE SEGMENT-LEVEL ESTIMATION APPROACH

(2) Direct proportion estimation approaches were receiving extensive attention by other SR researchers at NASA/JSC, so our emphasis provided a vehicle for development and evaluation of an alternative approach.

We now take a more detailed look at the segment-level estimation approach diagrammed in Figure 2.3. The proposed procedure begins with current spectral data $\bar{x}_s$, for each pixel, along with associated ancillary data, $\bar{x}_A$. The spectral data are from the current season's acquisitions available at the time of estimation. The ancillary data could include historical Landsat data, historical crop classifications, historical field (quasi-field) patterns and characteristics, historical crop prices, and quantifications of relevant government policy. The spectral data are first normalized (corrected for haze, sun angle, and sensor calibration) and then transformed to Greenness and Brightness features.

Next, the segment is stratified by spectral/spatial clustering into quasi-fields to approximate true target fields, based on $\bar{x}_s$ and $\bar{x}_A$. In particular, this procedure may initially utilize the previous year's field patterns which could be derived based on the full prior season of spectral data. The quasi-fields are then stratified by assigning each to one or more crop classes that it could belong to, based on spectral zones for $\bar{x}_s$ and on prior year information, including crop rotations. These spectral zones would be determined by planting date models and spectral appearance models, both of which are functions of meteorological parameters and other location-specific information, including prior-year characteristics. This step also is used, where possible, to identify substrata which are known to be of high single-crop purity, based on information such as planting date and crop rotation history. This information feeds the process of estimating expected crop signatures for the segment.

Then classification takes place for quasi-fields assigned to crop groups which contain target crops. Crop temporal-spectral profile models will be used in a Bayesian classification scheme with priors based on the ancillary data. This classification approach is discussed more fully in the next section.

Lastly, the classified quasi-fields are aggregated into segment-level acreage estimates for the target crops.

This approach could be generalized in a multisegment context to take advantage of information from neighboring segments.

2.2.4.2 Detailed Classification Approach

We contemplate using a Bayesian classification approach that incorporates temporal-spectral profiles in order to take full advantage of multidate Landsat data and our understanding of crop phenological differences and growth characteristics.

Two methods of using these profiles are identified here for later exploration and comparison. One method would fit expected crop profile shapes to current-year data values, e.g., along lines developed by researchers at ERIM [8]. This could have an advantage when a full season of data is not available. In a complex implementation, one might compute probabilities by first determining a continuum of expected profiles and tolerance limits with respect to planting date for each crop, based on meteorological conditions, i.e., there could be different shapes for different planting dates. In choosing the best fitting profile one obtains a planting or emergence date in addition to crop type and a quantification of fit or certainty.

The second method would fit a model form to the data and make decisions based on resultant values of the model parameters. One could apply some constraints when sufficient data acquisitions to produce stable fits are not available, to increase the applicability of the model. This method could be a modification of an approach being explored at NASA/JSC [9]. One would need to develop multivariate probability distributions of the parameters.

Mathematically speaking, the first method assumes a model of the form

$$\bar{x}_s(\bar{t}) = \mu_{\bar{x}_s}(\bar{t},\omega,\hat{\bar{m}}) + \bar{\varepsilon}_x(\bar{t}) \tag{3}$$

where

$\bar{x}_s(\bar{t})$ = Vector of observed spectral variables

$\mu_{\bar{x}_s}$ = The expected profile

$\bar{t}$ = Vector of acquisition times

ω = Class

$\hat{\bar{m}}$ = Estimate of meteorological (and other) parameters which help define the expected temporal-spectral profile

$\bar{\varepsilon}_x(\bar{t})$ = Error vector

We assume that we can get estimates of the density of $\bar{\varepsilon}(\bar{t})$ conditioned on the class ω and $\hat{\bar{m}}$. We designate this estimated density by

$$P(\bar{\varepsilon}_x(\bar{t})|\omega,\hat{\bar{m}})$$

We further assume that we have estimates of the prior probability of a class conditioned on the auxiliary information $\bar{x}_A$. Let these priors be designated by

$$P(\omega|\bar{x}_A)$$

Then the class ω is chosen to maximize the posterior probability,

$$P(\omega|\bar{x}_A)\ P(\bar{\varepsilon}_x(\bar{t})|\omega,\hat{\bar{m}})$$

In the second method, we would assume a model of the form

$$\hat{\theta} = \mu_\theta + \bar{\varepsilon}_\theta \tag{4}$$

where μ_θ is the vector of true profile parameters for the target crop and $\hat{\theta}$ is an estimate of the parameter vector for the chosen profile model, as derived from the spectral data $\bar{x}_S$. The error term $\bar{\varepsilon}_\theta$ has an estimated density

$$P(\bar{\varepsilon}_\theta|\hat{\bar{m}},\omega)$$

As before, the class ω is chosen to maximize the posterior probability,

$$P(\omega|\bar{x}_A)\ P(\bar{\varepsilon}_\theta|\hat{\bar{m}},\omega)$$

Of course, the estimation of priors and the error densities for either model will require a substantial effort. The advantage of this classification approach is that prior probabilities, estimated using data other than Landsat, can have a greater influence when Landsat discrimination is uncertain and assume a lesser role when Landsat offers discriminability.

2.2.4.3 Summary

The segment-level estimation scheme described above is one realization of the general concept we developed earlier. It lets agronomically based priors have the major weight until there is enough evidence spectrally to do otherwise. Thus, it merges the functions of prediction and direct observation, as outlined in Section 2.2.1 and in particular in Figure 2.1. Furthermore, the Bayesian classification approach provides for a continuous balancing of information gained from the ancillary and current-year spectral data.

2.2.5 AN ADVANCED APPROACH FOR THROUGH-THE-SEASON ESTIMATION

The preceeding discussion has shown how information from varied ancillary and collateral sources is important for the full extraction and utilization of information from Landsat data. A decision structure is needed that can effectively utilize data from disparate data sources having differing degrees of information content, accuracy, and precision. Furthermore, we believe that this structure should be flexible and adaptable, should allow for both machine-derived and human inputs, should maximize the efficiency of the human resource, and should be able to "learn" or build a knowledge base as it continues in use.

We have studied the opportunities for artificial intelligence, specifically knowledge engineering systems, and believe that they would serve as the desired vehicle for TTS decision making and utilization of remotely sensed data.

Figure 2.4 is an elaboration of the general TTS estimation diagram presented earlier in Figure 2.2. It presents the various elements in a form that would be amendable to implementation through a knowledge-engineering or rule-based inference approach. In such an approach, a knowledge base and inference structure are built so that, as each new fact or observation is introduced, a particular inference will become more certain. The chain of inferences leading to particular decisions can be based on the knowledge and experience of expert interpreters, analysts, and agronomists. These systems were first developed for medical applications.

A candidate prototype for the desired system is found in the PROSPECTOR system [10]. It differs from its predecessors, the EMYCIN and MYCIN systems [11], in that it uses Bayesian methods of estimation whereas the others use a more empirical, yet axiomatic approach. Both have provisions to grow and "learn" and incorporate new facts and data as they become available. Prospector was developed to help locate optimal drilling sites in prospecting for ore bodies for mining.

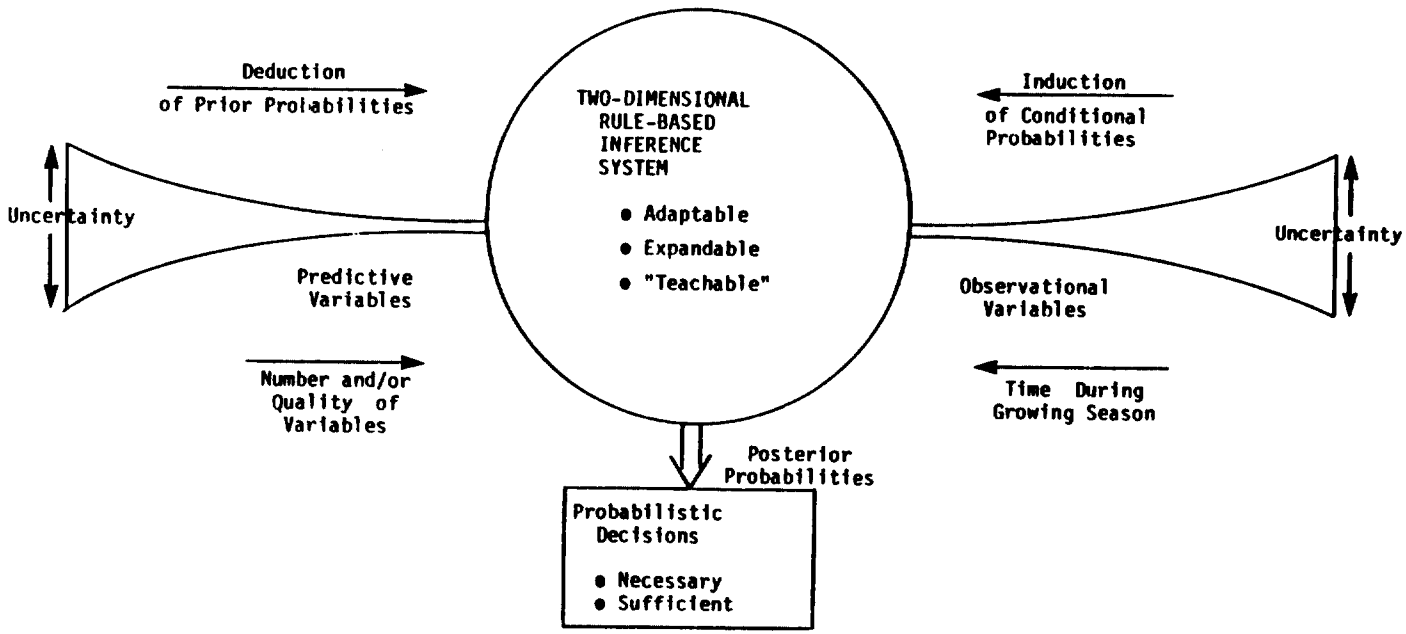


FIGURE 2.4. A CHARACTERIZATION OF TTS ESTIMATION THAT IS AMENABLE TO KNOWLEDGE-ENGINEERING IMPLEMENTATION

Although it will take some time to fully develop the knowledge engineering approach to TTS estimation, we recommend it as a desirable pursuit with a potentially large payoff in accuracy, efficiency, and automation.

2.2.6 SUMMARY OF THROUGH-THE-SEASON ESTIMATION RESEARCH

In conclusion, we summarize the main ideas and concepts that have been developed and expressed in this section. They are:

(1) The crop estimation process was characterized as being a time-varying combination of prediction and measurement (observation) processes through-the-season (TTS), with the balance swinging from prediction to measurement as time progresses through the growing season. It was shown how Landsat can contribute to both processes.

(2) Value was shown for merging traditional prediction variables (prices, government policy, etc.) with early season Landsat observation of the farmers' actions (gross crop group acreages) to produce improved early estimates of specific summer crop acreages. Quantitative results were presented for a simulation study based on historical USDA statistics for an entire state. Furthermore, potential was shown for models based on regional rather than the usual national levels.

(3) Field-by-field Landsat observations are seen as the appropriate and optimal basis for use in TTS estimation. It is by observing fields on multiyear basis that one can best interpret current-year Landsat observations of farmers' actions for crop acreage estimation.

(4) Predictive models of crop spectral appearance, which taken into account local weather and other factors, would be most beneficial for interpreting Landsat observations and maximizing the amount of measurement information extracted from them.

(5) Agricultural practices were identified which are observable by Landsat and could be of high interpretive value in TTS estimation.

These include the timing of field preparation, irrigation, predecessor crops and time of spectral emergence (related to planting date).

(6) Multiyear use of Landsat was shown to be important for establishing the expected crop spectral signature for a given area under a variety of conditions. Also, the interpretive keys discussed in (5) would be more readily and accurately used with a multiyear Landsat data base.

(7) A segment-level Bayesian estimation approach was presented for merging prior probabilities based on ancillary (predictive) variables with direct crop Landsat observation at the field level. The priors are based on predictive variables and indirect (prior year) Landsat observations. The current-season Landsat observations are used to produce direct spectral-based probabilities. An important property of this approach in early season is that the predictive priors can dominate the classification when direct observation by Landsat is of little value. As the season progresses and direct observations by Landsat are of much greater value, the current-season spectral-based probability dominates the classification. Thus we have a scheme which shifts in a continuous fashion from predictive acreage in early season to observed acreage in later season.

(8) For long-range development, we recommend investigation of knowledge engineering systems tailored to the TTS estimation problem. They seem well suited to handling the varied information sources available and have a potentially large payoff.

2.3 MULTISEGMENT ESTIMATION RESEARCH

2.3.1 BACKGROUND AND INTRODUCTION

2.3.1.1 LACIE

The bulk of the current Landsat-based crop inventory methods used in AgRISTARS are based on the multistage sampling techniques developed during LACIE. If one wished to estimate the proportion of a crop of interest within a given region with today's technology then one would go through the following steps:

(1) Partition the region into strata in such a way that the crop proportions varied little within a stratum yet these strata would still be large enough to allocate samples for the steps given below. APU's (agrophysical units) and CRD's (crop reporting districts) are examples of such stratifications.

(2) Partition the region of interest into 5x6-mile segments for data base purposes. We will simplify this discussion by assuming that this segmentation represents a refinement of the stratification defined above. The segments which survive cloud screening are the sample units.

(3) Choose a random sample of segments from each stratum. During LACIE this sample tended to represent about 1% to 2% of the total area. (This can be viewed as the stage-one sample.)

(4) Obtain an estimate of the proportion for each segment in the sample, based on a second stage of sampling. Two of the methods are:

(4a) Procedure 1 (Developed by NASA/JSC) [12]. Choose a deterministic sample of 60 to 100 pixels from the segment as the stage two sample units. The elements of this sample are called dots. These dots are divided into type one and type two dots.

Type one dots include only pixels deemed to be "pure" (single crop) by an analyst interpreter, whereas type two dots may be either pure or mixed.

The analyst labels each dot as crop of interest, 1, or not crop of interest, 0. A classifier is trained on the type one dot labels and then assigns labels to every pixel in the segment, including the type two dots. The labels of the type two dots are used to estimate the performance matrix of the classifier. This estimated performance matrix is then used to debias the mean of the classifier's labels.

(4b) Procedure M (developed by ERIM) [13]. The pixels within each sample segment are clustered using spatial and spectral variables into field-like patterns called blobs. These blobs are the stage-two sample units.

The blobs within a segment are clustered again using spectral/temporal variables. The resulting clusters were treated as strata for the stage-two sample. The Midzuno sampling technique is used to select blobs for labeling, because the blobs vary in size. About 100 blobs are sampled and labeled. The weighted proportion of the blob labels within a cluster gives the cluster proportion estimate. The weighted mean of the cluster estimates then gives the segment estimate.

(5) The sample segment proportion estimates are aggregated into stratum estimates and an overall region estimate in the normal manner.

2.3.1.2 AgRISTARS

Post LACIE research has been conducted in several areas, these include:

(1) Advanced Labeling Techniques. In LACIE about 50% of the standard deviation in the segment grain estimates and all of the bias in the estimates were due to labeling errors. There have been improvements but this component is still a major source of errors and cost. Labeling is being made more objective and hence more automatable.

(2) Multiyear Estimation. Procedures which take advantage of year-to-year correlation to improve sampling efficiency have been

developed. The level at which the multiyear procedures should be implemented at is not clear at this time.

(3) Through-the-Season Estimation. The most used procedures require acquisitions throughout most of the growing season. Procedures which give estimates throughout the growing season, especially early and midseason, are in the development stage. This topic is the subject of another section of this report (Section 2.2).

(4) Profile Based Techniques. Profiles are parameterized functions which map a day of year, t, into Greenness (and sometimes Brightness or other spectral variables) based on observations of crops. Because profiles allow the comparison of crops in segments which have different acquisition histories, profiles will most likely play a major role in multisegment estimation. The drawback of the current profile techniques is that at least three acquisitions are required in order to fit a good profile. The number of acquisitions required could be reduced if constraints were added on the parameter space such as a linear relationship within a subset of the parameters, or in a multistage procedure in which one set of parameters are estimated and then the remaining are fitted.

(5) Multisegment Estimation. In multisegment estimation the overall objectives are to increase sampling efficiency and reduce measurement cost without sacrificing accuracy. Sampling efficiency can be increased by reducing the segment size and increasing the number of segments. Sampling is discussed in Section 2.3.2. Measurement cost reductions might be gained by processing several segments together and/or by processing a few intensely and a larger number with a more economical but less accurate procedure. Though reduction in the scope and funding of our efforts precluded carrying out the research to fruition, we considered three methods of measurement. First signature extension is conceptually described in Section 2.3.3. In signature extension, labels measured from a few segments would be geographically extended to other segments thereby reducing measurement cost by eliminating the need to extract training from all segments. Secondly, regression methods are

discussed in Section 2.3.4. Such methods extend relationships between economically derived estimates and intensive estimates thereby achieving a higher level of accuracy at a reduced cost. Finally, the bin method is described in Section 2.3.5. Sufficient resources were available to evaluate this multisegment measurement scenario experimentally and it is so reported. The bin method extends the decomposition of the spectral distribution from a training sample to the entire segment. Due to the robustness of the method, a reduced training sample is required thereby achieving a cost reduction. In addition, judicious selection of features would enable the use of the bin method within a signature extension scenario.

2.3.2 MULTISEGMENT SAMPLING

2.3.2.1 Effect of Sample Size

Sample variance is known to increase as the segment size increases, assuming the product of the segment area and the sample size remains constant. Perry [14] showed that this effect could be approximated $V(x) = \alpha x^{\beta}$ where x is the segment size. LARS and UCB estimated, empirically, that the LACIE sampling efficiency was about 1/8 compared to simple random sampling. The choice of cluster (or two stage) sampling was made in LACIE for valid cost, data base, and measurement considerations. However, the present 5x6-mile segment size was just a first try. The increases in computer power per unit cost and advances in registration technology relax data base considerations and it appears that the segment size could be reduced significantly with very small impact on the measurement procedure. We developed plans with UCB to test a segment size of 64x64 pixels, extracted from full-frame Landsat data sets.

2.3.2.2 Sampling Vs. Segment Selection for Training

The optimal method of selecting segments depends on the estimator which is being used. Random sampling schemes are required in some procedures such as the regression method and Procedure M. When using the sample to train a classifier, it is more important to represent all of the major spectral classes in the region, randomizing only after these constraints are met. One multisegment procedure we postulated and planned to test is based on a profile classifier. The parameters of the profiles would be estimated for every pixel in every segment (or a large sample) in the stratum. The classification would take place in the parameter space. The problem is how to choose the sample which will best train the classifier.

The IBM Procedure-2 [15] experiment used a technique which first clustered the pixels (CLASSY) across segments and then used a factor-analysis-like technique for segment selection. Earlier ERIM Procedure-B experiments [16,17] also clustered targets (blobs) across segments using spectral/temporal variables. But the method of segment selection differed from that used by IBM. ERIM employed a pairwise selection procedure which chose the two remaining segments which best represented the major undersampled clusters. The pairwise selection continued until the sample budget was exhausted. These two segment allocations gave about the same results.

In the profile-based multisegment procedure, the profile parameters will form the feature space. The pixels will be clustered based on these parameters and the segments selected using either IBM's factor-loading or ERIM's pairwise-loading technique. Labels obtained for targets in the sample segments will be used to train the classifier. The classifier will be applied to every pixel of every segment with sufficient acquisition history.

Early multisegment experiments will use ground truth labels or will modify existing measurement technqiues. Later research will optimize measurement techniques in a multisegment environment.

2.3.3 SIGNATURE EXTENSION

2.3.3.1 Notation and Signature Extension Assumption

$R_{...}$ is the region of interest

$R_{i..}$ is the stratum i

$R_{ij.}$ is the j^{th} signature extension stratum of stratum i

R_{ijk} is segment K

$P_{...}$, $P_{i..}$, $P_{ij.}$, and P_{ijk} are the corresponding crop proportions

We assume the region $R_{...}$ is partitioned into clusters based on spectral/temporal attributes of the labeling targets. Denote these clusters as $\{S_\alpha\}$ so that $R_{...} = \bigcup_\alpha S_\alpha$. Let $Q_{ijk\alpha} = R_{ijk} \cap S_\alpha$ and $q_{ijk\alpha}$ as the corresponding crop proportion.

The signature extension assumption is that the distribution of the random variable $q_{ijk\alpha}$ is independent of k. (This assumption can be relaxed somewhat.) This assumption implies that all of the segments within $R_{ij.}$ can be processed using the same decision logic, and that a classifier which has been trained on a subset of segments which represents the S_α's within $R_{ij.}$ can be used to classify all of the targets within $R_{ij.}$.

2.3.3.2 Signature Extension Region

The signature extension experiment, described in [16,17], trained and applied a classifier across the state of Kansas. This was too large of a region to apply any one decision rule. There were Greenness/Brightness/Temporal signatures which represented pure grain on one side of the state and pure non-grain on the other. These Kansas signature extension experiments indicated that there are at least four signature extension regions in Kansas. A different decision rule is generally needed for each signature extension region.

Signature extension regions have to be small enough for the assumption to hold and have to be large enough to allow a large enough sample to develop a decision rule (train a classifier).

Research has been conducted in this area by UCB under the Dynamic Stratification Task.

2.3.4 MULTISEGMENT REGRESSION METHODS

2.3.4.1 General Regression Methods

We assume that there are two random variables X and Y with the following linear relationship:

$$(Y - u_y) = B(X - u_x) + e$$

where e is a random variable with mean zero. Two samples are taken. In the first sample, we observe $(X_i)_{i=1}^{n'}$ (i.i.d. X) and in the second we observe $(X_i,Y_i)_{i=1}^{n}$. Cochran [18] gives the estimate for u_y as:

$$\hat{u}_y = \overline{Y} + b(\overline{X} - \overline{X}')$$

where $\overline{X}'$ and $\overline{X}$ are the means of the first and second samples, respectively, and b is the least squares estimate for B, based on the second sample. This estimate is conditionally biased, i.e.,

$$E(\hat{u}_y - u_y | X') = B(X' - u_x) \; .$$

In most applications $n \ll n'$ because each Y observation is much more expensive than each X observation.

Cochran's figure 12.1 [18] gives a useful chart for comparing a one-phase simple random sample and a two-phase regression estimator.

2.3.4.2 A Multisegment Regression Procedure

The Baseline Corn/Soybean Procedure described in Section 3 is a two-phase procedure. The procedure provides two levels of corn/soybeans estimates. The first, called the stage one estimate is a nearly automatic procedure while the second is a more intensive and more accurate procedure. The stage-two estimator requires twice the computer time and five times as much analyst time as the stage-one estimator. This suggests that regression estimation methods might provide a lower varianced estimator for a fixed cost.

Let Y denote the stage-two estimator and X denote the stage-one estimator. Because of the nature of the Baseline Corn/Soybean Procedure, a stage-one estimate is obtained automatically for every stage-two estimate. This implies that $n' = 0$ is not an option. Hence the Baseline could be viewed as a special case of a regression estimator where $n = n'$.

An ITD experiment was carried out in order to determine if it would be cost effective to have a large number of stage-one estimates and, for a smaller subsample, to also have stage-two estimates. This experiment is reported in detail in Section 3.3.3. The experiment indicated that variance could be reduced by 25% to 50%, for fixed cost, by the use of regression estimates. This application of regression methods of estimation is more general than that discussed in Section 2.3.4.1 in the following ways:

(a) The quantity to be estimated is multivariate, i.e., the acreage of two or more crops (in this case, corn and soybeans) simultaneously.

(b) The cost constraints are more general, consisting of two or more linear constraints imposing limitations on several resources (analyst and computers).

2.3.5 AN EXPERIMENT USING THE BIN METHOD FOR SEGMENT PROPORTION ESTIMATION

The bin method is a direct proportion estimation scheme which has been researched in the past by JSC and for which there is current interest for use as an early season proportion estimator. We ran an experiment using the bin method in order to increase our understanding of it in a real life estimation scenario and to establish its applicability as a signature extension scheme for multisegment estimation.

For the experiment we had spectral data for 17 segments of which ten had been processed through the ITD Baseline Corn and Soybean Procedure for proportion estimation based on sampling and classification (Section 3). For purposes of understanding, all segments were processed as follows; targets (targets will be defined later on) were sampled (different sampling rates were tried), assigned their ground truth labels, and used as training data for the bin method. For the purpose of testing an alternative proportion estimation segments were run through the bin method using the sampled and labeled targets as training data. The purpose of this section is to outline the results of the experiment and understandings gained.

2.3.5.1 The Bin Method

The bin method is a direct crop-proportion estimation scheme that can use spectral data from several satellite passes. The basic idea is to divide the multitemporal spectral space into regions or bins and, based on the overall dispersion of the data across these bins, to determine the proportions of categories of interest. Specifically, the total joint density across the bins, denoted by f, is computed from the spectral data. If one also has $f(x \mid corn)$, $f(x \mid soy)$ and $f(x \mid other)$ then regression methods can be used to solve the model:

$$f(x) = cf(x \mid corn) + sf(x \mid soy) + of(x \mid other) + e,$$

where

c denotes the proportion corn;
s denotes the proportion soybeans; and
o denotes the proportion other.

If one has consistent estimates for $f(x \mid corn)$, $f(x \mid soy)$, and $f(x \mid other)$, then regression methods will give consistent estimates of c, s, and o. If the estimates are biased (may still be consistent), then the complete procedure will give slightly biased estimates for c, s, and o, however. But the results of the experiment give evidence that the bias is quite small. A slight problem is that the procedure does not restrict its estimates to the three-dimensional simplex; and indeed this experiment gave estimates above one and below zero. We will discuss this in more detail later on.

2.3.5.2 The Bins

The bins were derived by establishing thresholds on Greenness values measured for scene targets on three different dates for each segment. The targets were quasi-fields generated during other processings by an ERIM spatial-spectral clustering (or blobbing) algorithm.

Labeling Greenness measured on Day i as g_i, where i=1,2,3, two thresholds, t_{i1} and t_{i2}, were determined for each day. Then for every quasi-field there was a mapping

$$h: \; R^3 \rightarrow \{1,2,3\}^3$$

where h is defined as

$$h(g_1,g_2,g_3) = (b_1,b_2,b_3)$$

where

$$\begin{aligned} b_i &= 1 \quad \text{if} \quad g_i < t_{i1} \\ &= 2 \quad \text{if} \quad t_{i1} \le g_i \le t_{i2} \\ &= 3 \quad \text{if} \quad t_{i2} < g_i \end{aligned}$$

Thus, the mapping h defines 27 spectral bins and these bins were determined for every segment by setting the six threshold levels based on expected crop spectral responses on the given days of year. A listing of the Julian days with corresponding thresholds is given in Table 2.6. For seven segments, a supervised mode of blobbing was used in which the clusters were restricted to include only pixels of like ground truth. The other ten segments were run through the Baseline C/S Procedure.

The basis for the choice of acquisitions and thresholds was the logic used by the Baseline C/S Procedure in stratifying for summer crops and in separation of corn and soybeans. This led to selection of early and late acquisitions, which gave substantial separability between summer crops (corn and soybeans) and other crops, and a middle date where there appeared to be maximum separability between corn and soybeans in Greenness space.

2.3.5.3 Methods of Estimating $f(x \mid \text{corn})$, $f(x \mid \text{soy})$, and $f(x \mid \text{other})$

This experiment estimated the above conditional densities by training on a random sample of the data in each segment. The random sample was labeled with "ground truth" for Segments 107, 127, 809, 844, 854, 866 and 891. The sampling rates, denoted Q, were .05, .10, .15, .20 and .25. Baseline corn-soybean labels were used to obtain bin estimates for Segments 141, 202, 205, 800, 832, 842, 852, 853, 877 and 881.

TABLE 2.6. BIN DESCRIPTIONS

Method of Quasi-Field Generation	Segment	Greenness Thresholds and Acquisition Dates					
		Day 1	(t_{11},t_{12})	Day 2	(t_{21},t_{22})	Day 3	(t_{31},t_{32})
	107	208	(15,25)	226	(25,30)	271	(10,15)
	127	152	(5,10)	243	(20,25)	306	(5,10)
	809	164	(5,10)	244	(20,25)	308	(5,10)
Supervised	844	161	(5,10)	233	(25,30)	306	(5,10)
	854	161	(5,10)	234	(25,30)	306	(5,10)
	886	167	(5,10)	249	(25,30)	293	(5,10)
	891	168	(5,10)	249	(20,25)	267	(15,20)
	141	166	(5,10)	256	(15,20)	302	(5,10)
	202	167	(5,10)	221	(25,30)	302	(5,10)
	205	155	(5,10)	246	(25,30)	308	(5,10)
	800	164	(5,10)	246	(25,30)	300	(5,10)
	832	160	(5,10)	232	(25,30)	304	(5,10)
Unsupervised	842	160	(5,10)	232	(25,30)	304	(5,10)
	852	160	(5,10)	232	(25,30)	304	(5,10)
	853	160	(5,10)	232	(25,30)	304	(5,10)
	877	150	(5,10)	231	(25,30)	267	(15,20)
	881	159	(5,10)	231	(25,30)	303	(5,10)

Note: The year was 1978 for all segments.

Since the bin method sometimes gave estimates outside of the three-dimensional simplex, negative estimates were replaced by 0 and then the other estimates were normalized to add to one. The next section gives the results of this experiment.

2.3.5.4 Results

Figure 2.5 through 2.7 display true vs. estimated crop proportions determined for the seven segments for which supervised quasi-fields were available, for sampling frequencies Q = 0.05, 0.15 and 0.25, respectively. Each point represents the mean of 100 estimates produced from bin proportions generated by using the different training samples. Each figure has seven estimates each for corn, soybeans and other, except Figure 2.5 which is missing values for Segment 809. The bin method gave unbiased estimates for all of the sampling frequencies, on the average, for this source of labels.

It was expected that the standard deviation would depend on sample size in somewhat the same way as that of a simple random sample, namely proportional to the inverse of the square root of the sample size. Because the number of targets varied from segment to segment, the sample size also varied from segment to segment. Figure 2.8 gives the standard deviation vs. sample size least squares response function for corn and soybeans, where the response function is assumed to be of the form:

$$s = c/\sqrt{n}$$

where c is to be estimated by standard linear regression. The standard deviation drops rapidly from 10 to 50 samples after which the decrease slows significantly.

For the second part of the experiment, we analyzed about 100 targets per segment which were given analyst labels in an early test of the Baseline Corn/Soybean Procedure. The segments were: 141, 202, 205, 800, 832, 842, 852, 853, 877 and 891. These analyst labels are called

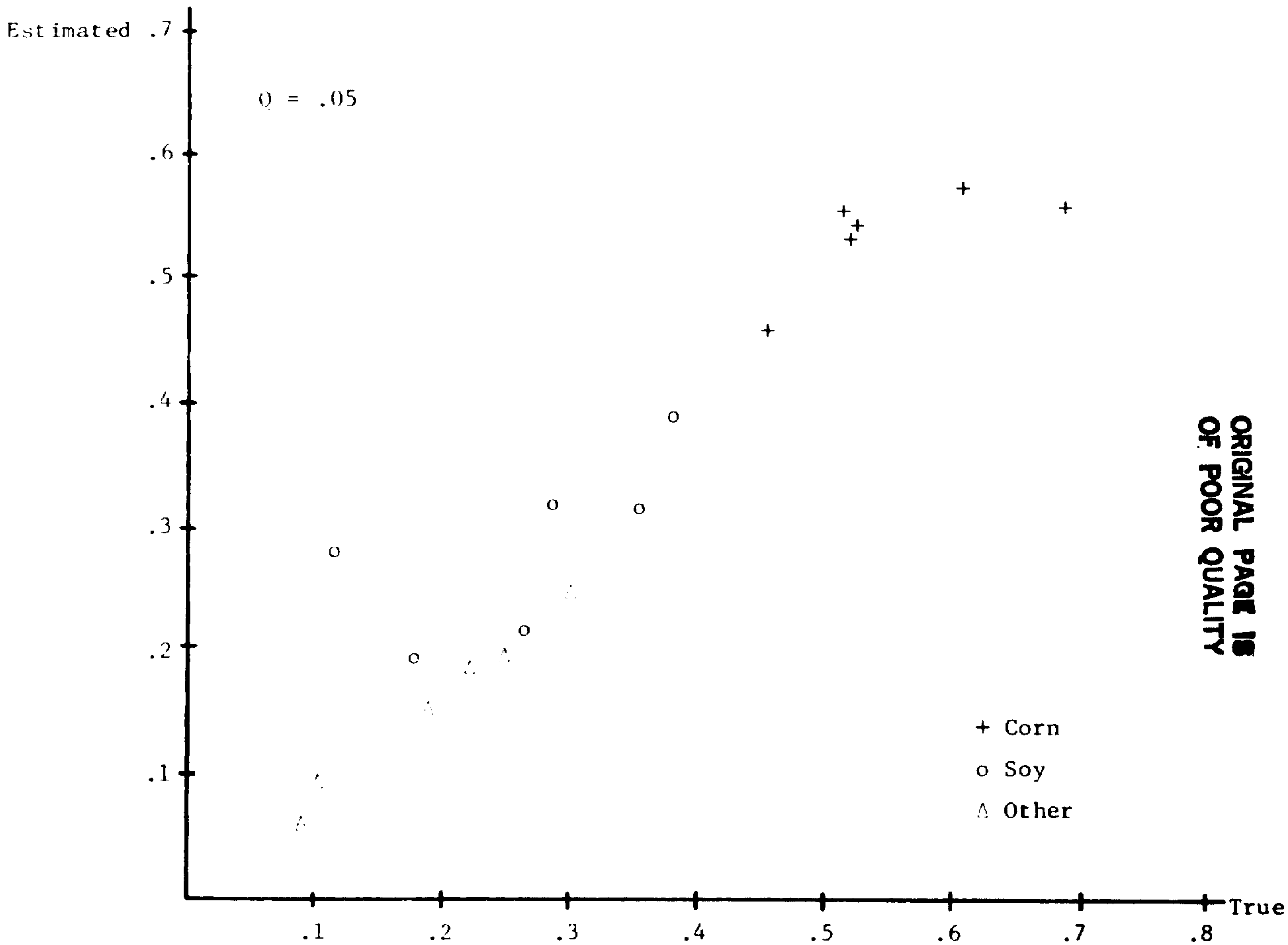


FIGURE 2.5. AVERAGE ESTIMATED PROPORTION VS. TRUE PROPORTION

ORIGINAL PAGE IS
OF POOR QUALITY

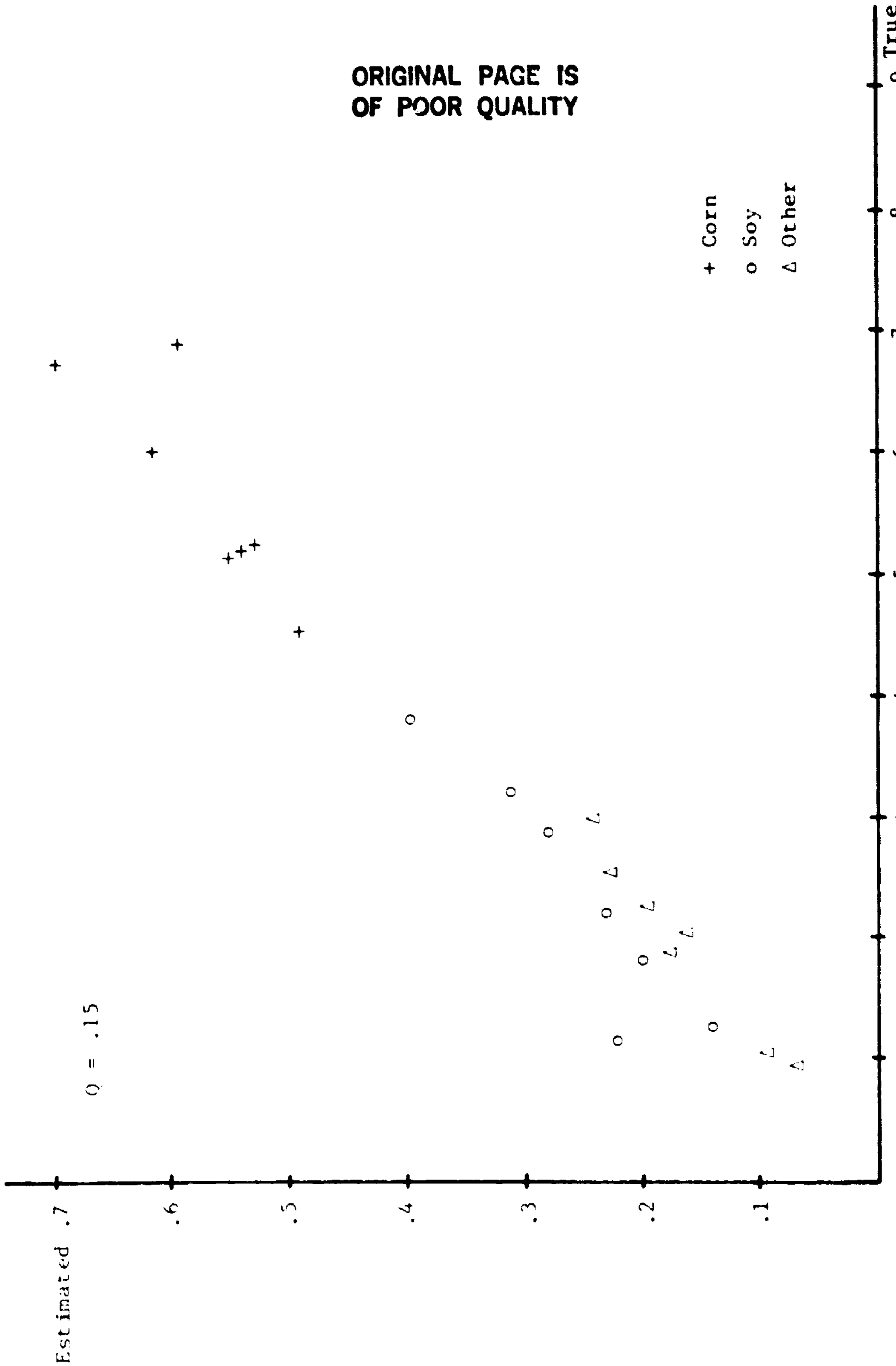


FIGURE 2.6. AVERAGE ESTIMATED PROPORTION VS. TRUE PROPORTION

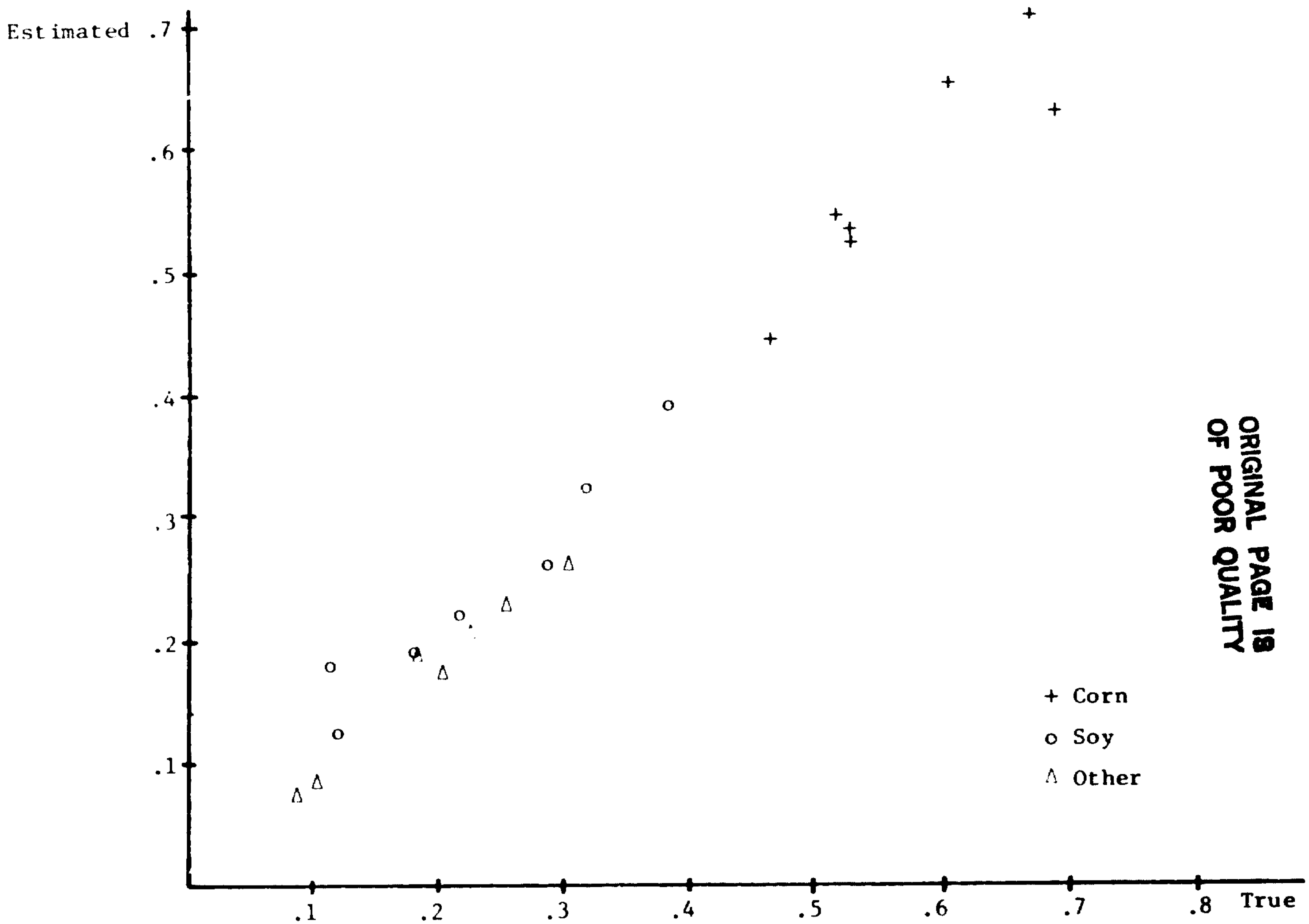


FIGURE 2.7. AVERAGE ESTIMATED PROPORTION VS. TRUE PROPORTION

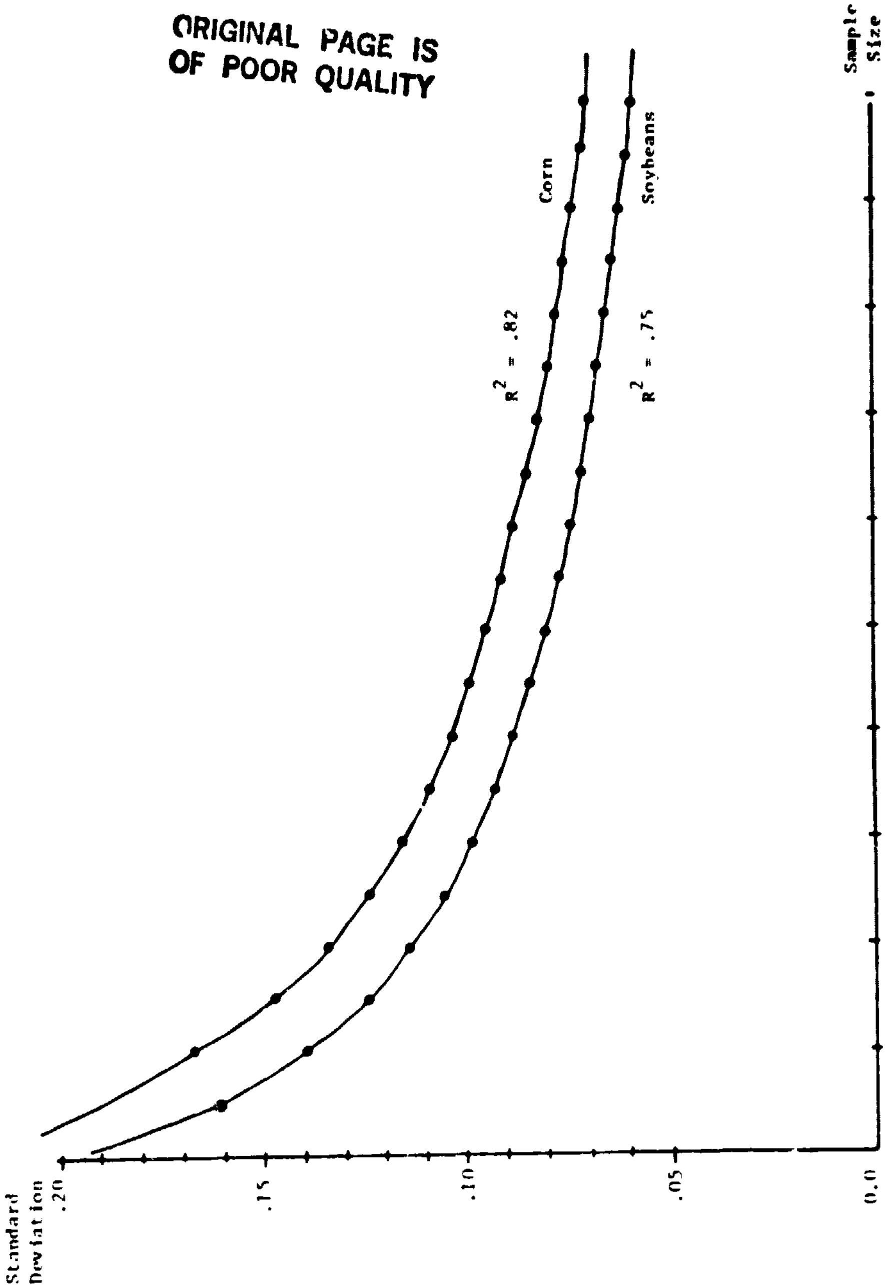


FIGURE 2.8. STANDARD DEVIATION OF ESTIMATED PROPORTION VS. SAMPLE SIZE

Stage 2 labels. The Baseline Corn/Soybean Procedure also gives semi-automatic labels, called Stage 1 labels, for every potential target. The samples which were used in the above baseline procedure test were used to estimate $f(x \mid corn)$, $f(x \mid soy)$ and $f(x \mid other)$ for these ten segments. Table 2.7 gives four values for each segment for each crop class. These were based on the ground truth, the mean of Stage 1 labels, bin estimates using Stage 1 labels and bin estimates using Stage 2 labels, respectively. Averages across all segments, standard deviations, and biases are also given.

The mean of the Stage 1 corn labels gives an unbiased estimate while the Stage 1 and Stage 2 bin methods give 6% and 3% bias, respectively. The mean of the Stage 2 soybeans labels gives a -10% bias while the bin method using Stage 1 labels gives only -6% bias.

2.3.5.5 Conclusions

The choice of the thresholds for the bins as outlined in Section 2.3.5.2 was made using prior knowledge of the distributions of Greenness for corn, soybeans and other. In an operational system, these thresholds would need to be based on one, two or three of the following:

- Historical Landsat data and ancillary data
- Histogram of all the pixels/blobs Greenness
- Identifiable subpopulations of specific crops

Intuitively it is appealing to choose bins which maximize the difference between probabilities of two cover types of being in each of these bins. The results are supportive of this in general but it seems that late in the season, the bins did not pick up much separation of crops. It also appears that not many quasi-fields had $b_i = 2$ for $i = 1,2,3$. Thus this may need to be looked at in the future also. The results of experiment on the BASELINE segments indicates that the bin method is a fairly unbiased way to use the labeled targets.

TABLE 2.7. BIN METHOD ESTIMATES FOR TEN SEGMENTS

	Corn				Soybeans				Other			
Segment	P	P_1	$\hat{P}_1$	$\hat{P}_2$	P	P_1	$\hat{P}_1$	$\hat{P}_2$	P	P_1	$\hat{P}_1$	$\hat{P}_2$
141	.233	.289	.338	.300	.182	.112	.140	.141	.578	.598	.520	.558
202	.244	.212	.267	.274	.326	.237	.247	.220	.428	.550	.484	.504
205	.188	.301	.426	.334	.540	.423	.436	.398	.271	.274	.137	.257
800	.603	.608	.681	.598	.259	.178	.206	.269	.136	.213	.111	.132
832	.244	.345	.381	.348	.393	.236	.243	.240	.361	.417	.374	.411
842	.487	.381	.428	.450	.277	.197	.241	.257	.235	.416	.330	.292
852	.293	.245	.306	.260	.301	.160	.241	.227	.405	.593	.451	.511
853	.383	.322	.387	.373	.292	.160	.244	.240	.323	.517	.367	.385
877	.058	.521	.593	.501	.222	.141	.186	.206	.268	.336	.219	.231
881	.474	.437	.458	.455	.068	.049	.048	.050	.457	.513	.493	.494
Ave	.366	.366	.427	.395	.286	.189	.223	.225	.346	.443	.349	.379
SD	0	.071	.086	.061	0	.040	.038	.055	0	.070	.068	.046
Bias	0	-.001	.060	.029	0	-.096	-.062	-.061	0	.096	.002	.032

P_1 = ground truth proportion

P_1 = bin method proportion estimate using ground truth labels

$\hat{P}_1$ = bin method proportion estimate using Stage 1 labels from BASELINE

$\hat{P}_2$ = bin method proportion estimate using Stage 2 labels from BASELINE

Our recommendation is that research be conducted to determine: (1) the effects of the choice of bins and (2) the optimal estimation scheme when the bin method gives proportion estimates greater than one or less than zero. Use of labeled targets as training data also should be explored further because of the relative unbiasedness of the results.

2.4 ARGENTINA-BRAZIL AGRONOMIC UNDERSTANDING

The principal reason for establishing this subtask was to help ensure an orderly transition from a U.S.-based technology development for corn and soybeans area estimation to one adaptable to foreign areas (Argentina and Brazil). As such, the subtask was designed to anticipate and/or respond to data and information needs so that techniques designed and developed primarily with U.S. data can be adapted to handle expected agronomic conditions found in Argentina and Brazil. This requires the collection, organization and summarization of a wide variety of information relating to country specific agricultural crop types, crop-livestock practices, the location and extent of agricultural regions, soils and climatic data and other factors that characterize the agricultural systems operating in Argentina and Brazil. Another critical aspect was the collection of ground information on crop types in segments in these countries for which Landsat data are being acquired. Initial emphasis was placed on Argentina due to its greater similarity to U.S. regions.

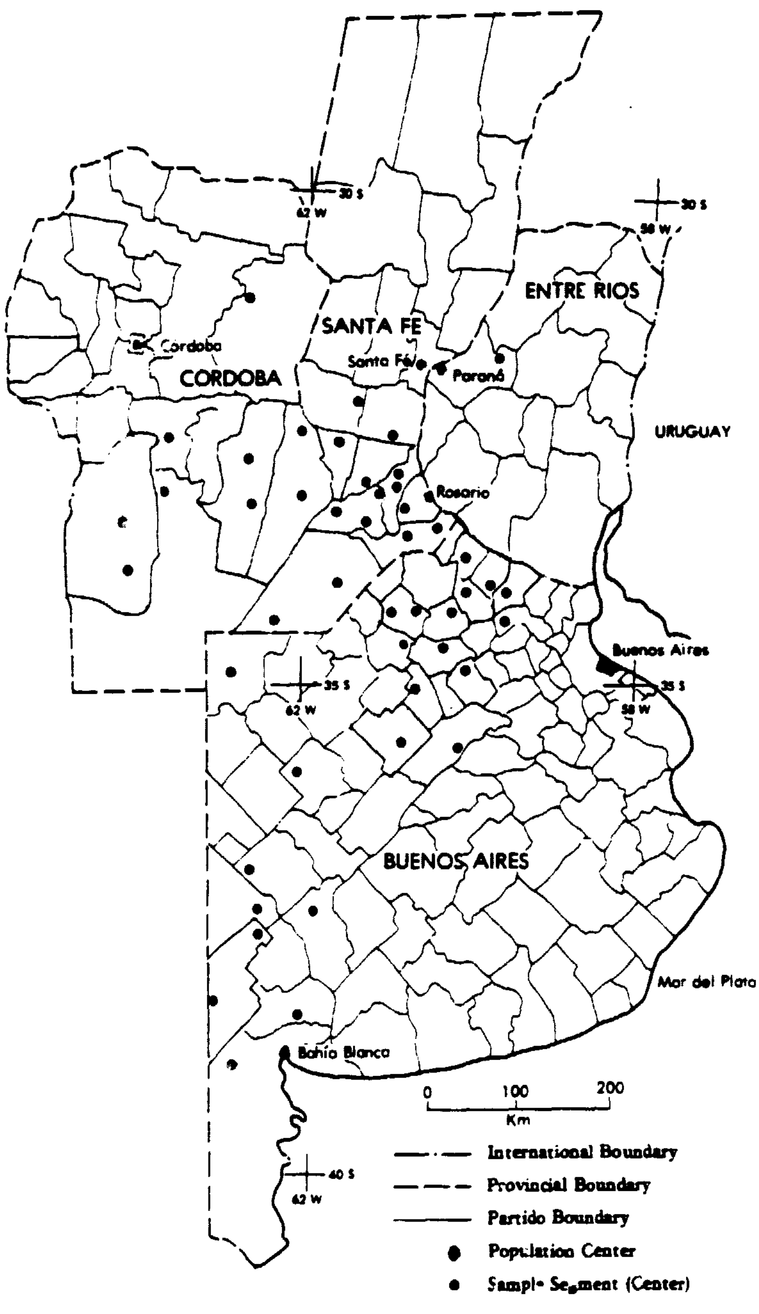
2.4.1 DESCRIPTION OF AGRICULTURE IN ARGENTINA

A separate technical report [19] has been written to give a detailed presentation of the information and understanding we gained about agriculture in Argentina. Related reports include References [20] to [24]. This section presents a summary and overview of that report.

2.4.1.1 Study Area Defined

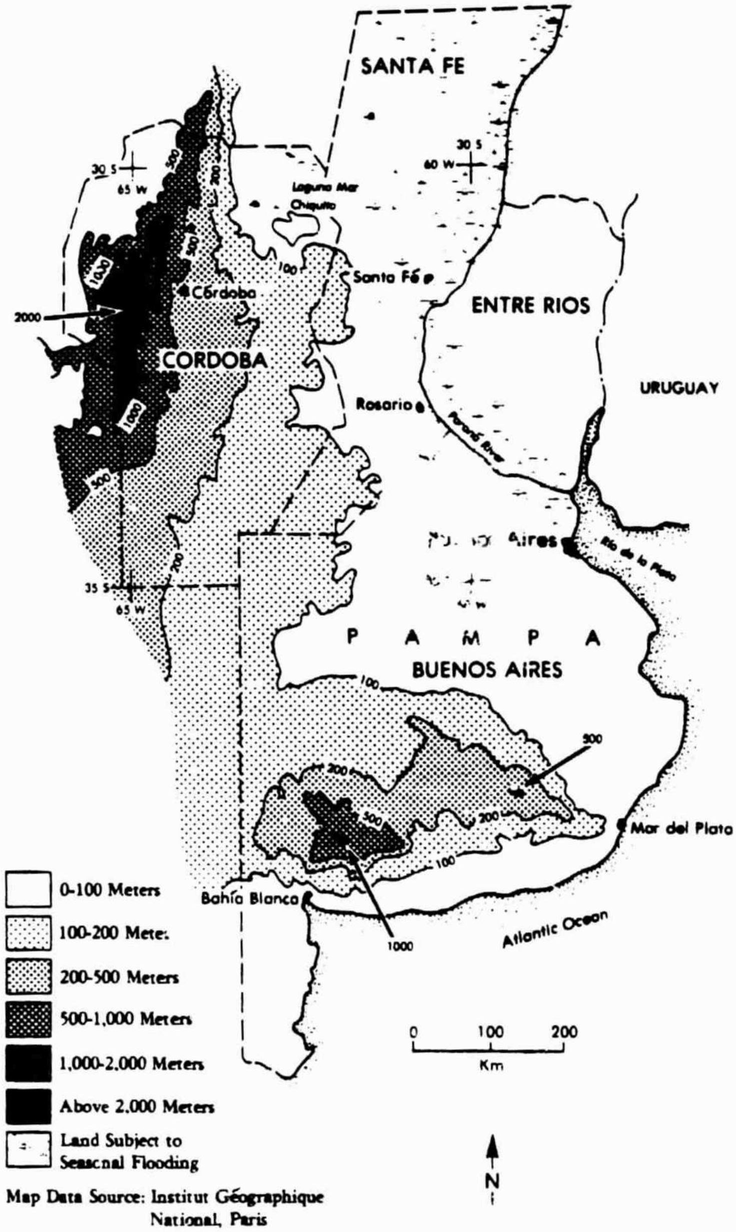
The AgRISTARS study area which had been selected in Argentina (Argentina Indicator Region) for the corn/soybean classification and area estimation technology experiment includes four provinces located in the east-central part of the country (see Maps 1 and 2). Three of the provinces, Buenos Aires, Cordoba and Santa Fe, comprise the Pampa heartland while a fourth province, Entre Rios, is located immediately to the

ORIGINAL PAGE IS
OF POOR QUALITY



MAP 1. AgRISTARS STUDY AREA IN ARGENTINA

ORIGINAL PAGE IS
OF POOR QUALITY



MAP 2. PHYSIOGRAPHY OF AgRISTARS STUDY AREA IN ARGENTINA

east. The study area is situated in the lower middle latitude zone of the Southern Hemisphere, roughly between 30 and 40 degrees South latitude and 59 and 65 degrees West longitude.

Fifty sample segments had been selected in the four provinces, 25 of which are former LACIE segments. Of the total number, about half (26) are found Buenos Aires province with diminishing numbers found in Santa Fe, Cordoba and Entre Rios provinces, in that order.

2.4.1.2 Overview

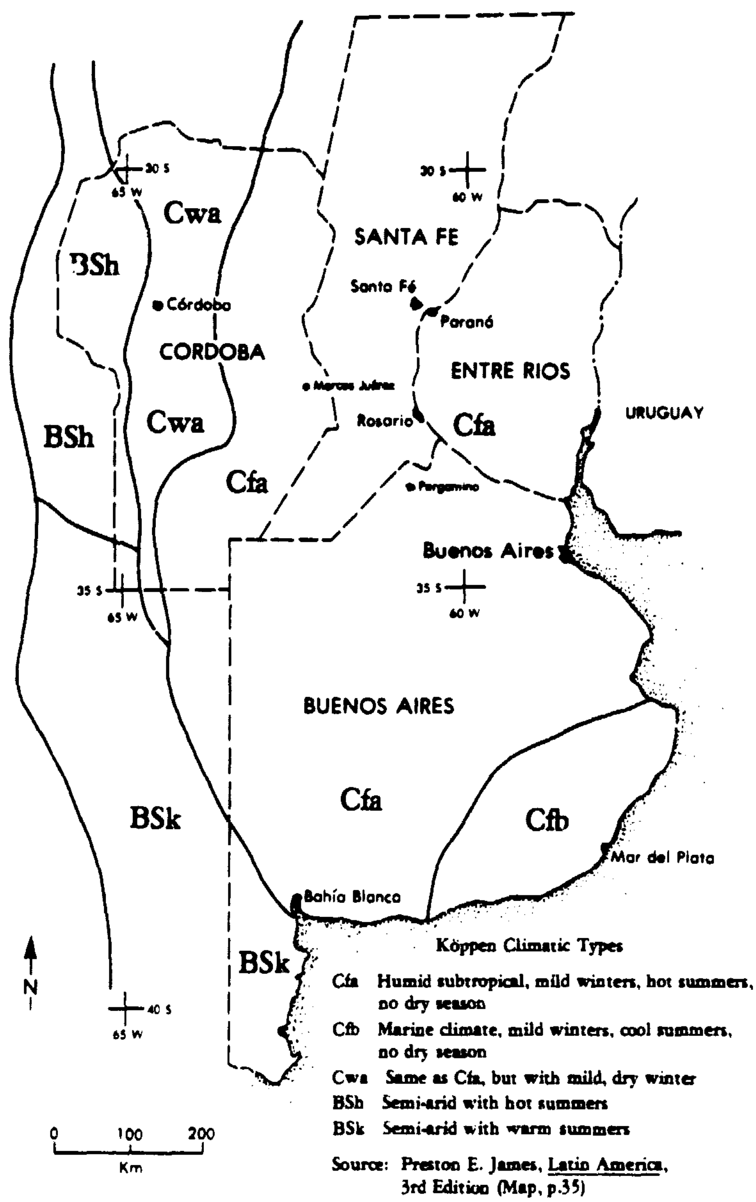
A variety of physiographic factors including nearly level terrain, mild climate, and fertile soil have been conducive to the development of agriculture within the study area. In the center, which covers northern Buenos Aires, southern Santa Fe and southeastern Cordoba, the amount and distribution of precipitation favor the cultivation of corn, soybeans, and other crops, but drought is a problem farther west and south. Conditions in southern Buenos Aires are favorable for wheat production. In Entre Rios, somewhat less favorable conditions for wheat prevail due to high humidity.

Topography and Drainage. The AgRISTARS four-province study area mainly lies within the borders of the Argentina Pampa, a very large, flat to slightly rolling plain that stretches westward into the interior from the east coast of Buenos Aires province, the Rio de la Plata estuary and the lower Parana River Valley (see Map 2). The Pampa extends westward and southwestward well beyond the borders of the study area and ultimately to the desert which separates it from the Andean mountain system. It extends north to the Chaco, a subtropical scrub woodland zone, and southwestward to northern Patagonia. Strictly speaking, the province of Entre Rios is not part of the Pampa, but is a flat plain broken by north-south aligned ridges.

Sedimentary materials cover nearly all of the Pampa, most of which is fine wind-blown loess which was transported from areas farther westward along the Andean front. Generally coarser rock materials are found in the western Pampa while the finer wind-blown materials were carried farther eastward. Two topographic divisions can be distinguished in the Pampa, although differences are subtle. The "Pampa ondulada" or Undulating Pampa exhibits slightly rolling topography such as portions of northern Buenos Aries, southern Cordoba and southern Santa Fe. In constrast, much of central Buenos Aires province to the south is low-lying and poorly drained and forms part of the "Pampa deprimida" (Depressed Pampa), especially to the west of the Parana River (central and northern Santa Fe) where numerous low-lying areas occur. Summer flooding is common in all of these areas and both crop and livestock losses occur. Such events often result in loss of feed for livestock and conversion of cropland to pasture or forage as an emergency measure. Nearly all of the study area, with the exception of a few isolated hill areas and the Sierra de Cordoba highlands in the far northwest, lies below 200 meters elevation as do 13 of the 14 segments visited for ground data collection purposes in 1981.

Climate. The study area exhibits considerable climiatic variation with respect to temperature, precipitation totals, and seasonality and variability of precipitation. The most critical factor in terms of agriculture : the occurrence of drought in interior farming zones. Temperature differences are also important, given the north-south extent of the study zone (1400 kilometers), as is distance from marine moisture sources.

Five climatic types occur within the study area (see Map 3). Most of the area lies within a zone of humid subtropical climate that extends southwestward from Brazil and Paraguay into Santa Fe, Entre Rios, Buenos Aires, and the eastern third of Cordoba. Farther west, a variant of this climate with dry winters and decreased, more unreliable summer rainfall is found. A similar climate prevails much farther south in southwestern Buenos Aires. In contrast, southeastern Buenos Aires has a



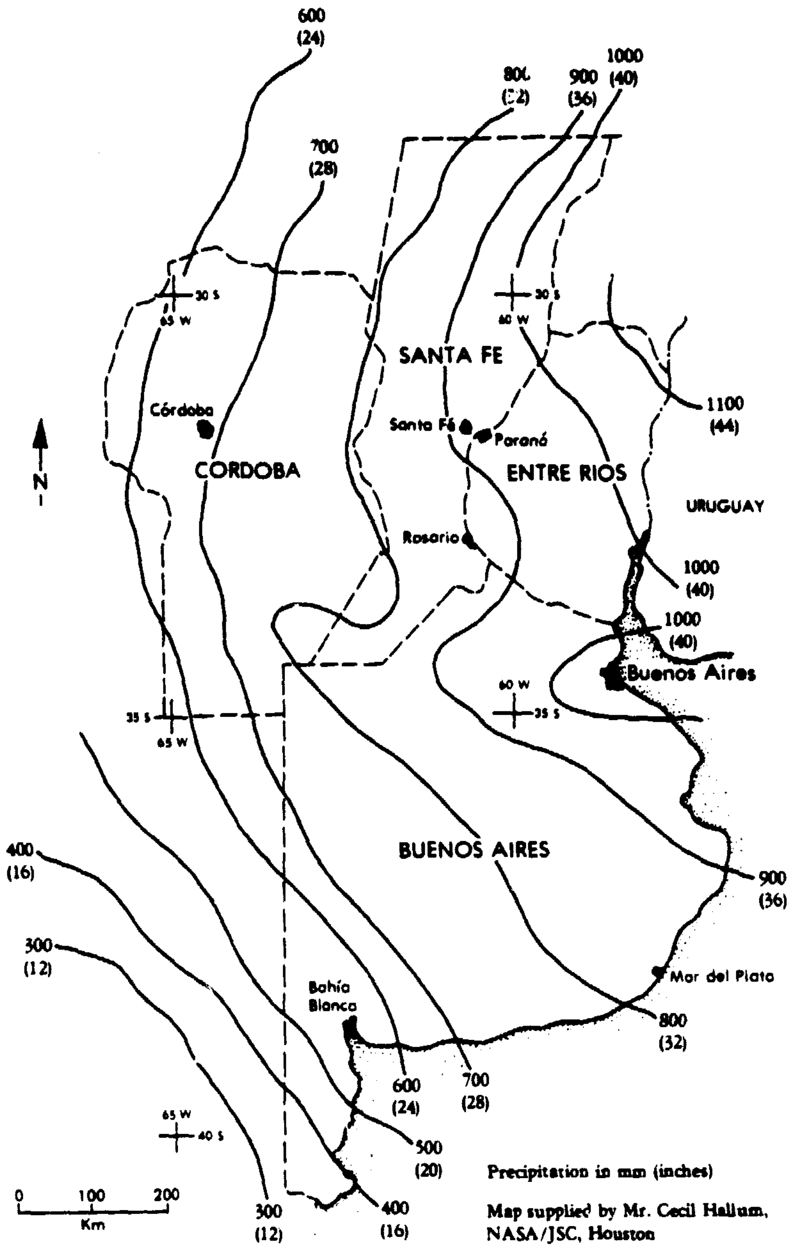
MAP 3. CLIMATE OF AgRISTARS STUDY AREA IN ARGENTINA

cool marine climate because of its proximity to cold offshort currents in the Atlantic Ocean.

Great differences in precipitation occur within the study area (see Map 4). Total precipitation decreases from east to west and from northeast to southwest. The seasonality of precipitation is also very important. Precipitation is more evenly distributed and reliable in northern Buenos Aires than in areas to the west and south, which is a key factor in agricultural land use. Rainfall in the Pampa of northern Buenos Aires is generally adequate for corn and soybean cultivation and is well distributed annually. To the west and south, rainfall decreases, while high temperatures produce high evapotranspiration rates which reduce precipitation effectiveness in the extreme north. In both areas, drought-resistant crops such as sorghum are grown rather than corn or soybeans.

Generally speaking, the region is characterized by long, hot, humid summers and mild winters. Chief climatic controls are landmass heating at subtropical latitudes and the nearby Atlantic moisture source. In more interior locations, the higher temperatures are ameliorated by lower humidity. Frost can occur during winter in interior areas, but snow is rare, and winter climatic conditions are less severe than those of the U.S. corn/soybean zone.

The Pampa region also can be divided into three zones arranged in concentric crescents around the city of Buenos Aires: the Humid Pampa, the Subhumid Pampa, and the Semi-arid Pampa, in order of increasing distance from that city. The Humid Pampa is the center of corn/soybean product and other crops having high moisture requirements while the Subhumid Pampa is used for wheat, alfalfa, sorghum and rye. The Semi-arid Pampa is mainly devoted to livestock raising due to low rainfall. Drought risk increases rapidly to the west of the Humid Pampa while high evapotranspiration rates as well as seasonal flooding adversely affect agriculture and livestock to the north of that same area.



MAP 4. ANNUAL PRECIPITATION IN AgRISTARS STUDY AREA IN ARGENTINA

Soils and Vegetation. Soils within the provinces of Santa Fe, Cordoba and Buenos Aires generally consist of fine wind-blown (aeolian) material transported from the arid west of Argentina with the soil particles of finest texture being transported farthest eastward. The fine wind-blown soil is powdery yellowish loess which is an extremely productive soil for agriculture.

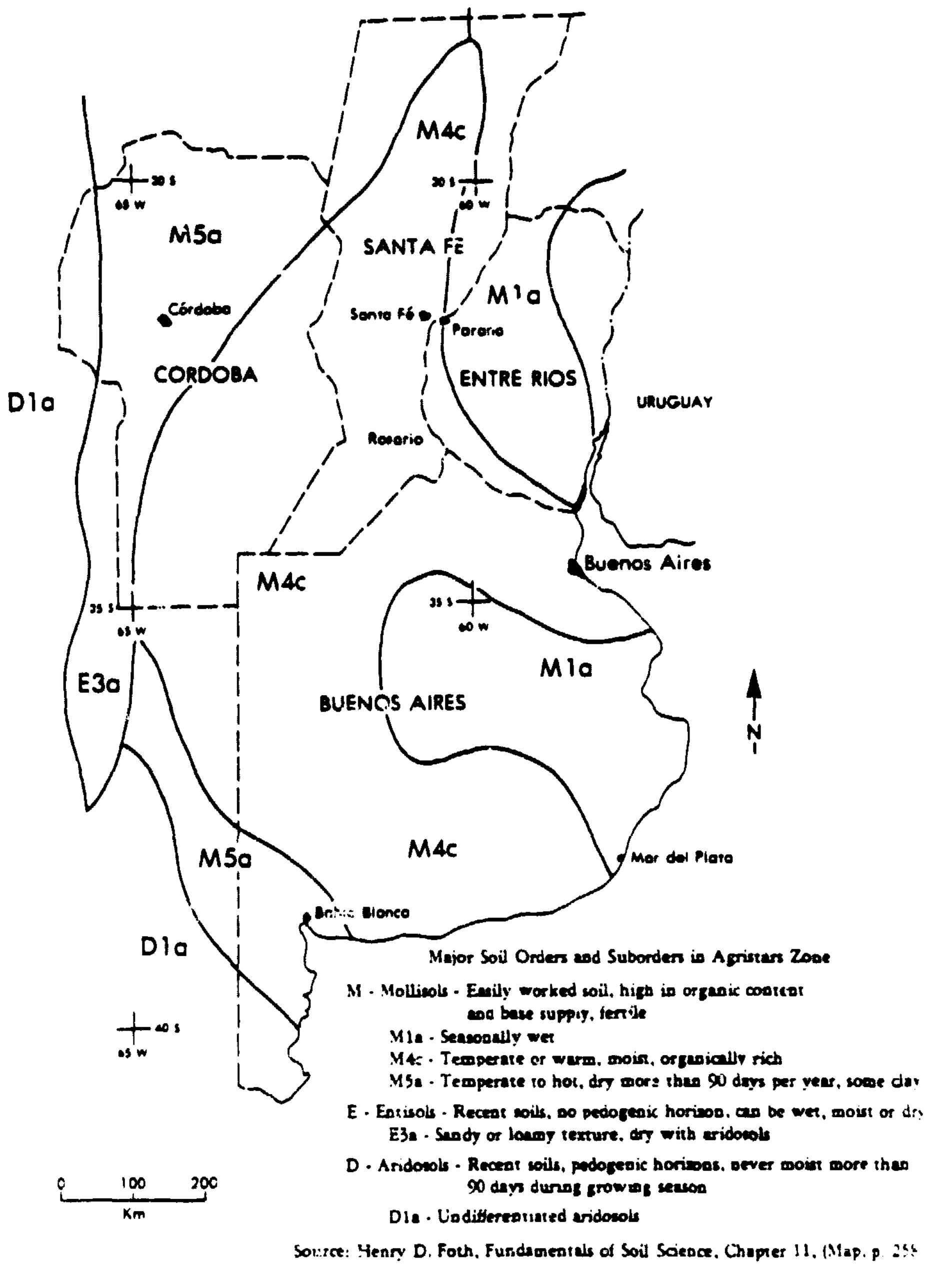
Most of the soils that occur throughout the Pampa region are classified as mollisols (see Map 5). These soils are easily worked, very fertile and are similar to those found throughout much of the U.S. Corn Belt.

Within the Pampa, several types of mollisols have developed due to parent material and climate. The most extensive types are the Udolls which occur in the Humid Pampa. These soils are moist, very high in organic matter and have great agricultural potential. To the west are Ustolls, a drier soil variant of the former type which have developed in areas that are dry for at least 90 consecutive days annually. In southern Cordoba, soils that are transitional between Udolls and Ustolls are found while, in the extreme southwest of Buenos Aires, conditions have favored the development of Aridosols, an even drier variant. The soils of Entre Rios are also Mollisols of the Alboll subtype. These soils are seasonally wet due to much higher precipitation and are also less permeable due to high clay content.

The original vegetation cover of the Humid Pampa was prairie grassland when the first Spanish explorers arrived. Tall plumed grasses covered most of the zone and marsh vegetation was also widespread, given the large number of poorly drained topographic depressions. As the Pampa was settled, this vegetation type was greatly modified through the planting of eucalyptus treees as windrows and woodlots.

Rainfall gradually decreases to the west and southwest of the Humid Pampa and the grassland windrow vegetation of that zone gradually gives way to short-grass steppe. In contrast, extreme northern Santa Fe and

ORIGINAL PAGE IS
OF POOR QUALITY



MAP 5. SOILS OF AgRISTARS STUDY AREA IN ARGENTINA

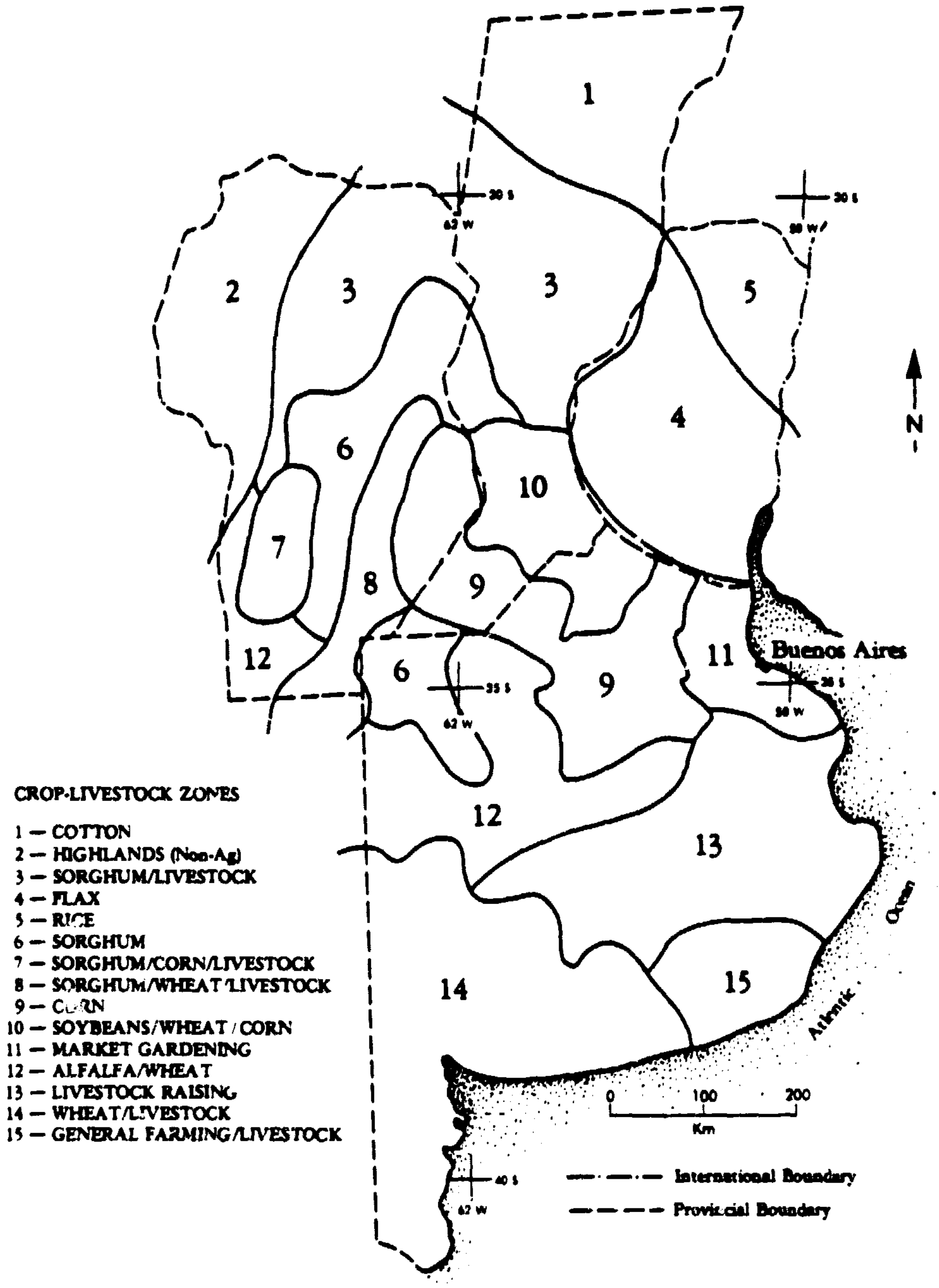
Cordoba lie along the southern margin of the tropical scrub woodland "Chaco" zone, while central Santa Fe and east-central Cordoba are transitional between the grassland-scrub of the Chaco margin and the Humid Pampa grasslands to the south. Scrub forests as well as marshland occur along the Parana River valley and extend far upriver out of the study area. However, marshland areas are also found immediately to the west in central Santa Fe province. Extensive marshland zones also occur in the low-lying poorly drained "Depressed Pampa" of central Buenos Aires as well as in some areas of southeastern Buenos Aires. Other types of vegetation are also found. A "parkland" vegetation type consisting of scattered trees and grassland typifies much of southern Entre Rios.

In general, existing vegetation closely corresponds to precipitation amounts received, evapotranspiration rates and topography. Decreasing precipitation is reflected in the southwestern and western short grass steppes, while high evapotranspiration rates and poor drainage are major factors that influence vegetation in the far north.

2.4.1.3 Crop/Livestock Zones in the Argentina Study Area

Despite the relative physiographic homegeneity of the Pampa region which characterizes most of the AgRISTARS study area in Argentina, very substantial differences in agricultural land use, crop mix and practices exist, due mainly to differences in rainfall amount and distribution (see Map 6).

Zone 1 - Cotton. The cotton area shown in northern Santa Fe is a southward extension of Argentina's major cotton production zone which also covers parts of the provinces of Formosa, Chaco and Santiago del Estero. Moderate rainfall, high evapotranspiration, poor drainage and sporadic flooding of cotton plantings characterize the zone. The zone is geographically remote from all 50 segments in the study area and is therefore not directly relevant to the corn/soybean agronomic understanding efforts of this subtask.



MAP 6. CROP-LIVESTOCK ZONES IN ARGENTINA INDICATOR REGION

Zone 2 - Highlands. This highland zone in extreme northwestern Cordoba (Sierra de Cordoba) is a non-agricultural zone and is likewise not of direct concern to the corn/soybean agronomic understanding effort.

Zone 3 - Livestock/Sorghum; Zone 6 - Sorghum; Zone 7 - Sorghum/Corn/Livestock; Zone 8 - Sorghum/Wheat/Livestock. These four zones represent various crop mixes, but in all cases, sorghum cultivation is significant. The zones are all located in the Subhumid Pampa, west and northwest of the Humid Pampa centered on northern Buenos Aires. In all four zones, sorghum along with beef livestock raising is the chief rural activity. Zone 3 covers northern Cordoba and central Santa Fe. Livestock pasture is the chief land use in this zone with most sorghum grown being forage sorghum. The sorghum plant's resistance to drought makes it the chief crop as very little corn or soybeans are in the far north due to moisture limitations and drought prevalence. Still, the amount of sorghum grown in Zone 3 is much less than in Zone 6 due to high evapotranspiration which reduces precipitation effectiveness, except for northeast Cordoba where more sorghum is grown. Zone 6 is a slightly more humid area than Zone 3 and is Argentina's major sorghum production zone. The largest portion is located in central Cordoba, while the remainder is located in extreme western Buenos Aires. Livestock raising remains important, but the percentage of land devoted to sorghum is much greater in Zone 6 than in Zone 3. In addition, some soybeans are grown in the zone. Zone 7 is similar to Zone 6, but corn is also a major crop. Zone 7 is the largest producer of corn in Argentina outside of the Humid Pampa for reasons not clearly understood, given the low average annual precipitation for the zone, 700 mm (28 in). However, livestock activities for forage sorghum production remain important. Zone 8 is similar to Zone 7 except that wheat production is also important. Precipitation is also slightly higher, 750 mm (30 in). Wheat production is greatest in the northern portion of Zone 8 and gradually decreases southward. Also, the zone accounts for less of the Argentine wheat total

C-2

than in the past as newer production zones in southwestern Buenos Aires have become more important. The northern part of Zone 8 is relatively densely populated, by Argentine rural standards, and has been an important agricultural zone since about 1900.

Agricultural practices within the four zones are fairly uniform. Irrigation is virtually non-existent and many sorghum fields were weed-infested due in part to the high organic content of the soil and the lack of herbicide application which would discourage weed proliferation. Furthermore, fertilizer use remains low due to high prices and high natural soil fertility. Crop rotation is practiced but no consistent, organized system exists. Land left in pasture for several years is generally planted to forage sorghum with the decision to plant being made in a real-time context because of weather and changing market prices. Most pastures are unimproved in the north but alfalfa becomes more important in Zone 6. Also, the flooding of forage crops in low-lying areas may necessitate sudden new plantings of sorghum or oats planted for livestock ground forage.

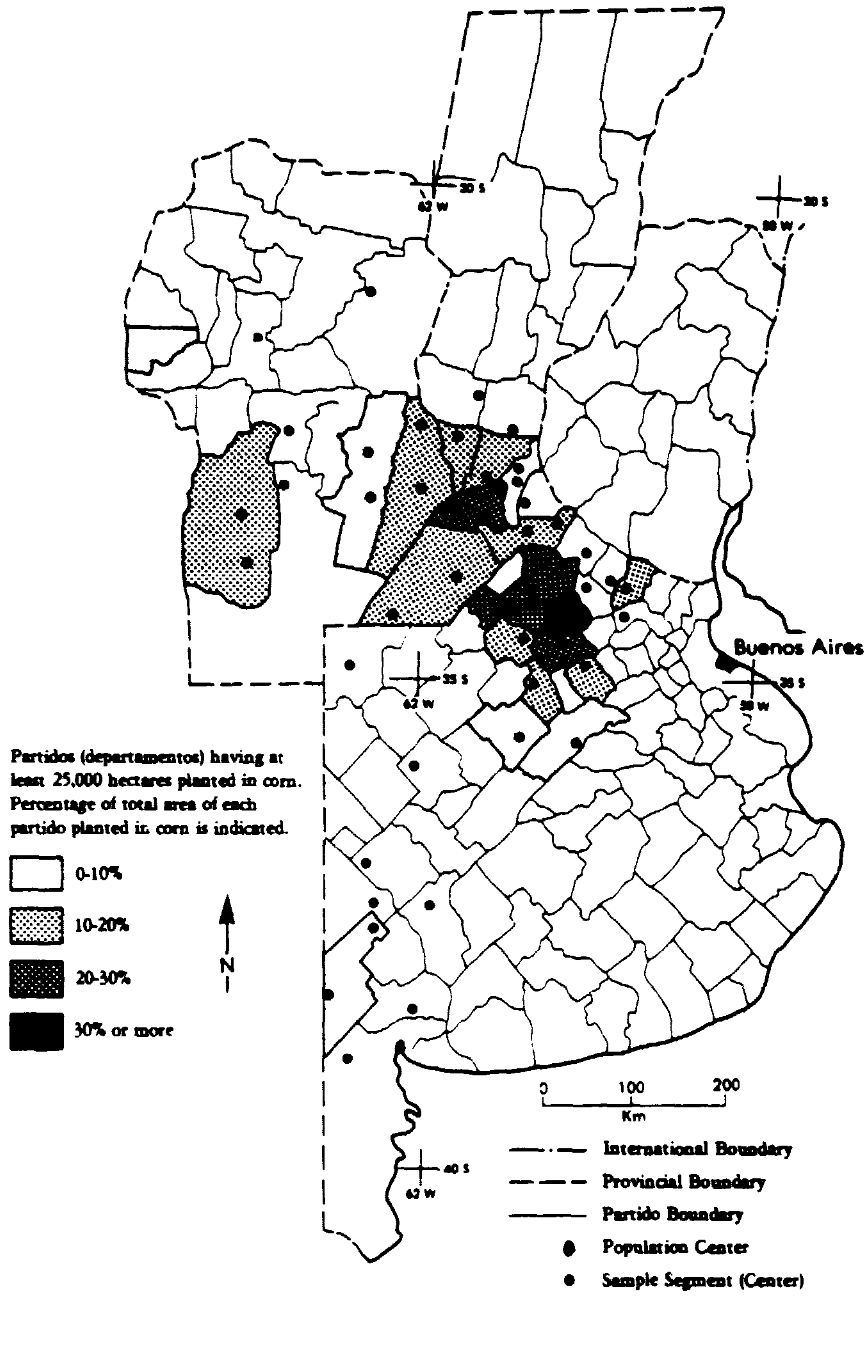
Zone 4 - Flax. Zone 4 covers most of Entre Rios province except the extreme northeast. Flax is the chief crop grown in the zone with the heaviest concentration being in central and southern Entre Rios. Livestock raising is of some importance, as are corn and soybeans in the extreme west-central portion. Although one segment is allocated to Entre Rios, Zone 4 is somewhat peripheral to corn/soybean technology development for Argentina, as flax and linseed oil production dominate the zone's economy.

Zone 5 - Rice. Zone 5 is a southern continuation of Argentina's major wet rice production zone, most of which islocated in Corrientes province to the north outside the study area.

Zone 9 - Corn; Zone 10 - Soybeans/Wheat Corn. Zones 9 and 10, located in the Humid Pampa of northern Buenos Aires, southeastern Cordoba and southern Santa Fe, arethe chief areas of interest relative to the Argentina agronomic understanding subtask. Zones 9 and 10 account for approximately 80% of Argentina's corn, while Zone 10 accounts for over 90% of the nation's soybeans. Climatic conditions within the zones are very favorable for the cultivation of both crops, but soybean production is geographically concentrated in the northeastern portion of the larger corn production zone (see Maps 7 and 8). Corn and alfalfa production along withlivestock raising is important in Zone 9, as is sunflower cultivation. Zone 10 is also important for corn cultivation but soybean/wheat double cropping surpasses corn in area planted and is the chief agricultural activity. About 75% of the soybeans grown are double cropped with wheat but this percentage may vary about 10% above or below this figure for different years.

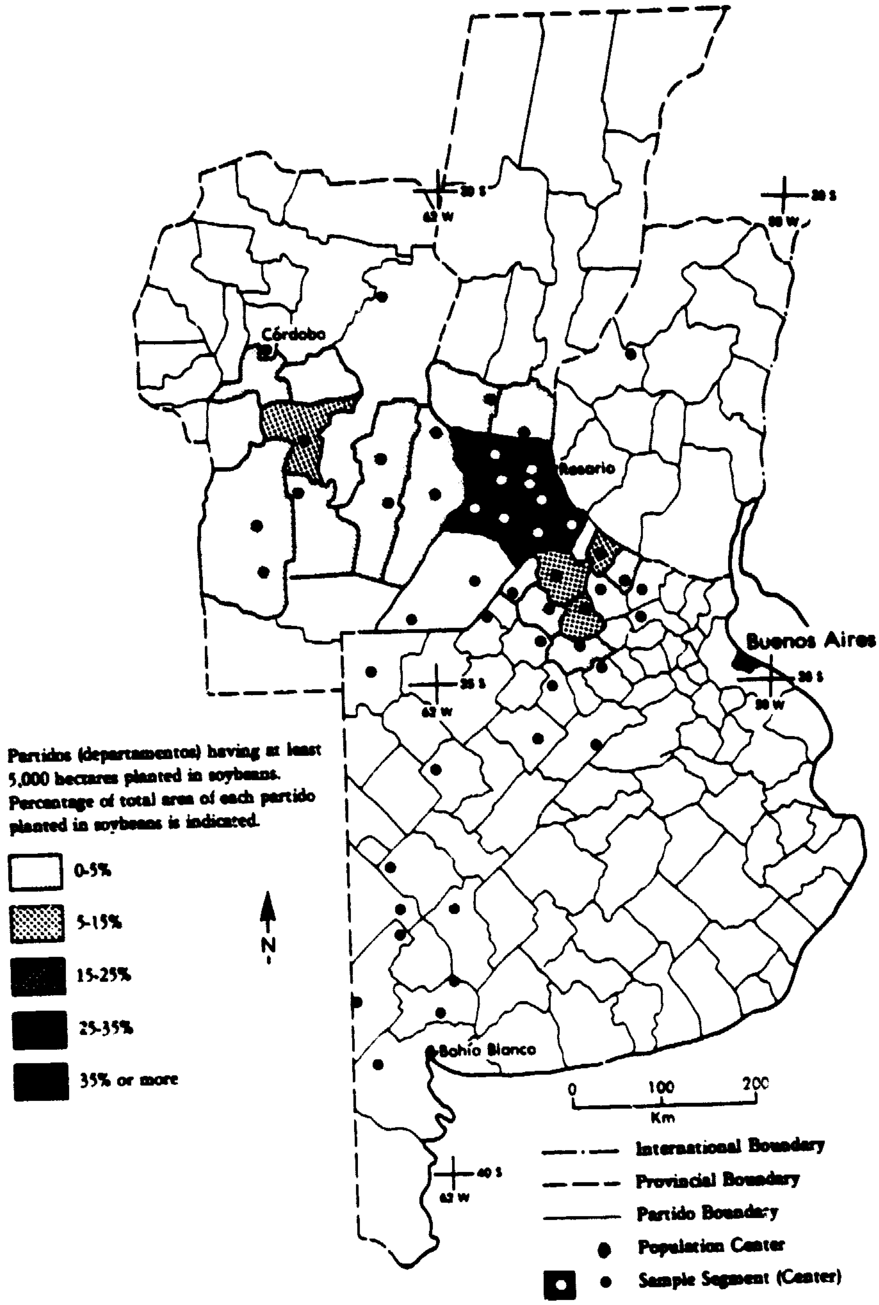
Mechanized agricultural production is widespread in Zones 9 and 10. Although mechanization levels are lower than in the U.S. Corn Belt, they are nevertheless high by Latin American standards. Three-to five-bottom (moldboard) plows are used on smaller farms, while ten-to fifteen-bottom implements are used on large properties. No-till planting is not widely practiced since plowing is considered a weed control measure.

Planting times are governed by temperature, drainage conditions and moisture availability. Corn is normally planted from mid-September to mid-October in both zones and harvested in March. However, planting and crop growth dates, as well as harvest dates, vary with weather and location. Soybean planting and harvest dates vary substantially depending on whether the fields are single-cropped or double-cropped after wheat harvest. Row width for corn, soybeans and grain sorghum is 70 cm, and that of forage sorghum and winter wheat is 15 cm.



MAP 7. DENSITY OF PLANTED AREA IN CORN 1977/78 CROP YEAR

ORIGINAL PAGE IS
OF POOR QUALITY



MAP 8. DENSITY OF PLANTED AREA IN SOYBEANS 1977/78 CROP YEAR

Several other agricultural practices deserve mention. In some areas of the zone, wheat and alfalfa are intercropped in the same field. Planted wheat is mature after about 125 days and following the harvest, the alfalfa is left for beef cattle pasture. Two major rotation patterns are also practiced. In many cases, fields may remain in pasture for five or six years after which time a row crop is planted such as corn, grain sorghum, or soybeans. Should single-crop soybeans be planted, the land would revert to a fallow condition following harvest. In cases where second-crop soybeans are planted, winter wheat is again sown in the field following the soybean harvest. After one or two years of row crops, the land would be left to pasture once again and an adjacent field planted in row crops. A second rotation pattern is the planting of corn, followed by rye, and then corn once again, after which time alfalfa is planted for three years.

Zone 11 - Market Gardening. Zone 11 is a zone of intensive vegetable and fruit production serving the city of Buenos Aires. The zone, which forms a crescent around metropolitan Buenos Aires on its northern, western and southwestern margins, is located outside the major corn/soybean production zone and is not directly relevant to this agronomic understanding subtask.

Zone 12 - Alfalfa/Wheat. This, the major alfalfa/wheat production zone in the Argentina study area, is located to the southwest of the principal corn/soybean growing areas. Despite its proximity to the corn/soybean zone, corn production is much less and soybean production is negligible due to decreased annual precipitation and erratic and unreliable rainfall patterns. Drought is a major risk in the zone and farmers therefore plant alfalfa or wheat. Sunflowers are also of some importance. Alfalfa is planted in March as winter forage throughout the zone, and is cut in May, July and September. In October, alfalfa is usually planted for a second time and the process is repeated. Unlike the U.S. Corn Belt, feedlot fattening of livestock is not commonly practiced in Argentina. Rather, alfalfa is the principal livestock feed, along with forage sorghum. Winter wheat is also grown, but production is generally less than in eastern Cordoba to the north, or areas farther south in Buenos Aires. In some areas of the zone, wheat and alfalfa are intercropped in the same field. Also, alfalfa is sometimes rotated with rye to restore soil moisture. Despite drought risk, irrigation is not practiced in the zone.

Zone 13 - Livestock Raising. Zone 13 located in central Buenos Aires is a low-lying, poorly drained area devoted mainly to beef livestock raising. Corn and soybean production are not important within the zone, due principally to poor drainage and flood risk. However, annual precipitation is sufficiently high, 800-900 mm (32 to 36 in), to support their cultivation. Oats, barley and rye are grown within the zone as cattle feed but many cattle are sent to alfalfa producing areas in Zone 12 for fattening prior to marketing. Some wheat is also grown but, as in the case of Zone 12, the amount grown is much less than in southern Buenos Aires.

Zone 14 - Wheat/Livestock. Argentina's largest and most important wheat growing region is located in southwestern Buenos Aires, south of a diagonal line separating it from Zones 12, 13 and 15. Pasture, wheat cultivation and some forage sorghum dominate rural land use but wheat is by far the most important crop produced. Precipitation decreases steadily from northeast to southwest to the extent that corn and soybean production is preculded in the southwest. Wheat is normally planted in June and harvested in late December. Following harvest, oats are normally planted in wheat stubble as forage for cattle. Also, several varieties of pasture grass are planted, but alfalfa plantings are of little importance, unlike areas farther north. Irrigation is rarely practiced and many pastures are unimproved and weedy.

Zone 15 - Livestock/General Farming. In southeastern Buenos Aires, the crop mix is considerably different from all other zones in the Pampa. Total annual precipitation is nearly double that of southwestern Buenos Aires and relative humidity is much higher. In addition, the soils of southeastern Buenos Aires are very high in organic matter (16%) and are among the most productive in Argentina. However, poor drainage and salinity are problems in some locales. Durum wheat, potatoes and pasture (alfalfa) used for livestock raising rather than fattening, dominate land use in Zone 15. Although, potato production is favored by the cool, moist climate as is rye and barley cultivation, the cooler temperatures discourage the production of corn and soybeans within the zone despite rich soils. Potatoes, which are the chief crop, are normally planted for two years followed by the planting of wheat, and then oats.

2.4.1.4 The Argentina Agricultural Economy

In 1981 the Argentine agricultural economy was adversely affected by poor weather in some crop zones as well as severe inflation. However, positive indicators resulted from the conclusion of several new bilateral trade agreements which will guarantee markets for agricultural products. The nation's major cotton production zone in the far north suffered serious flooding as a result of heavy rains in January and February 1981. Also heavy rains in April and May 1981 delayed the harvest of corn, soybeans and sorghum. Secondly, the agricultural sector of the economy was beset by high inflation which triggered successive monetary devaluations and rapidly increasing farm production costs. Consequently, some export rebates paid to farmers to stimulate production subsequently had to be rescinded since they were inflationary. High production costs continue to hold back the purchase of new farm equipment and the implementation of new approaches. Consequently, farmers opt to reduce costs by using traditional farming methods. The lack of irrigation in areas where needed, poor maintenance of some fields, and lower fertilizer consumption are examples of this situation.

About 75% of Argentina's exports are agricultural products, mainly wheat, corn, sorghum and soybeans. Given this, market guarantees for these crops are a critical issue. In addition, Argentina chose not to participate in the U.S.-sponsored Sovietgrain embargo initiated in 1980. In that same year, Argentina concluded a five-year agreement with the USSR. The agreement calls for annual Soviet purchases of three million metric tons of corn, 2.4 million metric tons of wheat, one million metric tons of sorghum, and 500,000 tons of soybeans. A re-negotiated agreement with the People's Republic of China was also concluded in 1980 which calls for the annual sale of one million to 1.5 million metric tons of corn, soybeans and wheat to the PRC. A third agreement between Argentina and Mexico was also signed in 1980 covering the 1981 and 1982 calendar years, during which time Mexico will purchase one million tons of corn, soybeans, sorghum and sunflower seed. A major task now confronting Argentine producers is to be able to meet the new export commitments given the high production and transportation costs involved.

2.4.2 FIELD DATA COLLECTION

Integral parts of the Argentina/Brazil Agronomic Understanding subtask were the collection of ground data in Argentina during February 1981, participation in an in-country evaluation of the USDA Brazil Sampling Frame (also conducted in February 1981), and the preparation of a ground data collection plan for Argentina for the 1981-1982 crop year.

2.4.2.1 Ground Data Collection in Argentina During 1981

During February 1981, a trip to Argentina was made by members of a consortium composed of staff from the Environmental Research Institute of Michigan (ERIM) and the Space Sciences Laboratory of the University of California at Berkeley (UCB). The general objective was to begin to gather and synthesize a wide range of agronomic information that could be used as a data base by AgRISTARS researchers working on research, development, and testing of technology for application in Argentina. Preparations for the trip begin in late 1980 and February was chosen as the time frame for field work since both the corn and soybean crops would be in advanced stages of phenological development at that time. A full trip report is contained in a separate technical report [20]. A summary follows.

The trip had several specific interrelated objectives:

(a) To become familiar with the problems as well as the opportunities for collection of ground data in support of AgRISTARS program needs.

(b) To collect crop identification data for a limited number of fields in 14 5x6-mile sites located throughout the corn, soybean, and wheat growing areas of the Argentine pampa, and to acquire collateral data such as crop calendars, and historical agronomic statistics.

(c) To meet with public officials representing the agronomic and remote sensing community of Argentina in order to familiarize them with our goals and gain their collaborative support for this ground data collection expedition.

(d) To encourage these public officials to consider future involvement in the AgRISTARS program that would be mutually beneficial.

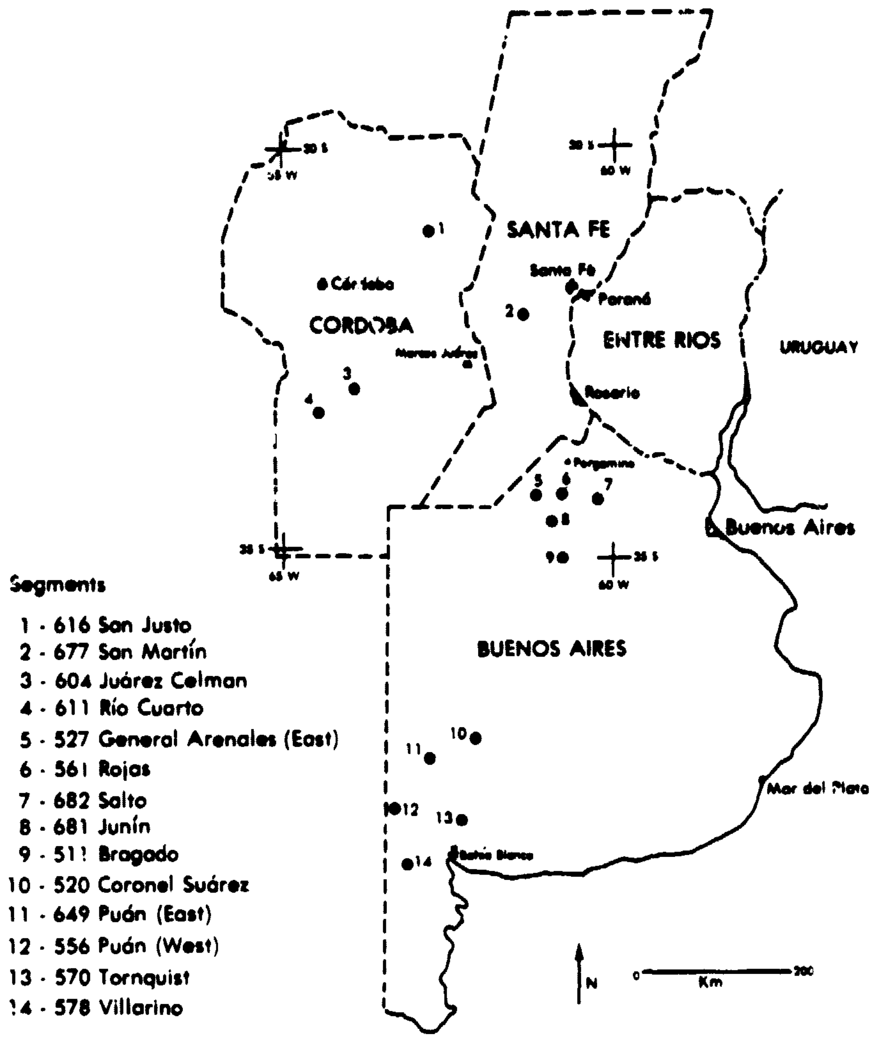
All of the objectives, in our opinion, were achieved. The Agronomic Understanding Task team is satisfied that its first-year data-collection goals in Argentina were achieved. The data collected and observations made will provide a useful foundation for future activities. Perhaps more important is our impression that there is considerable interest among key agency officials in Argentina in making productive use of contemporary remote sensing technology in agriculture. They graciously provided support to our field trip and appear open to future participation in AgRISTARS-related activities. Also important to the success of the field trip were the timely planning assistance of NASA/JSC, their rapid response to our data needs, and the assistance and coordination of USDA staff in securing introductions in Argentina and providing other needed support.

During the 14-day period of field work, 14 segements were visited (see Map 9), with assistance provided by the State Secretariat of Agriculture and Livestock Raising (SEAG) and the National Commission for Space Investigations (CNIE). Roadside observations of crop identification and condition were annotated on enlarged color Landsat imagery of the sites, as were field boundaries. In the case of two of the segments, aircraft overflights made possible the identification of additional crops in fields inaccessible by road. Over 500 ground and air photos were taken during the inventory to provide information for subsequent study and crop identification information for 629 fields was obtained. Two soil samples and a small quantity of hybrid flint corn seed were gathered and transmitted to other AgRISTARS researchers at Purdue University. In addition, historical crop calendar data and crop acreage statistics were obtained for three provinces.

The trip report contains descriptive information, maps of sample segment areas visited, and an annotated graytone Landsat image of each segment showing crop identification codes, field boundaries and pertinent remarks about individual fields where warranted. In addition, a few copies also contain annotated color Landsat images as well as color slides with commentaries.

The annotated crop identification data for each inventoried field in the 14 segments were digitized and merged with Landsat data at ERIM under ITD support, as discussed in Section 3.3.5 of this report.

ORIGINAL PAGE IS
OF POOR QUALITY



MAP 9. LOCATION OF ARGENTINA SEGMENTS WHERE GROUND DATA WERE COLLECTED - 18-26 February 1981

2.4.2.2 Brazil Sampling Frame Evaluation

The United States Department of Agriculture, Economics and Statistics Service (USDA/ESS, Fairfax, Virginia), developed a sampling frame for future use in a Brazil Corn/Soybean Pilot Study. In February, two of their personnel conducted a trip to evaluate it. Dr. David R. Hicks of ERIM was invited to accompany them since he had extensive agronomic field experience in southern Brazil and spoke Portuguese.

Previously annotated Landsat images showing percentage of land under cultivation and the percentage of land devoted to corn/soybean production were brought into the field by team members so that their accuracy could be assessed through ground truth checks. In addition to assessing the accuracy of prior percentage estimates of agricultural land use, the team paid special attention to the problem of small field detection.

Trips were made to six cities in the southern Brazilian states of Parana, Santa Catarina and Rio Grande do Sul. From those cities visits were made to selected outlying agricultural areas for the purpose of evaluating the annotated Landsat imagery as a potential sampling frame. The sampling frame evaluation proved to be generally successful, i.e., the percentage data shown on the annotated Landsat imagery were quite accurate upon being compared with ground truth checks. However, the percent of land classified as agricultural on the Landsat in the plateau escarpment area west of Curitiba in Parana state was greatly overestimated. Secondly, the detection of small fields on Landsat images was not possible, as was anticipated. The results of this evaluation appear in a subsequent USDA trip report [21], as well as in Notes for Brazil Sampling Frame Evaluation Trip published by ERIM in August 1981 [22].

The trip, in addition to its original purpose, served as an opportunity to obtain a general understanding of crop-livestock systems in southern Brazil. Some agronomic data also were obtained as were numerous soil samples. This reconnaissance should provide a useful background for future studies of and visits to Brazilian corn/soybean zones, should a cooperative program be developed.

2.4.2.3 Argentina Ground Data Collection for 1981-1982 Crop Year

A key objective within the Argentina/Brazil Agronomic Understanding Subtask was to identify research needs and establish requirements for future data collection missions in Argentina that could build on information already obtained in and from that country. In response to this need, a collection plan for 1981-1982 crop year was prepared at ERIM [23]. This same plan was subsequently translated into Spanish and also published that same month [24]. The document outlined plans for data collection and field research in Argentina for 1982 through 1984 and proposed steps to be taken by United States and Argentine researchers and government agencies to achieve mutually beneficial objectives.

2.5 INFORMATION EXTRACTION TECHNOLOGY RESEARCH

This section describes work carried out during FY81 to better understand the temporal-spectral development patterns (profiles) of corn and soybeans. To that end, a technique was defined for deriving standard profile features from spectral data collected at different times, years, and/or intervals. The technique was then applied to field reflectance data collected at the Purdue Agronomy Farm by personnel from the Laboratory for Application of Remote Sensing (LARS), after which changes in those features as a function of treatments applied to experimental plots were quantitatively assessed, and compared to expectations derived from review of relevant literature in the area of agronomic research.

This work, summarized in Sections 2.5.2 and 2.5.3, represents the initial phase of an overall data analysis approach described in Section 2.5.1. Details of the analyses are available in Reference [25].

2.5.1 OVERALL APPROACH

The evaluation of crop spectral characteristics as viewed by Landsat is hindered by a number of largely external factors. First, atmospheric effects, illumination geometry, and similar phenomena result in variations in signal values entirely removed from the characteristics of the crop being viewed. Second, misregistration and ground truth errors can create substantial problems with regard to obtaining a pure sample of a crop. Third, and for the present purpose most important, environmental conditions, cultural practices used, crop development stages, and similar pieces of data are unavailable and/or imprecise for the majority of Landsat data.

As a result of all these factors, conclusions drawn with regard to crop spectral characteristics, crop separability, or classification techniques which are based largely or entirely on Landsat data will be extremely dependent on the particular set of data employed. A better approach to deriving information about crop appearances in Landsat data is to begin as close to the plants themselves as possible and, in effect, to step back by increments, moving farther away from the plants or field at each increment, but utilizing the results of the previous higher-resolution steps as a context in which to evaluate information obtained at the present level.

This approach recognizes that the basic elements of interest in classification or interpretation of Landsat data for agricultural applications are not pixels, but rather collections of biological entities. The better we understand workings at the plant or plant population level, the better able we will be to understand and utilize Landsat data in deriving crop-related information.

In practice, this approach to crop spectral understanding consists of some or all of the following steps:

1) Determining relevant physiological, cultural, and environmental influences on those characteristics of plants or plant populations likely to influence their spectral appearance. This involves review of literature in the field of agronomic research and, frequently, gleaning of pertinent information from reports of experiments whose purposes are far removed from remote sensing interests.

2) Modeling the effects of these influences on crop spectra. A model such as that described in Section 2.6.1 provides a means of assessing the spectral expression of particular changes in crop characteristics while keeping all other factors constant.

3) Evaluating field reflectance data to determine or confirm the effects of key factors on crop spectral characteristics. This step provides the crucial link between the modeled data and the real world, but maintains a fairly high degree of control over confounding effects.

Results of modeling, and the plant-level information gathered at earlier steps, provide a context in which to understand the results obtained through field data analysis.

4) Evaluating Landsat data to adjust expectations and conclusions formulated at the other levels. Having established a foundation and context through the previous analyses, one can analyze Landsat data, in conjunction with whatever associated information is available (crop labels, weather data, etc.), and better understand and explain what is seen there. The quantitative results of the previous levels are combined with a Landsat data set that is probably larger, more geographically widespread, and more variable in terms of crop mix and growing conditions, to allow more comprehensive evaluation of crop spectral characteristics.

2.5.2 CURVE-FITTING TECHNIQUES FOR ANALYSIS OF CROP SPECTRAL DEVELOPMENT PATTERNS

Analysis of crop spectral data collected at discrete intervals, and particularly at irregular discrete intervals, is often restricted by the absence of observations at key times in the crop development cycle. In addition, comparison of data from different plots or locations is hindered by the temporal mismatch of observations between plots. Even when all plots are observed on the same days, planting date differences cause a mismatch of data with respect to some sort of 'effective day' time scale (e.g., days since planting). In order to make meaningful comparisons among several plots, some method must be devised by which the spectral characteristics of the plots may be described in a standard fashion.

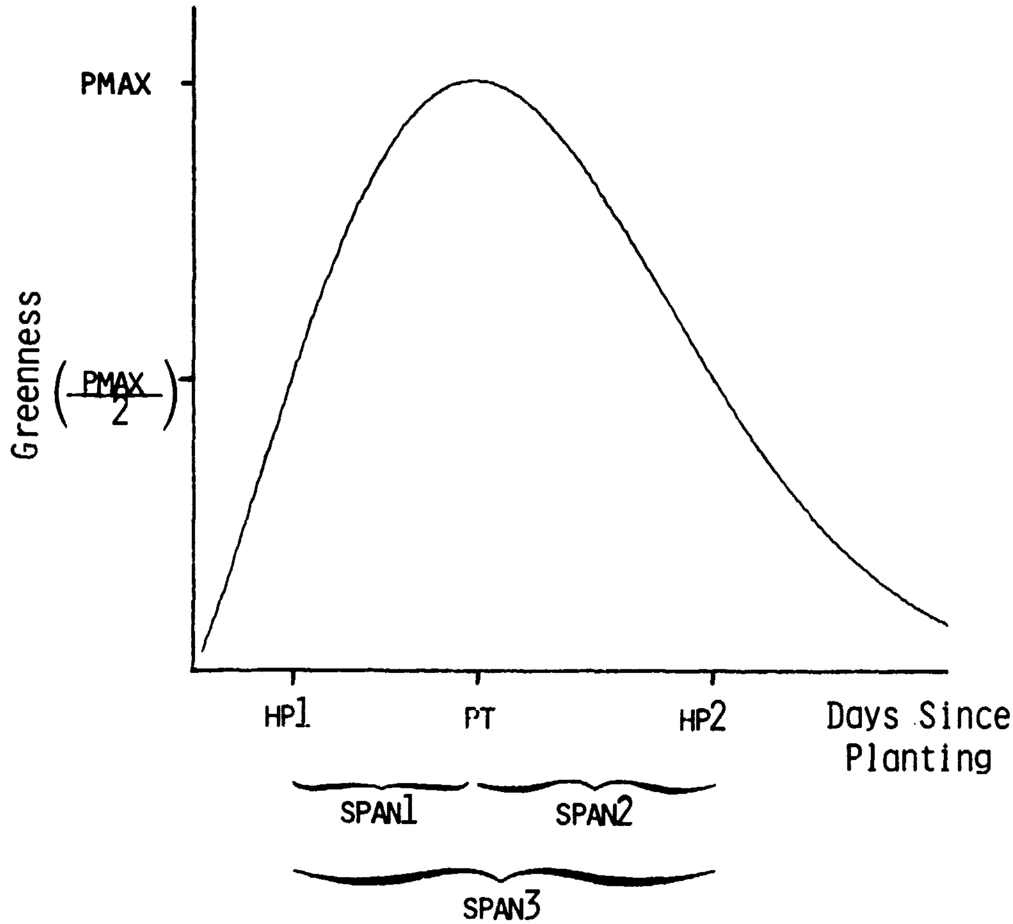
The technique developed at ERIM for this purpose consists of two elements: a standard set of features, and a curve-fitting technique for deriving those features for any particular plot.

Profile Features. Analysis carried out in FY81 used Tasseled-Cap Greenness as the spectral variable. The Tasseled-Cap transformation, and its adaptation to reflectance data, are described in Section 2.5.3. Figure 2.9 shows a typical, simple Greenness profile, and illustrates the set of features used in this analysis. These features represent a basic set of parameters to describe any simple curve of more or less a bell shape. Particular crops may warrant additional features, although this standard set should still be appropriate. For example, corn data tend to appear as a flattened bell shape (Figure 2.10). This shape has been observed both in spectral data [26,27] and in other agronomic variables (e.g., leaf area index) correlated to Greenness [28]. While additional features were not used in the analyses described in Section 2.5.3, some possible additional features are described in Figure 2.11. Use of a spectral variable other than Greenness would simply require that a new set of features be defined.

Curve-Fitting Technique. In order to use the profile features just described, the intermittent spectral data must be transformed into a smooth, continuous curve.

An approach which offers some smoothing of irrelevant data variation without the complexity of empirical modeling is the use of a curve-fitting function to derive a new set of smoothed data based on the original observations. As long as one can be reasonably confident that the majority of data taken over a particular plot is free from major external effects, that is, that the outliers in a set of observations are the contaminated rather than the pure data, then a curve-fitting technique can provide some more or less-precise correction for major externally-induced variations.

Work toward selecting a smoothing technique involved less an exhaustive evaluation of all possible approaches and more an evaluation of a few particular techniques which were readily available and comprised



Feature	Description	Associated Agronomic Characteristic
PMAX	Maximum profile value	Maximum amount of green vegetation
PT	Time of PMAX	Rate of vegetative growth
HP1	Time of first half-peak	Rate of emergence and early vegetative growth
SPAN1	Time from HP1 to PT	Rate of later vegetative growth
HP2	Time of second half-peak	Rate of development from planting to senescence (overall development time)
SPAN2	Time from PT to HP2	Rate of senescence
SPAN3	Time from HP1 to HP2	Rate of development after emergence

FIGURE 2.9. GREENNESS PROFILE FEATURES

ORIGINAL PAGE IS
OF POOR QUALITY

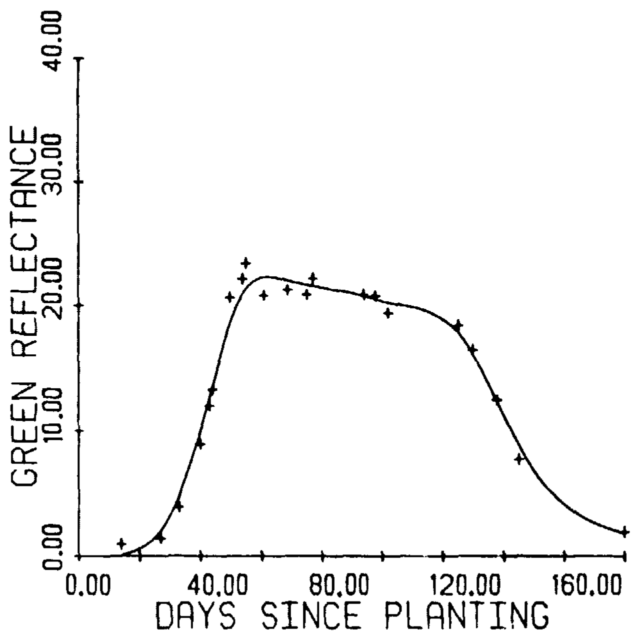
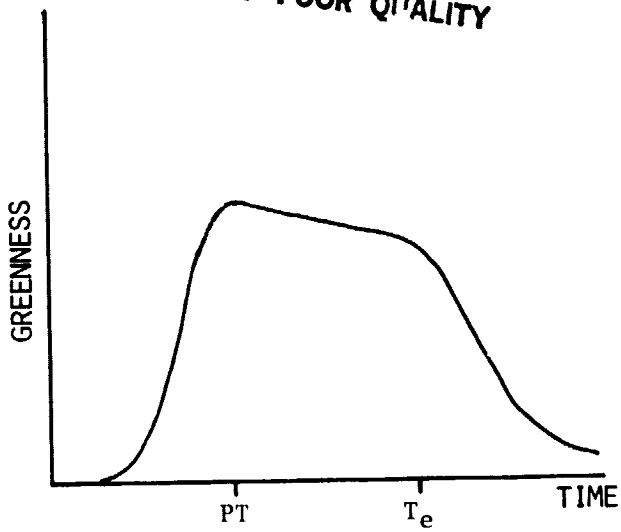


FIGURE 2.10. TYPICAL CORN GREENNESS PROFILE SHAPE



1) T_e = Time of plateau end

2) Duration of plateau = T_e - PT

3) Slope of plateau = $\frac{PMAX - G(T_e)}{PT - T_e}$

FIGURE 2.11. ADDITIONAL FEATURES FOR CORN GREENNESS PROFILE ANALYSIS

something of a sample from the range of possible approaches. Because the corn Greenness profile is a more complex shape and therefore a more challenging problem for curve-fitting, corn data were used in the comparison of curve-fitting approaches. The simpler nature of the soybean Greenness profile can be well described with a number of techniques.

Six techniques were evaluated: polynomial regression, least squares approximation using cubic splines with variable knots, a cubic smoothing spline, a non-linear filtering algorithm developed at ERIM called the Rolling Ball algorithm [29], a three-parameter profile model originally developed for small grains [30,31], and a five-parameter model developed at ERIM specifically for corn.

Evaluation of the techniques took a number of forms. All the techniques were applied to the set of corn reflectance data described in Table 2.8 of Section 2.5.3 (118 total plots from 3 years), with the previously described set of profile features computed in each case. Evaluation criteria included overall performance and stability, residual errors, ability to detect significant treatment effects on the experimental data, and ability to reproduce the flattened peak of corn.

It should be noted that the spline techniques and the Rolling Ball algorithm, as well as the polynomial technique to some extent, are usually used in an interactive mode, with parameters tuned for each individual curve fit. However, to be of use in the evaluation of many plots (as in this application), the techniques must be automated. Thus the degree of the polynomial, number and spacing of knots, smoothing parameter, and ball diameter sequence were all fixed, based on results of a more intensive interactive application of the techniques to a subset of the data.

<u>Comparison of Techniques</u>. While all the techniques tended to detect most of the same treatment effects in the profiles, the profile models, or at least the non-linear least squares techniques used to fit them, were more likely to fail in attempting to find a solution for any individual data set. All the other techniques successfully fit most or all of the data. Figure 2.12 provides an example of results obtained using the six curve-fitting techniques on the same set of data; residual errors are plotted vs. time from estimated peak in Figure 2.13 for the entire data set analyzed. These data provide a clear example of the flattened peak of corn, and include observations spaced throughout the growing period of the crop. The results displayed illustrate many of the findings of the curve-fitting comparison.

First, both polynomial regression and least squares approximation by cubic splines with variable knots tended to catch some of the flatness, but included extra loops or dips, particularly in the tails of the profile. Reducing the complexity of the curves (degree or number of knots) eliminated these extra slope inflections, but also reduced the ability of the functions to reproduce the flattened peak.

The Rolling Ball algorithm avoided the dips or ringing at the tails, but tended to smooth out the fairly sharp corners associated with the beginning of the flattened peak. The 5-parameter or Corn model, on the other hand, tended to produce too sharp a corner and, in addition, tended to overestimate data values early in the season (not as clearly illustrated in this particular plot, but readily apparent in the residual plots in Figure 2.13(f). The simple 3-parameter or Wheat model failed to provide a flattened curve, since it has no mathematical mechanism to allow for such a result. This shortcoming is highlighted in Figure 2.12(e).

Of the six techniques evaluated, the cubic smoothing spline algorithm produced the most intuitively appealing results, captured the flattened peak most often, and accurately fit the data throughout the season.

ORIGINAL PAGE IS
OF POOR QUALITY

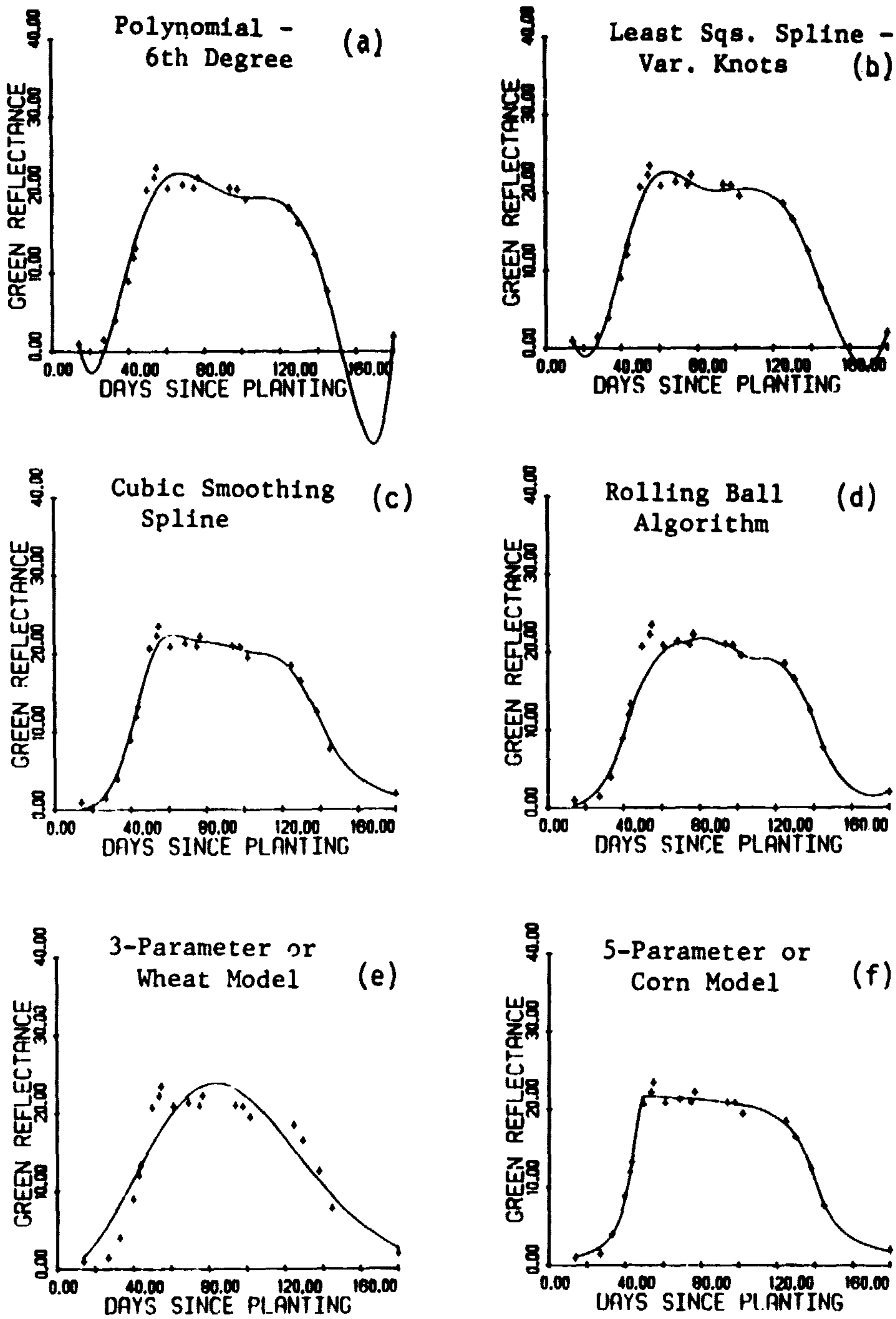


FIGURE 2.12. EXAMPLE CURVE FITS - PLOT 44, 1979 CORN CULTURAL PRACTICES EXPERIMENT

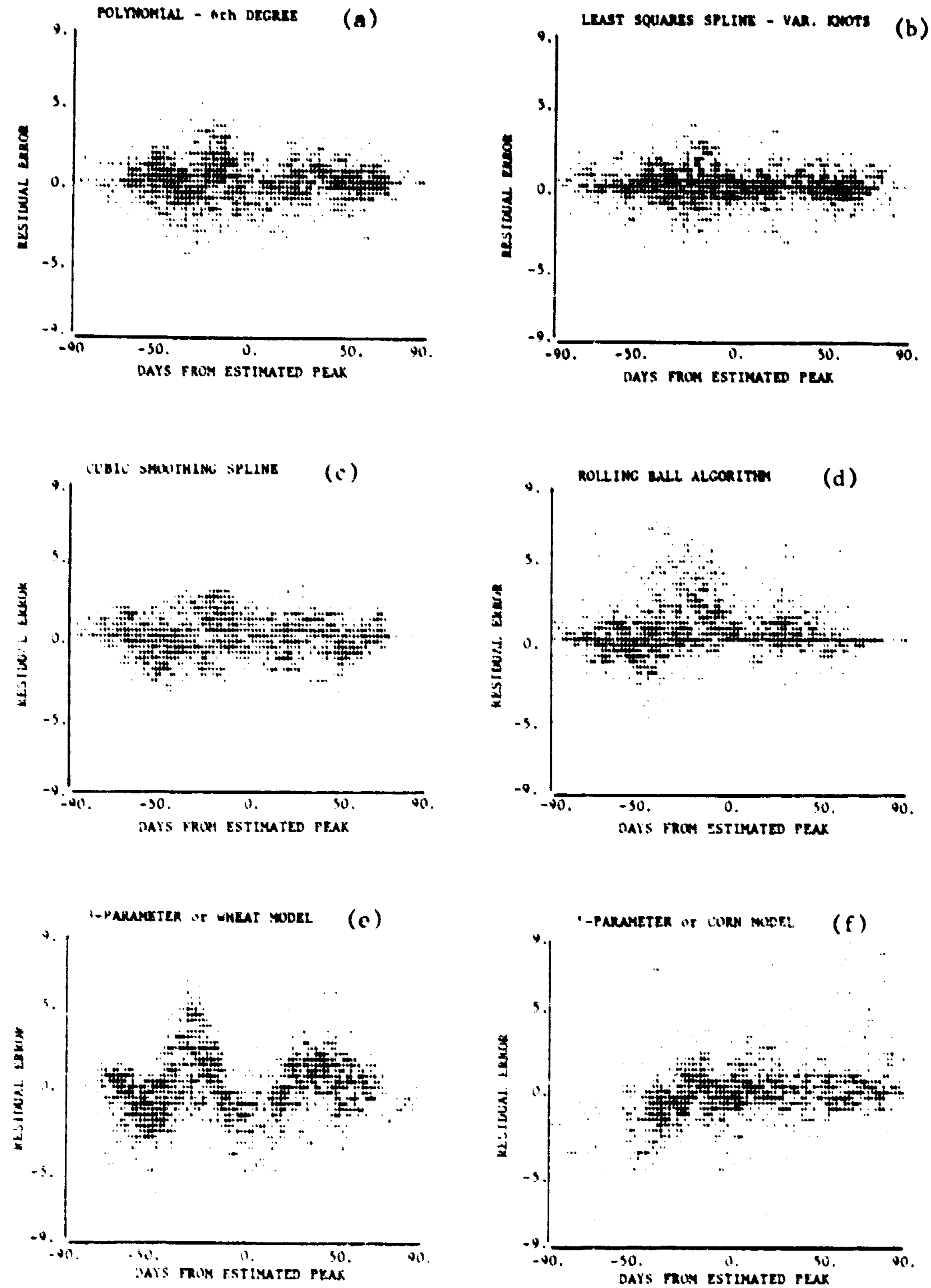


FIGURE 2.13. RESIDUAL ERRORS FROM CURVE FITS - 1979 AND 1980 CORN DATA

As a result the cubic smoothing spline was selected for use in subsequent analyses of field reflectance data. The same cubic smoothing spline technique was evaluated, in a more abbreviated fashion, for the soybeans data, and found acceptable. In the analyses reported in the following sections, all curve-fitting was done with this technique.

2.5.3 CULTURAL AND ENVIRONMENTAL EFFECTS ON CORN AND SOYBEANS SPECTRAL DEVELOPMENT PATTERNS

The curve-fitting technique described in Section 2.5.2 was applied to reflectance data collected over corn and soybeans plots by and at Purdue/LARS. Included were data collected using an Exotech 100 Landsat band radiometer as well as data collected using an Exotech 20C spectroradiometer. Exotech 20C data were converted to Landsat band reflectances by multiplying by Landsat sensor relative spectral response curves and integrating over wavelength. Multiple observations of a single plot on a single day were represented by their mean.

In order to simplify analysis of the spectral data, and to provide spectral variables that are readily associated with physical phenomena, a transformation was used which captures the majority of data variability over agricultural regions in two variables. It was based on a transformation, derived for Landsat data, which is termed the Tasseled-Cap transformation [32], and produces two variables which typically contain more than 95% of the total data variation in an agricultural scene. Brightness, the first variable, corresponds to the spectral direction in which the majority of soil brightness variation is found. The second variable, Greenness, is orthogonal to Brightness, and is an indicator of the amount of green vegetation present in the scene.

A rotation of the principle components plane of the field reflectance data was used to provide Tasseled-Cap equivalent values. The final transformation determined to derive Tasseled-Cap equivalent variables from the raw Landsat band reflectances is:

$$\begin{bmatrix} .3298 & .3996 & .5910 & .6182 \\ -.4778 & -.6486 & .0932 & .5851 \end{bmatrix} \times \begin{bmatrix} \text{Refl. Band 4} \\ \text{Refl. Band 5} \\ \text{Refl. Band 6} \\ \text{Refl. Band 7} \end{bmatrix} = \begin{bmatrix} \text{Reflectance Brightness} \\ \text{Reflectance Greenness} \end{bmatrix}$$

A small degree of subjective data screening was also carried out. A few observations that were clearly abnormal were deleted, and several entire plots were deleted, either because they showed substantial noise overall or because they lacked acquisitions in a large and significant portion of the growing period. Elimination of plots with excessive noise or too few observations resulted in a data set consisting of 118 corn plots and 171 soybean plots in eight experiments from 1978 through 1980, as detailed in Table 2.8.

After applying the techniques previously described, a series of oneway analyses of variance was carried out to determine the significance of effects of the various experimental treatments on the derived profile features. The following sections provide a summary of the results of these analyses. Details may be found in Reference [25].

2.5.3.1 Corn Results - Summary

The effects of Nitrogen fertilization, planting date, and plant population were evaluated with regard to their impact on features of corn Greenness profiles. All were found to significantly affect the Greenness development of the test plots.

Addition of Nitrogen (which promotes vegetative development) to a plot increased the peak Greenness values and the length or duration of the flattened portion of the profile. Both of these effects are indicators of more lush, vigorous vegetation. A 25% (5 count) difference in peak Greenness was observed from lower to higher fertilization levels.

Planting date differences were spectrally expressed in the height and time of occurrence of the peak profile value. Later planting always

TABLE 2.8. CORN AND SOYBEAN REFLECTANCE DATA USED IN ANALYSIS

Year	Experiment Name	# Plots
1978	Corn Nitrogen	13
1979	Corn Nitrogen	9
1979	Corn Cultural Practices	34
1979	Corn Soil Background	10
1980	Corn Cultural Practices	52
1979	Soybean Management	69
1979	Soybean Cultural Practices	46
1980	Soybean Cultural Practices	56

caused the peak value to occur sooner, as emergence and early growth were promoted by warmer temperatures. The effect on the magnitude of the peak, however, was variable with time. Peak Greenness values increased from very early to more medium planting dates, probably as a result of the colder, less conducive environment encountered by the very early-planted plots. As planting was delayed later, peak Greenness values tended to decline again, probably an indication of the stresses encountered by later-planted crops in the heat of the summer. Peak Greenness variation was similar to that observed in the Nitrogen experiment, with 27% (4 counts) variation, while planting delays hastened the time of peak by as much as 15 days.

Plant population also affected the height and time of occurrence of the peak Greenness value. Increasing the number of plants per hectare resulted in an earlier peak value, a reflection of the increased competition and accompanying increase in development rate, and also produced a higher profile peak. The higher peak was most likely the result of increased Green biomass, and reduced shadow and soil background in the sensor field of view. Not detected was an earlier decline in Greenness, which would be expected when the increased competition and associated increase in growth rate causes the plants to use up the available nutrients and water. This may have been an indication of the favorable growing conditions encountered by most of the plots during most of the vegetative phase (the latest planting dates were not included in this analysis).

Population-related peak Greenness variation ranged from 41 to 62% (7 to 8 counts) in 1980, but only 22 to 32% (4 to 6 counts) in 1979. Variations in time of peak were 11 to 33% (9 to 18 days) in 1980, and 14 to 32% (10 to 23 days) in 1979. Other profile features were found to be significantly affected by population in only one of the years.

2.5.3.2 Soybean Results - Summary

The effects of variety, planting date, row spacing, and plant population on Greenness profile features were examined. All had some degree of impact, with population effects of least significance.

Soybean varieties differ considerably in growth habit, length of growing period, response to environmental changes, and other characteristics. Four varieties were available for comparison including samples from two maturity groups, a semi-dwarf determinate variety, and a "thin line" variety.

Although a seasonal effect was evident between 1979 and 1980, the class III (later maturing) varieties generally showed a slower Greenness decline than the class II (earlier maturing) varieties. The semi-dwarf, determinate, class III variety reached higher peak Greenness values and exhibited a more rapid green-up rate than the larger, indeterminate, class II varieties. The bushy class III variety also achieved a higher peak than the thin line class II variety. In addition, differential responses to row spacing and plant population were noted and are discussed later. Varietal peak Greenness differences ranged from 6 to 12% (2 to 4 counts), and occurred as much as 5 days apart.

These results are consistent with the described characteristics of the varieties. The later-maturing varieties stayed green longer, the more compact semi-dwarf cast fewer shadows and thus reached a higher peak Greenness, and the bushy varieties filled in the space better than the thin line variety, and so achieved a higher peak value.

Planting date effects are, as previously indicated, strongly connected to temperature and its effects on emergence and vegetative development. Later planting tended to increase peak Greenness values, although very late planting was accompanied by a reduction in the profile peak. The time of peak was substantially influenced, occurring much earlier for later planted plots. Some indication of a reduced effect on maturity date as compared to vegetative development was seen in a lengthening of the Greenness profile after the peak for later planted plots, as would be expected. Planting-date-related variation in peak Greenness was about 16% (5 counts), while plots planted in early July reached their peak value in 42 fewer days than those planted in early May.

Increasing the row spacing in a soybean plot reduced peak Greenness, since more soil and shadow was in view. The rate of green-up was reduced, and the rate of Greenness decline increased, again largely due to the percentage of the field of view occupied by non-green components. A hastening of the time of peak Greenness was observed with narrower rows. This was probably due to an earlier achievement of complete canopy closure. If so, it should be noted that for soybeans, the time of peak Greenness cannot be clearly associated with any particular development stage. Varietal differences were observed. Peak Greenness values varied some 12% (4 counts), with 8 to 11 day delays in the profile peak.

The impact of population should be of a similar nature to that of row width. However, possibly as a result of the soybean plant's tendency to fill in the available space, very little effect was detected. Peak Greenness values tended to increase with population, but the variability present at the highest populations rendered the increase statistically insignificant.

2.5.3.3 Evaluation of Curve-Fitting Technique

Overall, the technique described in Section 2.5.2 performed as desired. The cubic smoothing spline technique fit the soybean data, and much of the corn data, very well. The extraction of standard profile features allowed ready comparison of plots with different planting and/or observation dates, and characterized the continuous profile in a manageable number of variables. With these variables, quantitative analysis of experimental effects was greatly facilitated.

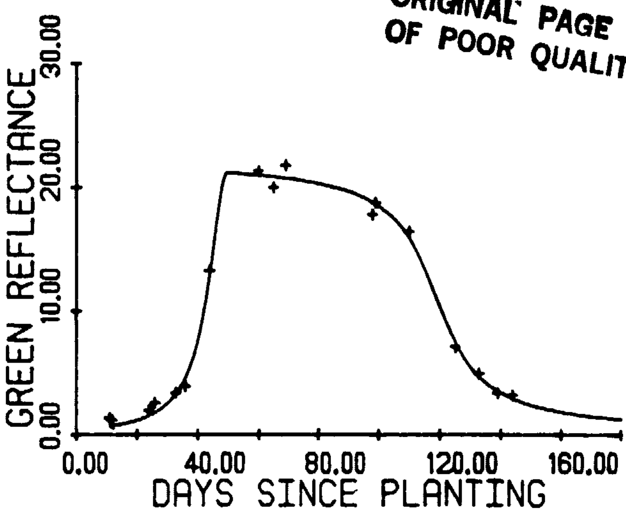
In the course of analysis, two improvements to the procedure were suggested. First, even the cubic smoothing spline algorithm failed to detect the flat peak of corn data when insufficient data points were available, especially when the sparse data occurred just before or on the plateau. Given the expectation of a flattened peak, one could often see such a feature in the data when the spline technique had not.

The 5-parameter corn model, which is designed to function with a similar expectation, also detected flat peaks when other techniques did not (Figure 2.14 provides an example), although that model had other weaknesses. Most desirable would be a curve-fitting function with the flexibility of the cubic smoothing spline, but also the prior expectation of crop development that would allow it to draw a "corn-like" or "crop-like" profile even with sparse data. Development of such a function would greatly increase the power of this analysis technique for corn data.

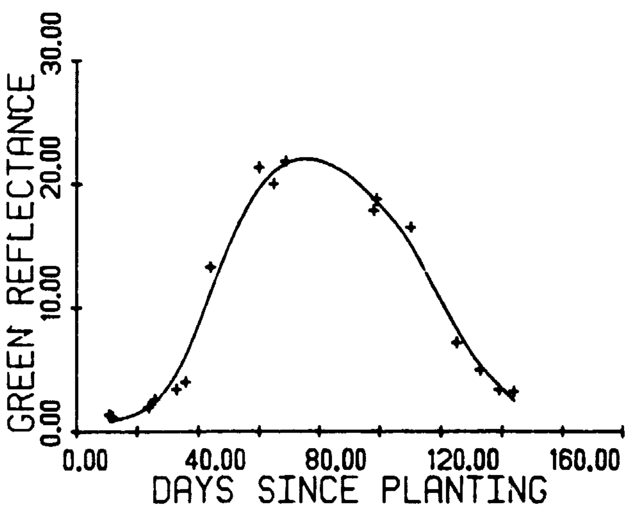
The second suggested modification to the analysis technique regards the rate-related features. As described, half-peak values are used as critical points in measuring time intervals. However, in some cases it appeared that treatment effects were missed because of significant increases in the peak value, which of course resulted in increased half-peak values. Time intervals related to half-peaks were thus based on the achievement of substantially different Greenness thresholds, and rate differences between treatments were, at least to a degree, normalized. While half-peak values may provide useful information, rates might better be computed, or at least also be computed, based on fixed thresholds, i.e., compute absolute rates of change in Greenness $\frac{\Delta G}{\Delta t}$ as opposed to relative rates $\frac{\Delta G/G_{max}}{\Delta t}$.

2.5.4 CONCLUSIONS AND RECOMMENDATIONS

The analyses of field reflectance data presented in the previous sections provide a clear indication that a number of commonly varying



a) 5-Parameter or Corn Model



b) Cubic Smoothing Spline

FIGURE 2.14. COMPARISON OF CURVE-FITTING TECHNIQUES - DETECTING THE FLATTENED PEAK OF CORN

field characteristics can exert a substantial influence on the spectral appearance of crops. Such key features as the maximum Greenness value and rate of green-up can be altered significantly by varying any one of a number of parameters including Nitrogen fertilization, planting date, variety, and plant spacing. In a real-life situation where any or all of these characteristics may vary, the likely effects on crop spectral appearance will be considerable. Such variability must be taken into account in any crop identification technique, whether carried out by human analysts or computer algorithms. In addition, this type of information is of critical importance in the design and implementation of accurate, useable simulation systems.

The work presented is, however, only a first step. Expanding the Greenness profile analysis for corn to include the new features described in Figure 2.11, which specifically relate to the flattened peak or "plateau" observed in corn Greenness data, and applying a similar analysis technique to the understanding of Brightness profiles and their sensitivity to cultural and environmental factors, will provide still more insight. The derived profile features could also be used to determine, again on a quantitative basis, the similarities and differences between corn and soybeans profiles, and the effect of the various treatments on their separability.

Finally, of course, the insights gained through field data analysis must be applied to real Landsat data. The loss of control over crop parameters, the inclusion of an atmosphere, the degradation of resolution, and the mixing of the independently evaluated factors, as well as others not even considered, will likely cause some of the observed and/or predicted effects to be reduced, while others will be intensified.

Controlled experimentation provides a foundation and a context, but it cannot completely replace real data, nor can crop inventory techniques be derived from field data alone. It is the progression from physiological understanding through modeling and field data analysis to Landsat data analysis that brings the experimental data and understanding into the real world, while at the same time anchoring the uncertain real world to some reliable and stable points of reference.

2.6 SIMULATION, MODELING AND ANALYSIS

Simulation models are designed to capture one's best understanding of how the "real world" operates and can be used for many purposes. They can help rank the importance of multiple factors, predict the nature of responses to those factors individually and in concert, help in analysis of existing measurements and empirical data sets, make predictions for unmeasured conditions and situations, and guide the specification of new measurement and analysis efforts. They can be used in the design of new sensors and to develop preliminary analysis procedures and predictions of performance in advance of new sensor operations.

Past simulation models have not adequately represented the full range and character of factors that affect remotely sensed data. For example, in agricultural applications such as AgRISTARS, the effects of crop physiological parameters, meteorological variables, and atmospheric and sensor characteristics on spectral observations currently are not well enough understood. Field measurements are not practical under all the observation conditions and situations necessary to fully explore the nature and range of variation, so improved simulation models are appropriate.

This section describes three substantial developments in simulation modeling capability. The first two relate to a simulation tool that ERIM is developing named the "Seed-to-Satellite Model" [33]. Its purpose is to help analysts better understand factors that affect the observable spectral responses of crops, analyze data sets that have been acquired by Landsat, and develop improved information extraction techniques. It has modules to model crop reflectances, atmospheric effects, and sensor spectral responses, modules that have been used in previous analyses [34,35]. It can also help in preparation for Thematic Mapper data and data from other sensors.

PRECEDING PAGE BLANK NOT FILMED

The first development involved incorporating, for the first time, a meteorologically driven, physiological growth model for a crop and interfacing it with a bidirectional reflectance model for vegetation canopies.

The second substantial development was modification of the Suits bidirectional reflectance model for vegetation canopies to incorporate row effects as observed in many agricultural crops.

The third development was of a capability to simulate the spatial and spectral effects of Landsat when viewing agricultural scenes. This capability includes representation of the temporal-spectral profiles of crops and variations of planting dates and crop vigor on a field-by-field basis. It also incorporates the full two-dimensional point-spread function of the Landsat MSS to permit detailed simulation and analysis of mixed pixels and field boundary effects.

2.6.1 SIMULATION OF THE SPECTRAL APPEARANCE OF WHEAT AS A FUNCTION OF ITS GROWTH AND DEVELOPMENT

The objective of this simulation was to provide an understanding of the connection between important agronomic features of an agricultural crop and the satellite signals that are received from that crop.

The agronomic features of general interest are crop type, crop vigor, and ultimate yield at the end of the growing season. On the ground, the crop type can be determined from the taxonomy of the individual plants. Crop vigor and yield predictions can be inferred from the size and morphology of the plants and the size, number, weight, and color of plant components - such as, leaves, stems, flowers, and heads of grain. The same plant components and plant morphology also partially control the signals received by satellites, by way of their radiometric properties.

A simulation, which incorporates a physiological growth model for a crop as an intermediary, can supply that output signal which can be used for vigor and yield estimates as well as estimates of plant component number, sizes, color and morphology for signal calculations that are important for crop identification procedures. Laboratory measurements of the radiometric properties of actual components, a canopy reflectance model and atmospheric scattering model can then be used to predict the corresponding signals received by the satellite. In this way, the connection between agronomic features and satellite signals is made by means of the growth model and the other models.

During the reporting period, the problem of incorporating a crop physiological growth model into the Seed-to-Satellite Model and interfacing it to the Suits reflectance model was addressed. Wheat was selected as the first crop to be investigated.

2.6.1.1 Summary Description of the Simulation for Wheat

The block diagram showing the logical structure and information flow through the wheat simulator is shown in Figure 2.15. The wheat growth model is the November 1979 version by Ritchie [36]. The growth model requires a number of input parameters representing genetic influences, environmental influences (soil-moisture and weather parameters), and planting density. Growth occurs through several stages that can be identified with Feekes scale numbers. The day-by-day outputs of the growth model are green leaf area index, number of active tillers, change in leaf area, and grain weight (where appropriate).

Since all of the plant components which are radiometrically significant are not supplied as outputs, a canopy geometry interface is required to complete the physical description of the crop. For our purposes, we derived quantitative relationships from field data collected for wheat by Jackson and Pinter [37] and scaled them to the

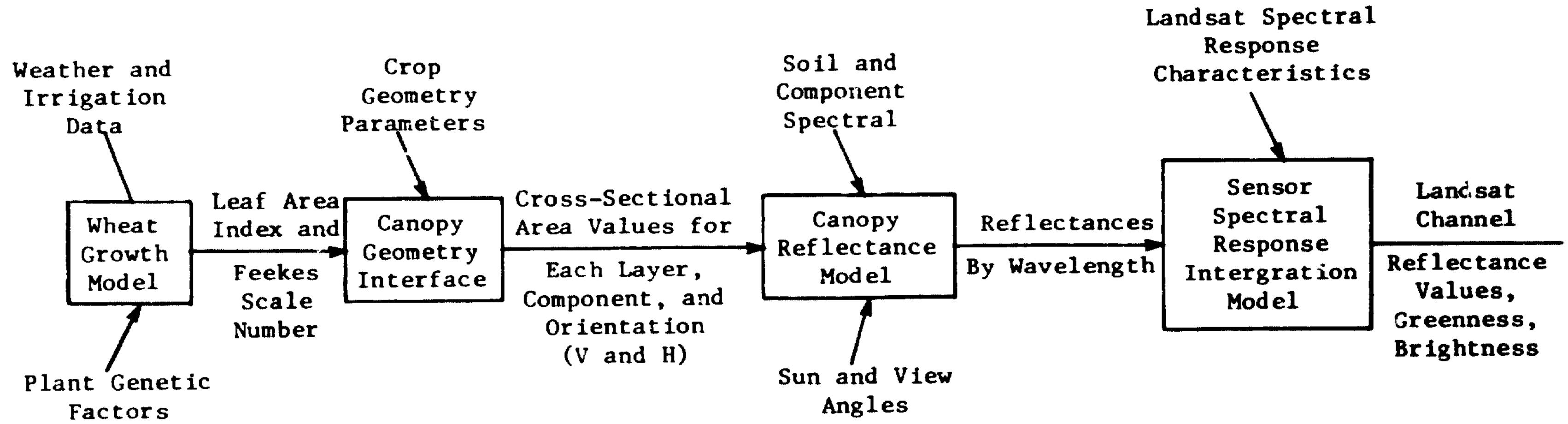


FIGURE 2.15. INFORMATION FLOW THROUGH THE WHEAT REFLECTANCE SIMULATOR

growth model LAI and active tiller number output at the equivalent Feekes scale of the growth model. New leaves differ spectrallly from old leaves. Sloughed off dead leaves and some dead tillers are part of the growth process. They are radiometrically important and do not disappear from the field but, rather, record by their presence the characteristics of the growth process predicted by the growth model.

The spectral properties of the components of the wheat plant were obtained from laboratory measurements made previously at ERIM of samples of Kansas wheat. There are likely to be some varietal differences in such spectra, particularly between wheat suited to different moisture conditions. An average soil spectrum from measurements made by Condit was utilized. The spectra of soil upon which the crop is planted is often an important cause of crop reflectance variation which is purely coincidental with the crop development. Such variation can make the connection between agronomic features and received signals more obscure.

The size, number, orientation and spectral properties of the plant components are the inputs to the canopy reflectance model. The uniform canopy reflectance model of Suits [38] was used in this simulation; three layers were employed. A fixed sun angle and a nadir view angle were used for the simulation parameters for Landsat.

The atmospheric scattering model has not yet been introduced into this simulation. The spectral responses of the Landsat channels were used to determine the relative signal values which would be received by Landsat if perfect corrections were made for atmospheric attenuation and path radiance. These signal values were also converted into reflectance-space-equivalent Tasseled-Cap transformed signals, i.e., Reflectance Brightness and Reflectance Greenness.

2.6.1.2 Initial Results of Simulation

Simulation of a single growing season for wheat was made. The time locus of points on the reflectance Brightness-Greenness plot shows the characteristic path of wheat in the Tasseled-Cap plane; in Figure 2.16, Feekes scale indications are given for selected times. Feekes 2 is the beginning of tillering where the vegetation cover is nearly undetectable. The progression through tillering and stem extension to Feekes 9 corresponds to the rapid vegetative growth of the canopy. Between Feekes 9 and Feekes 10 the LAI continues to increase with the flag leaf at the top of the canopy becoming fully extended and mature.

Between Feekes 10 and Feekes 11 the wheat goes from boot stage to a full development and extension of the head over the flag leaf. At Feekes 11, stem and head fill the top layer (Layer 1) of the canopy, stem and mature leaves occupy Layer 2, the next layer down, and dead leaves, any dead tillers from previous growth, and active green stem occupy Layer 3 next to the soil. This particular growing season's weather into the model resulted in very little dead tissue in this lowest third layer. From Feekes 11 to 11.4 the heads ripen and the wheat leaves and stem die and change color. Feekes 12 represents the harvested field where only the dead stubble and tissue in Layer 3 remain. Everything above Layer 3 has been cut and carried away. The position of Feekes 12 will change with harvesting practice.

The sharp cornered transitions in the plot are artifacts of the simulation where all wheat in the field developed in perfect synchronism and the leaves died off abruptly. In actual fields, there is a spread in the stages of development which will cause these corners to be rounded. This spread will have to be introduced in later simulations.

The simulated maturation of the new and upper leaves of the canopy between Feekes 9 and Feekes 10 also contributed to the path shown on

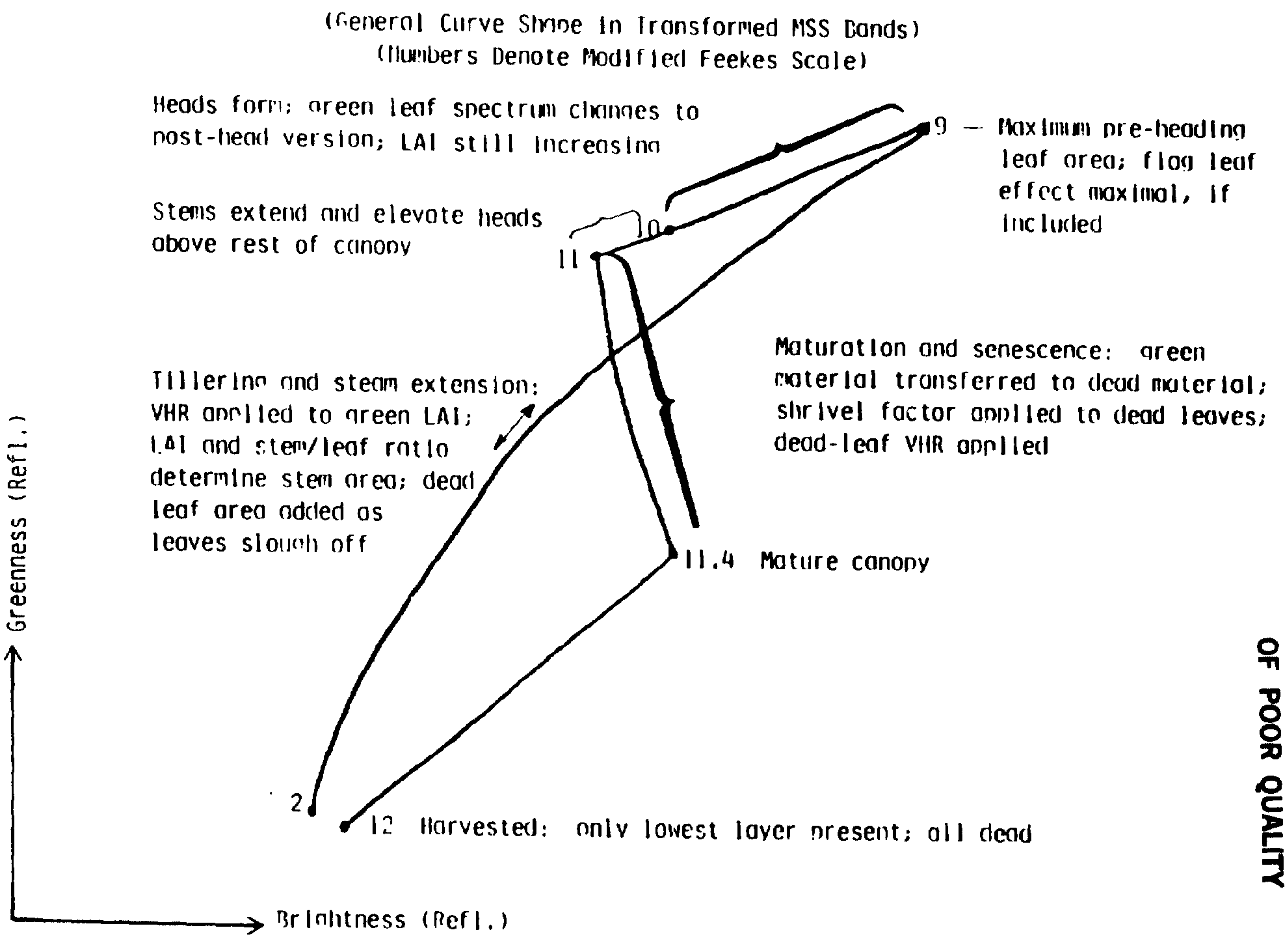


FIGURE 2.16. PHENOMENA ASSOCIATED WITH WHEAT REFLECTANCE CHANGES

the Tasseled-Cap diagram. The choice of when and how fast to change from immature to mature green leaf spectral properties is a decision required for the simulation. Unfortunately, the growth model is mute on the significance of this detectable transition. The yield predictions by the growth model depend, in part, upon total leaf area rather than leaf area in a particular portion of the canopy. The remotely sensed signal on the other hand depends largely upon the scattering in the upper portion of the canopy when LAI is near maximum. In subsequent simulations, we have let leaves mature in a fixed number of days after their emergence.

Figure 2.17 shows one of several parametric studies we performed on various canopy parameters. This was a study of the effect of head size on the time trajectory in the Tasseled-Cap plane. The head length of 8 cm was taken as "normal" and a variation in head length from 6 cm to 10 cm shows that the variation in the Feekes 11 to 11.4 transition is clearly affected. The results emphasize that even small plant components which are consistently located at the top of the canopy have a much larger effect than one might suspect purely from the size and number of such components.

2.6.1.3 Summary and Conclusions

A number of other parametric studies were made to determine the sensitivity of each separate parameter upon the Tasseled-Cap and MSS5-MSS7 plots of Landsat-equivalent reflectances. The variation of each parameter revealed the timing and magnitude of the variation in Landsat signals which could be expected. However, while each parameter in the canopy geometry interface was at all times consistent with the Growth Model outputs, the Growth Model outputs were insufficiently detailed for determining all of the needed parameters in the Canopy Geometry Interface. Consequently, we had to use the empirical

ORIGINAL PAGE IS
OF POOR QUALITY

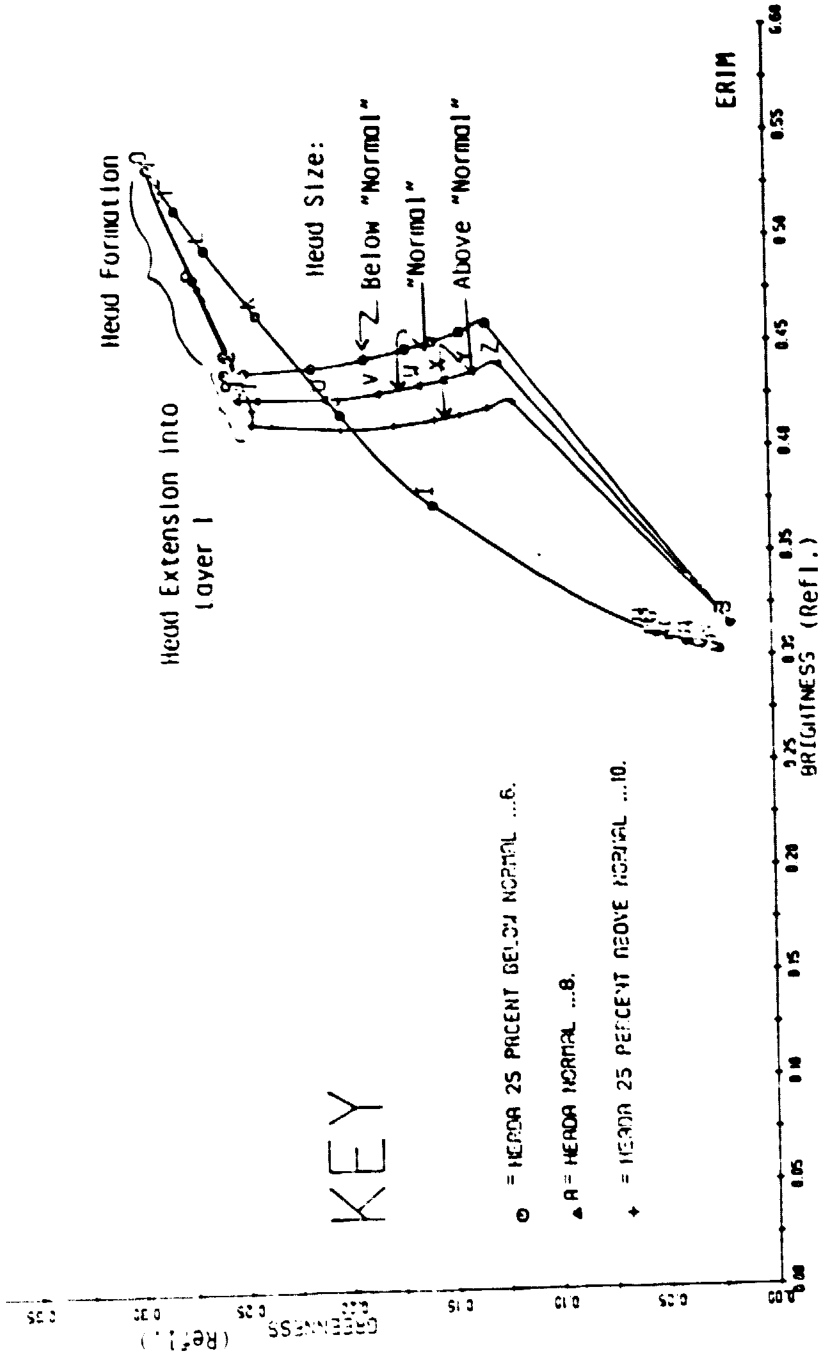


FIGURE 2.17. EFFECT OF HEAD SIZE ON SIMULATED REFLECTANCE OF WHEAT CANOPY

observations of Jackson and Pinter with scaling rules to complete the canopy description.

Clearly, the Canopy Geometry Interface is the weak link in the simulation. The parameters in the Interface should be causally connected to the same physiological growth process as are the agronomic features but the growth model was designed to predict the agronomic features rather than the concurrent expression of the growth and conditions of plant components within the canopy that control satellite signals. The causal connection between the Growth Model and Geometry Interface is incomplete.

The situation is comparable to the position of a practicing physician who utilizes various symptoms of the patient to arrive at a diagnosis of a disease. Medical research could fully explore and understand the disease process and the manner in which the disease causes destruction of vital organs - the central issue of course. Yet the same disease process can also produce concurrent symptoms which, by themselves, are not directly involved with the destruction of vital organs but could be used as causal connections or symptoms of the impending destruction. If medical research ignores the latter connection, the physician has little or no diagnostic power.

We, in remote sensing, are attempting to diagnose agricultural fields using satellite signals as symptoms. Our modeling has traced the causal connection down to the Canopy Geometry Interface but the growth model, addressing the central issues of economics, fails to connect completely the growth process with the concurrent features which we use as symptoms. Our use of empirical observations and scaling relations are not necessarily causally connected to the growth process.

2.6.2 THE EXTENSION OF A UNIFORM CANOPY REFLECTANCE MODEL TO INCLUDE ROW EFFECTS

2.6.2.1 Introduction

Many crops are planted in rows by machinery. Upon emergence of the plants, the bare soil between rows is still the dominant feature which reflects incident daylight. As growth continues, the vegetation grows both higher and spreads out over the inter-row regions, covering the bare soil. At some time during the growing season, the soil is covered enough that the bare soil between rows is no longer a dominant feature. The vegetation canopy becomes essentially laterally uniform in its radiation scattering properties. The alteration of incident daylight can be understood and calculated by a previously developed uniform canopy reflectance model [39] at this stage of growth.

However, for a considerable time during the early part of the growing season, the strips of bare soil between rows and the increasing density of vegetation along the rows become equally important in their contributions to canopy reflectance. One may intuitively understand that the direction of sunlight relative to the row direction will change the relative influence of vegetation and bare soil. When the sun is directed along the row direction, the bare soil is fully illuminated but, when the sun is directed across rows, the soil is largely in the shadow of the standing vegetation along the rows. Thus, Landsat can receive different signals due only to the way the rows trend relative to sunlight. An inference that such altered radiation is due to a change in some important agronomic feature could be in error.

The following text reviews the concepts, nomenclature, and symbols of the uniform canopy model in order to form the logical basis for its modification to incorporate the "row effect". The concept of density modulation is introduced to account for the row structure of a canopy and the manner of calculation using such a concept is described.

The extended model is applied to wheat in rows. The results are similar to those of field measurements. The red band, Landsat MSS Band 5, is most sensitive to row direction because of the usual large contrast between vegetation and soil. Reflectance in this band may easily vary by a factor of two with changing row direction. The IR bands, Landsat Bands 6 and 7, are least affected by row direction because of low contrast between soil and vegetation and because of the large amount of diffuse flux scattered to soil by the vegetation.

2.6.2.2 Review of the Uniform Canopy Model

The uniform canopy reflectance model consists of a number of infinitely extended horizontal layers or strata as illustrated in Figure 2.18. Within each layer, the plant components of the canopy are considered to be randomly distributed and homogeneously mixed. The plant components are the identifiable parts of the plant, such as, stems, leaves, branches, flowers, and pods or heads.

Collimated radiation from the sun enters the top of the canopy. This collimated flow of radiation is called specular flux in the following text. That specular flux which is intercepted by a plant component is diffusely scattered and partially absorbed. The remaining specular flux, steadily diminished by such scattering, proceeds on to the soil making "sun flecks" upon the soil surface.

The diffuse flux created by scattering may be produced by reflection from a component or by transmission through a component. Some of the diffuse flux is scattered towards the top of the canopy; the remainder is scattered towards the soil. As the diffuse flux moves through the canopy, some of the diffuse flux will be intercepted and scattered again with some of the rescattered flux going up and some going down and so forth.

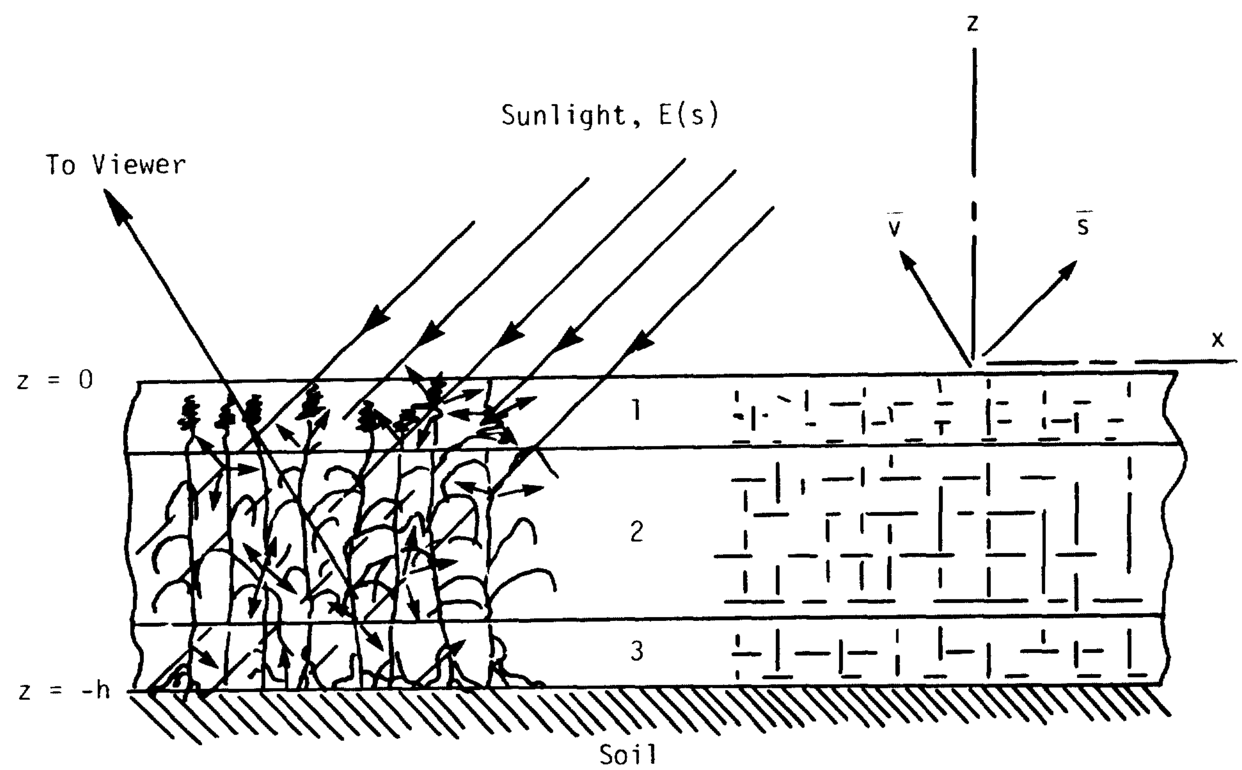


FIGURE 2.18. UNIFORM CANOPY MODEL. The actual crop is illustrated on the left. The corresponding model equivalent components are shown on the right. Vertical lines represent the vertical components oriented azimuthally at random. Horizontal lines represent horizontal components.

The lateral average flux density on a horizontal plane, of specular flux and upward- and downward-welling diffuse flux, varies with depth in the canopy. Allen, Gayle, and Richardson [40] showed by experiment that the flux densities could be derived using Duntley's differential equations for scattering in diffuse optical media. The scattering properties of any particular medium are specified by the values assigned to five independent parameters in these equations. These differential equations are shown in relations (5), (6), and (7),

$$dE(+d)/dz = -aE(+d) + bE(-d) + cE(s) \tag{5}$$

$$dE(-d)/dz = aE(-d) - bE(+d) - c'E(s) \tag{6}$$

$$dE(s)/dz = k(Es) \tag{7}$$

where $E(+d)$ = upward welling diffuse flux density,

$E(=d)$ = downward welling diffuse flux density,

$E(s)$ = specular flux density

a = extinction coefficient for diffuse flux,

b = backscattering coefficient for diffuse flux,

c = backscattering coefficient for specular flux,

c' = forward scattering coefficient for specular flux,

k = extinction coefficient for specular flux.

The five parameters, a, b, c, c', and k for each layer plus the boundary conditions of soil reflection at the bottom and sunlight at the top are all that is needed to specify how much flux goes which way. What remains unknown is the relationship between these parameters and the plant components that are present within the canopy.

The uniform canopy model provides a systematic and logical method of calculating approximate values for these parameters given the number, orientation, and spectral properties of the plant components in a canopy. This method conceptually replaces a particular plant component with three-plane orthogonal projections of that component. Each plane projection (hereafter called a model equivalent component) is assigned the same hemispherical spectral reflectance and transmittance as that of the actual plant component. The concept of projections is illustrated in Figure 2.19.

The five unknown parameters can now be calculated using model equivalent components.

Equipped with the values for the five parameters for each layer, one may solve relations (5), (6), and (7) for each layer and, hence, for the flux within the canopy. This flux is the illuminant for objects within the canopy which one can see from some direction of view. The final computation now is simply to determine the radiance, L, (radiometric brightness) of each component in the canopy and what fraction of these components can be seen without obstruction. The model equivalent components are again used to calculate the expected radiance of the components.

The reflectance is the ratio formed by dividing πL by the irradiance on the top of the canopy.

2.6.2.3 Extension to Include Row Effects

The fundamental concepts, nomenclature, and procedures of the uniform canopy model will be used with certain modifications to incorporate the effects of a row structure in agricultural crops. These modifications are introduced in such a way as to reduce to the uniform canopy model as row structure disappears from the crop due to overgrowth of the area between rows by the natural growth of the crop during the growing season.

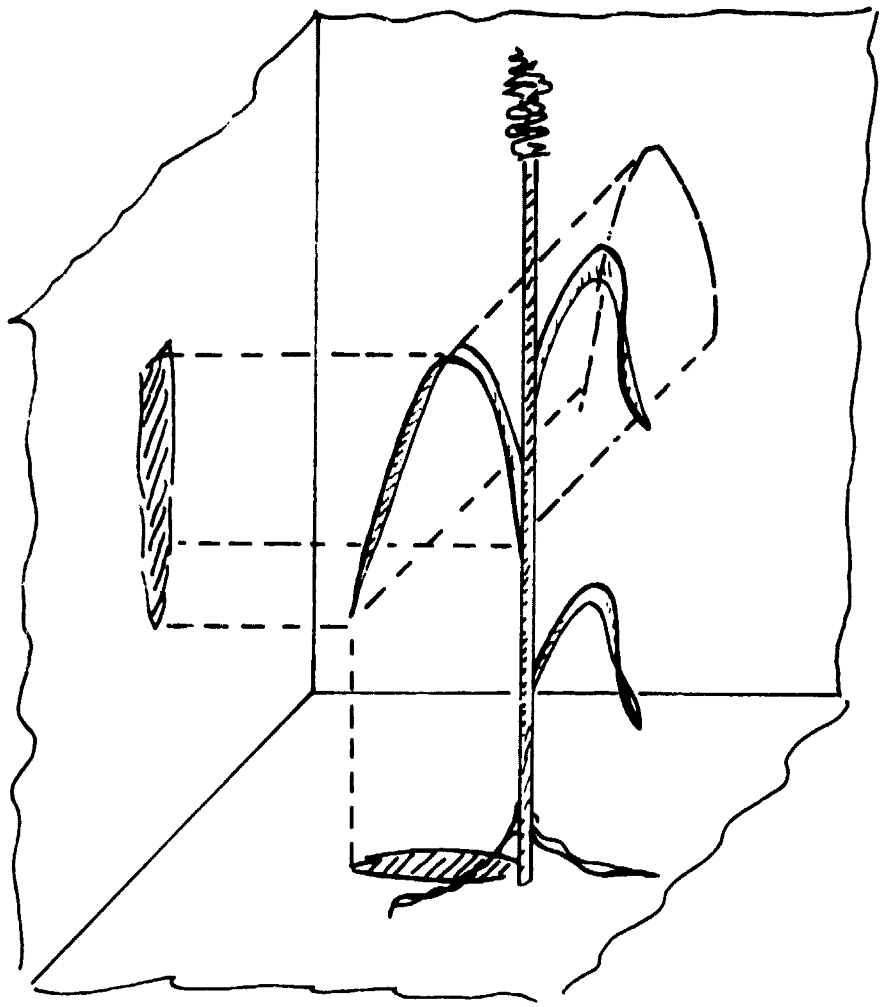


FIGURE 2.19. CONCEPT OF MODEL EQUIVALENT COMPONENTS.
Three orthogonal projections of a leaf component are shown for illustration.

The Concept of Density Modulation. In the uniform canopy, the density of components are the mean values for a patch of field the size of the instantaneous field of view (IFOV). Locally, the densities can be expected to vary due to the randomness of the distribution. Random distributions are expected to be clumpy but without any order as to where the clumps occur. One could consider any narrow strip of field and determine the mean density of components within that strip. The mean density would be the same as the IFOV mean, given sufficient strip length for any direction the strip might take over a uniform canopy.

However, in the case of a canopy with row structure, the strip mean will converge to a different mean density for strips parallel to the row direction depending upon the lateral displacement, δ, of the strip from the row center. The variation of strip means would be periodic for displacements of the strip in the across-row direction with large values on the row centers and small values between row centers. This variation in strip means, $M(\delta)$, relative to the IFOV mean is hereafter called density modulation. Density modulation is the evidence for the existence of row structure and is the measure of the amount of row structure.

Computation Method. In the extension of the uniform canopy model, the density modulation will be the same for all layers so that a particular profile would not be evident to the eye as illustrated in Figure 2.20. The use of the same density modulation, $M(\delta)$, for all layers simplifies the calculations but should still lead to the essential features of the row effect on canopy reflectance.

Let the five parameters, a, b, c, c', and k, be the IFOV mean values. Then the five parameters for strips required for row structure must be simply the IFOV means multiplied by the modulation, $M(\delta)$, since all parameters vary in direct proportion to component density.

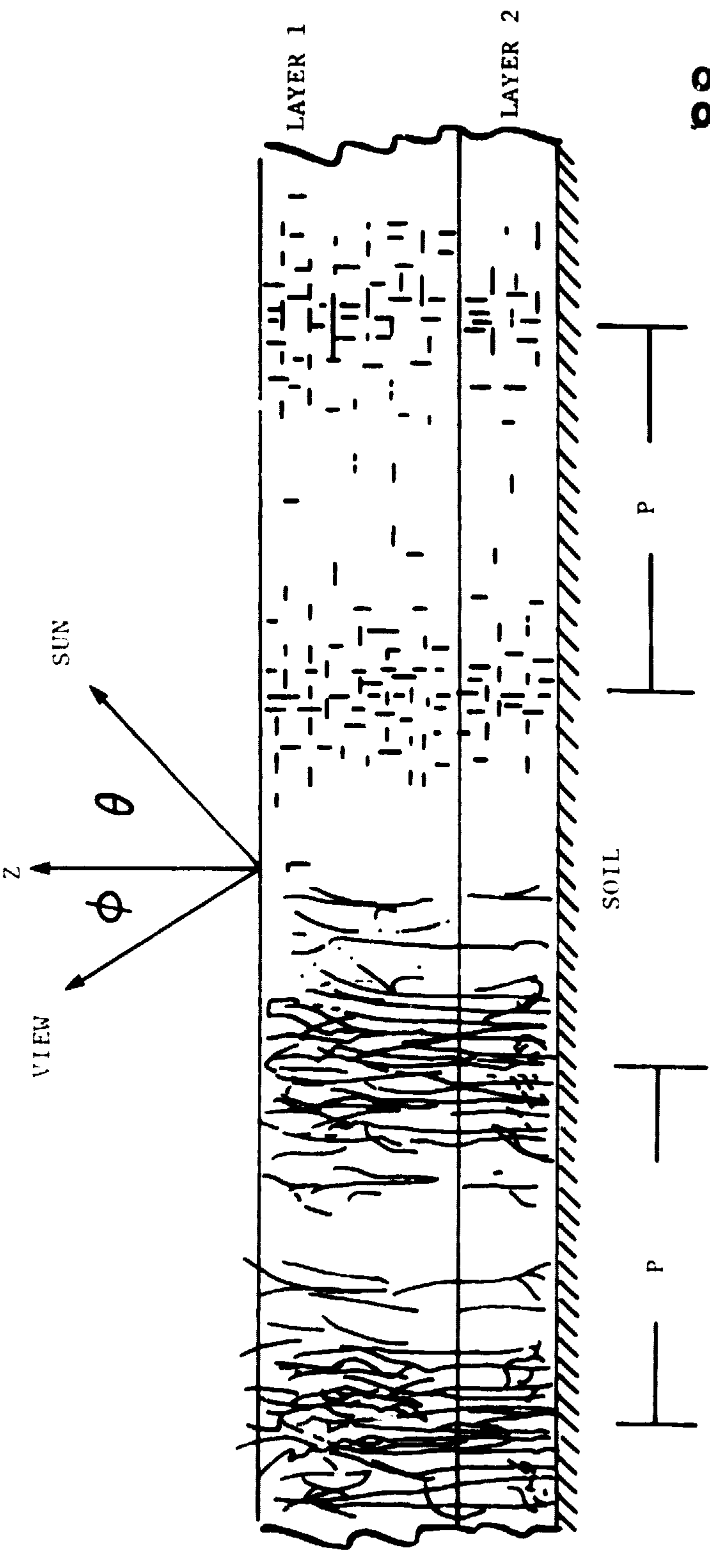


FIGURE 2.20. SCHEMATIC OF ROW CANOPY

Now, using the same differential equations as before but with the five parameters required for row structure, one obtains

$$dE(=d)/dz = -M(\delta)\ aE(+d) + M(\delta)\ bE(-d) + M(\delta)\ cE(s) \qquad (8)$$

$$dE(-d)/dz = M(\delta)\ aE(-d) - M(\delta)\ E(+d) - M(\delta)\ c'E(s) \qquad (9)$$

$$dE(s)/dz = M(\delta)\ kE(s) \qquad (10)$$

for each strip at level z in the canopy displaced from the row center by distance, δ.

The relations (8), (9), and (10) are to be solved for each displacement, δ, assuming that the diffuse flux is still approximately laterally uniform across rows. Then the lateral average of radiance over [illegible] d[illegible]-placements, δ, must be calculated to find the average radiance [illegible] direction of view.

2.6.2.4 Row Model Predictions for Wheat

Two wheat development stages were modeled: Feekes 5 and Feekes 8. The row modulation was taken to be a "rectangular prism" modulation which might be suitable at Feekes 5 but limited inter-row growth was assumed for Feekes 8. Figures 2.21 and 2.22 each show polar plots of reflectance for three band-center wavelengths -- 550, 650, and 750 nm. Row direction is North-South in the plot and the direction of view is the nadir in all cases. Because of the symmetry due to the nadir view, only one sun azimuth quadrant for each band center is necessary to illustrate all of the important variations. Along with solar azimuthal variations shown on the polar plot, three different polar sun angles (zenith angles) were used. The solid polar plot is for a 25° sun polar angle, the long dash plot is for a 45° polar angle, and the short dash is for a 60° polar angle. The radial scale for 750 nm plot is different from the scale for the 550 and 650 nm plots.

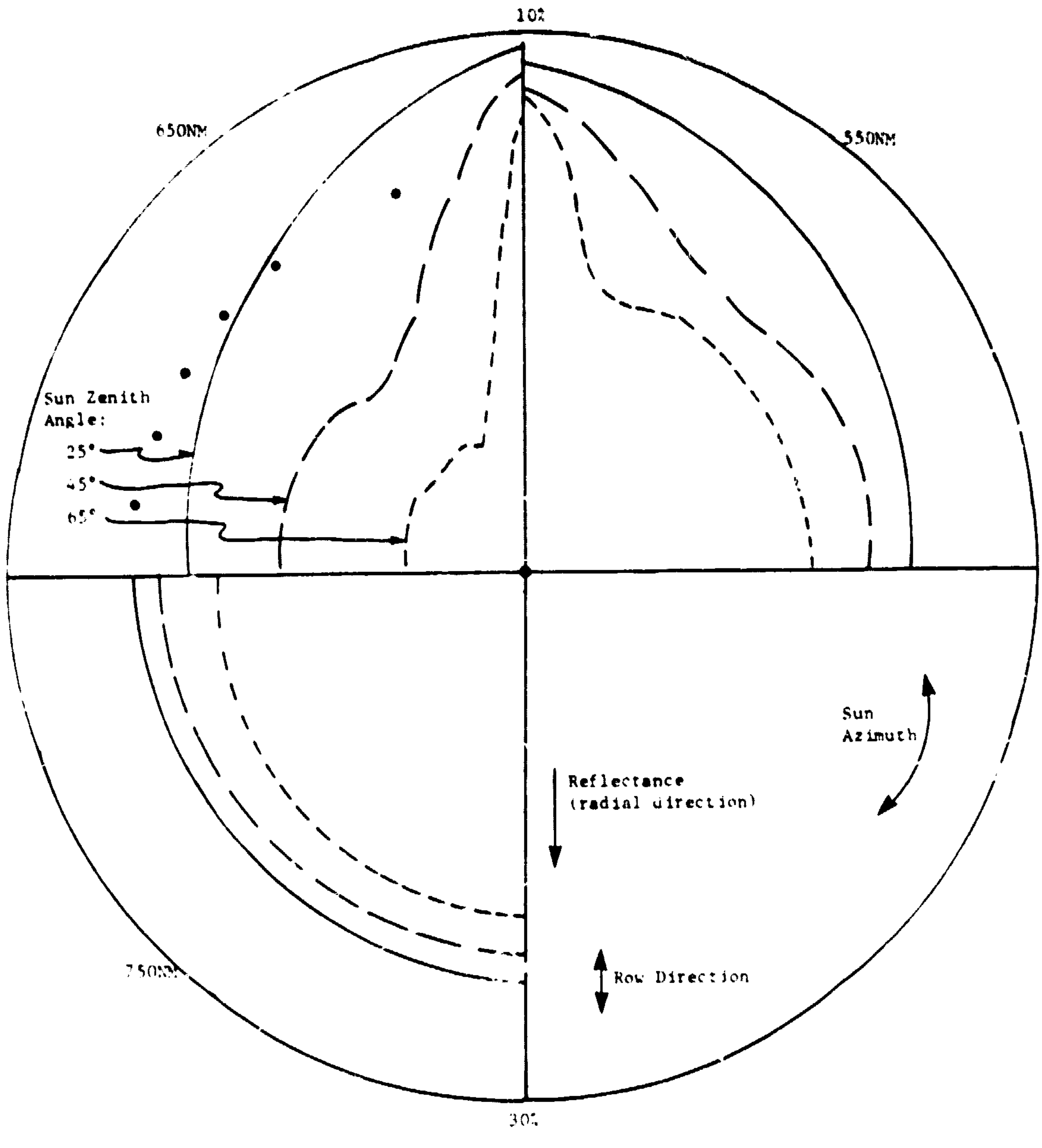


FIGURE 2.21. POLAR PLOT OF REFLECTANCE OF WHEAT AT FEEKES 5 WITH RECTANGULAR ROW STRUCTURE AS A FUNCTION OF SUN ANGLES (WITH NADIR VIEW)

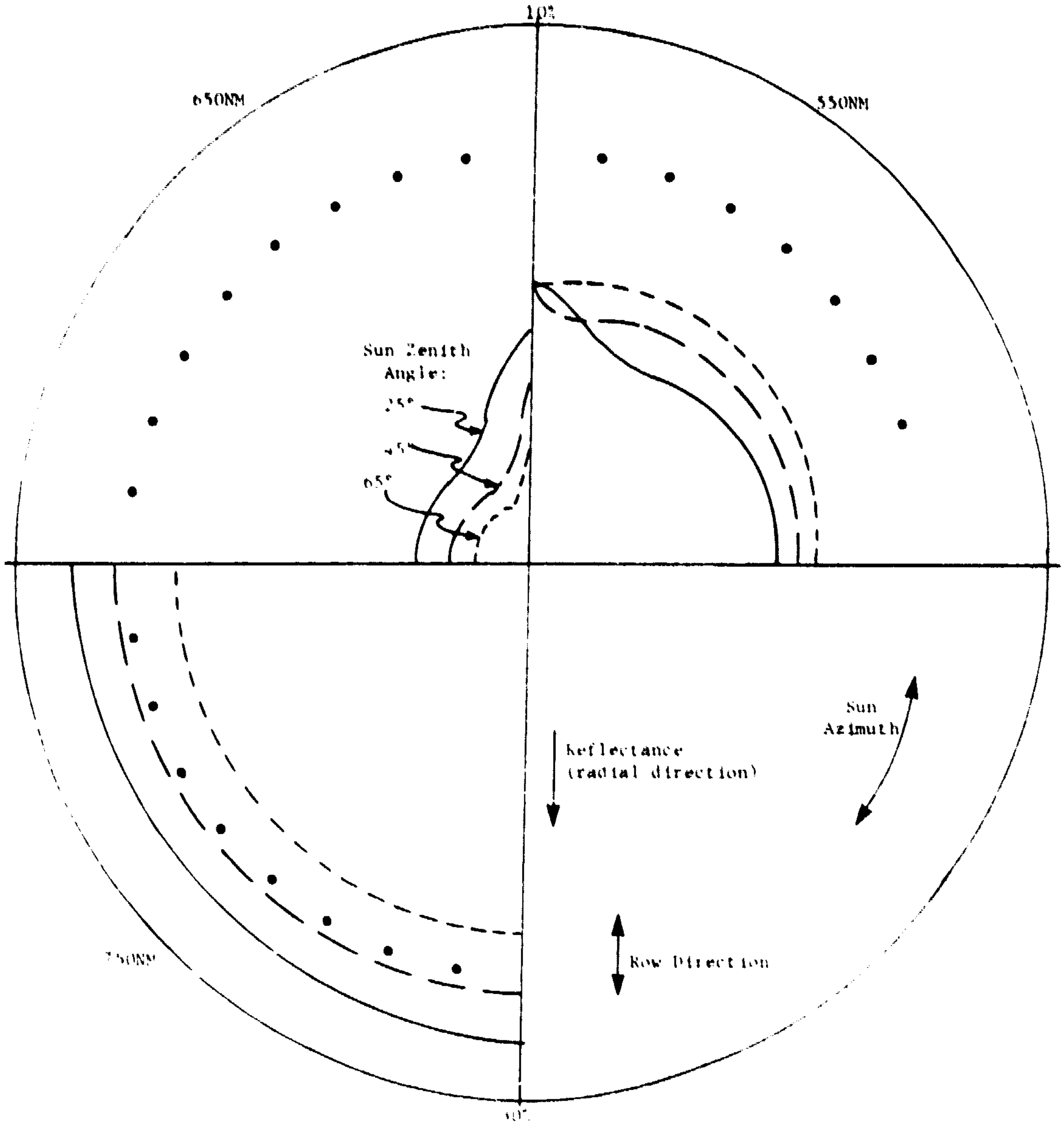


FIGURE 2.22. POLAR PLOT OF REFLECTANCE OF WHEAT AT FEEKES 8 WITH MODIFIED ROW STRUCTURE AS A FUNCTION OF SUN ANGLES (WITH NADIR VIEW)

Figure 2.21 shows the results for Feekes 5 wheat. The greatest effect is in the 650 nm band center and the effect becomes more significant as the polar sun angle increases, while the infrared 750 nm band center is only moderately affected. One can see that the infrared-to-red ratio, which is often used as a crop vigor measure, will be significantly altered merely by sun to row angle conditions. These calculations are for direct sunlight alone. The addition of skylight will tend to reduce the extreme variations for the setting sun.

The case for Feekes 8 wheat for sunlight alone is shown in Figure 2.22. The row structure was modified to allow 5% of the peak on-row concentration to appear at mid-row. Notice that the row effect is still significant but is much more subdued. It would not take much more vegetation in the inter-row region to reduce the row effect to negligible proportions.

The impact of row direction on Landsat signals from the latter field was estimated for a 45° sun angle and a nominal amount of path radiance. The resulting MSS7/MSS5 ratio and Greenness measures are shown in Table 2.9 for sun down-row and sun across-row directions.

TABLE 2.9. ESTIMATED EFFECT OF ROW DIRECTION ON LANDSAT RESULTS (Feekes 8, 45° sun zenith)

	Across-Row	Down-Row
MSS7/MSS5	2.0	1.33
Greenness	47.5	42.1

The down-row direction gives an indication of a much less vigorous field. An underestimation of crop vigor and biomass could result purely from a chance row-sun relation. However, the cross-row direction does not lead to a serious overestimation. The reflectance for the cross-row direction is not greatly different from that of the uniform canopy.

2.6.3 SPATIAL AND SPECTRAL SIMULATION OF LANDSAT AGRICULTURAL DATA

This section summarizes the development of a scene simulation capability which is described more fully in a separate technical report [41].

2.6.3.1 Introduction

The signal which the Landsat multispectral scanner generates is a function of many variables, few of which we have any control over. The ideal method of understanding a process is to hold all of the variables constant, except those under consideration. This method fails for the most part in the study of the Landsat signal-generation process with its seeming contradiction of vast amounts of data at the pixel level but a scarcity of data with unique combinations of factors such as scan angle, day of year, crop, field pattern, etc. Simulation is a tool which allows one to use combinations of assumed or known effects to infer the composite effect. The uses of a simulation include:

(1) The study of the interaction of known first order effects,

(2) Tests of procedures on data generated under known conditions, and

(3) Empirical estimation of model parameters when fitted to "real data."

The major motivation for the simulation model described here was the need for a capability to investigate, in detail, the effects of various factors on pixel values from small fields, boundaries between fields, and misregistered pixels. Both spectral and spatial properties were of interest. With this model any desired polygonal field pattern can be simulated and spectral characteristics can differ from field to field, with within-field variances being included.

2.6.3.2 The Model

Consider the point (x,y) on the ground at time t. Except for a set of area zero, (x,y) will be contained in the interior of a field. Denote this field as k. The main effect which a sensor could detect is that of the crop at point (x,y). We denote the crop in field k as C_k. We use crop development profiles in Greenness and Brightness to simulate the mean crop response as a function of time since planting. Reference 41 gives the empirically estimated profiles used, while Figure 2.23 illustrates those for corn, soybeans, small grains, pasture, etc.

Denote the profile for crop c as $P_c(.)$. Note that two fields with the same crop would not in general have the same profile value at time t due to different planting days. Denote the planting date for field k as T_k. The model further assumes that there are field effects beyond crop type and planting date due to soil characteristics, crop variety, fertilizer, etc. These additional between-field, within-crop sources of variability are viewed as geometric noise factors which scale each profile. Denote the scale factor for field k as U_k, where U_k is a random variable with a mean of 1. The profile at (x,y) is

$$g(x,y,t) := U_k P_{ck}(t-T_k) + \varepsilon_{txy}$$

where

ε_{txy} is assumed to be a bivariate normal with mean of zero.

The model assumes that the covariance of ε_{txy} is a function of crop and time. This is reasonable if the dominant effect in within-field variation is due to crop-field effects. If sensor noise were the real dominant effect, then variances of the Landsat Bands 4, 5, and 6 would be proportional to the signal and the variance would be constant in Band 7.

One of the major problems encountered in multitemporal Landsat data is spatial misregistration between dates. The coordinate system changes between passes of the satellite. The point (x,y) in the satellite's coordinate system does not correspond to the same ground point. The relationship between the ground coordinate system and that of the sensor's is non-linear. There are registration procedures which reduce

ORIGINAL PAGE IS
OF POOR QUALITY

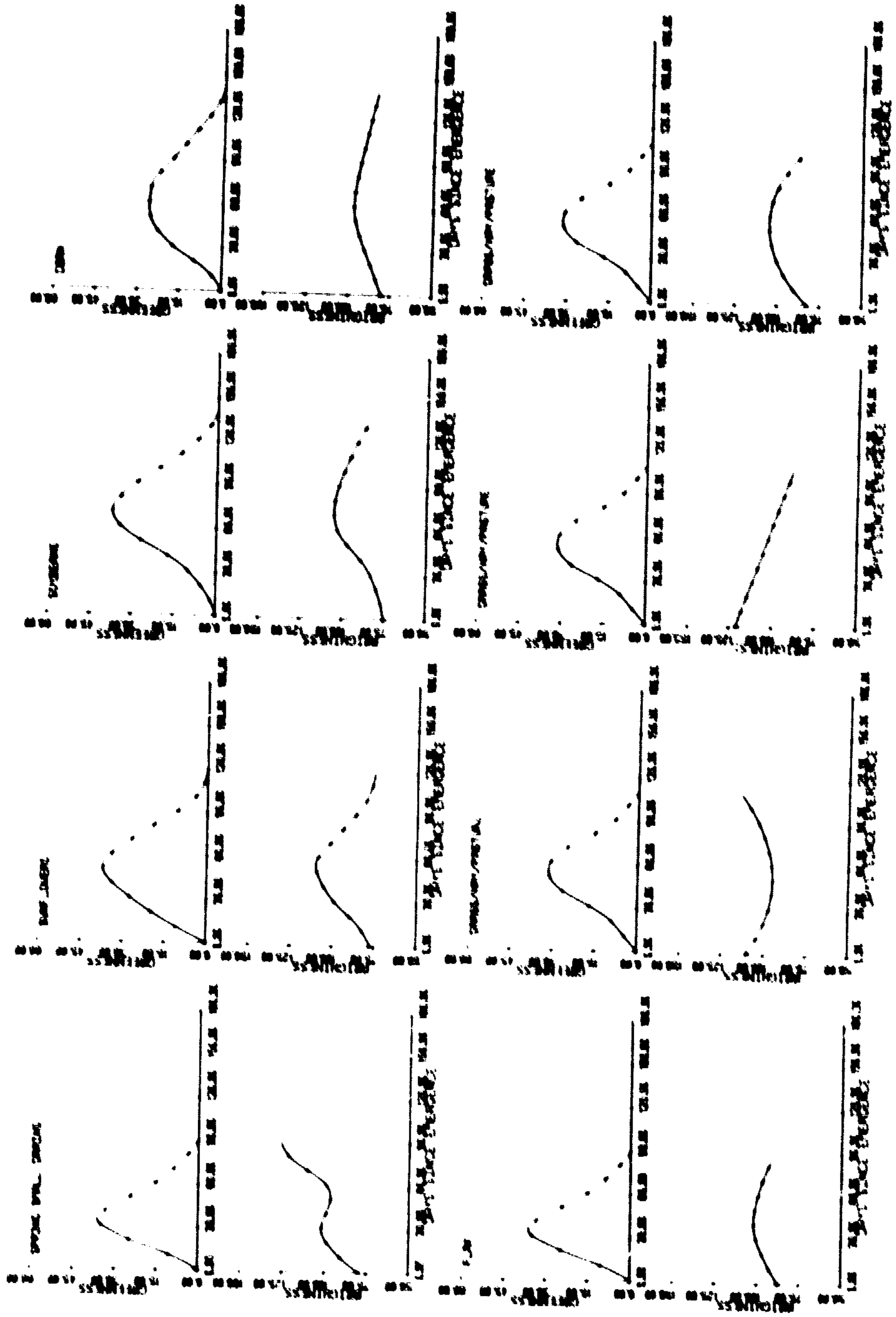


FIGURE 2.23. GREENNESS/BRIGHTNESS CROP PROFILES

the differences in coordinate systems; however, there is always a residual error in registration procedures. The model assumes the sensor coordinate system changes only by a translation between passes. If the ground coordinates are (x,y) then the sensor's coordinates at time t are $(x+x_t, y+y_t)$. This form of misregistration is suitable for most applications using simulation. A more general form of misregistration could be simulated by warping the coordinates which define the fields.

The signal which the sensor receives is not $g(x,y,t)$ but rather

$$f(x,y,t) = \iint g(x+x_t - r, y+y_t - s, t)p(r,s)drds$$

where

p is the Landsat point spread function.

p was derived in Reference 43 using the sensor's size, blur circle and properties of its three-pole Butterworth filter. Figure 2.24 gives a three-dimensional drawing of p and Figure 2.25 gives plots of p along the scan line and along track, at pixel center. The signals which the sensor allows us to observe are

$$\{f(x + idx, y + jdy, t)\} \qquad \begin{matrix} i=1,N_x \\ j=1,N_y \end{matrix}$$

Values for a 5x6-mile AgRISTARS segment are dx = 79M, dy = 57M, N_x = 196, and N_y = 117.

2.6.3.3 Implementation

The Field Geometry. Each field is stored in the computer as a polygon. The vertices of all of the fields are contained in arrays, say $\{U_{kj}, V_{kj}\}$. Polygon (field) k is defined by the vertices $k_1, k_2, \ldots, k_{nk}$ such that the points $\{U_{kj}, V_{kj}\}_{j=1}^{N_k}$ circumscribe field k

ORIGINAL PAGE IS
OF POOR QUALITY

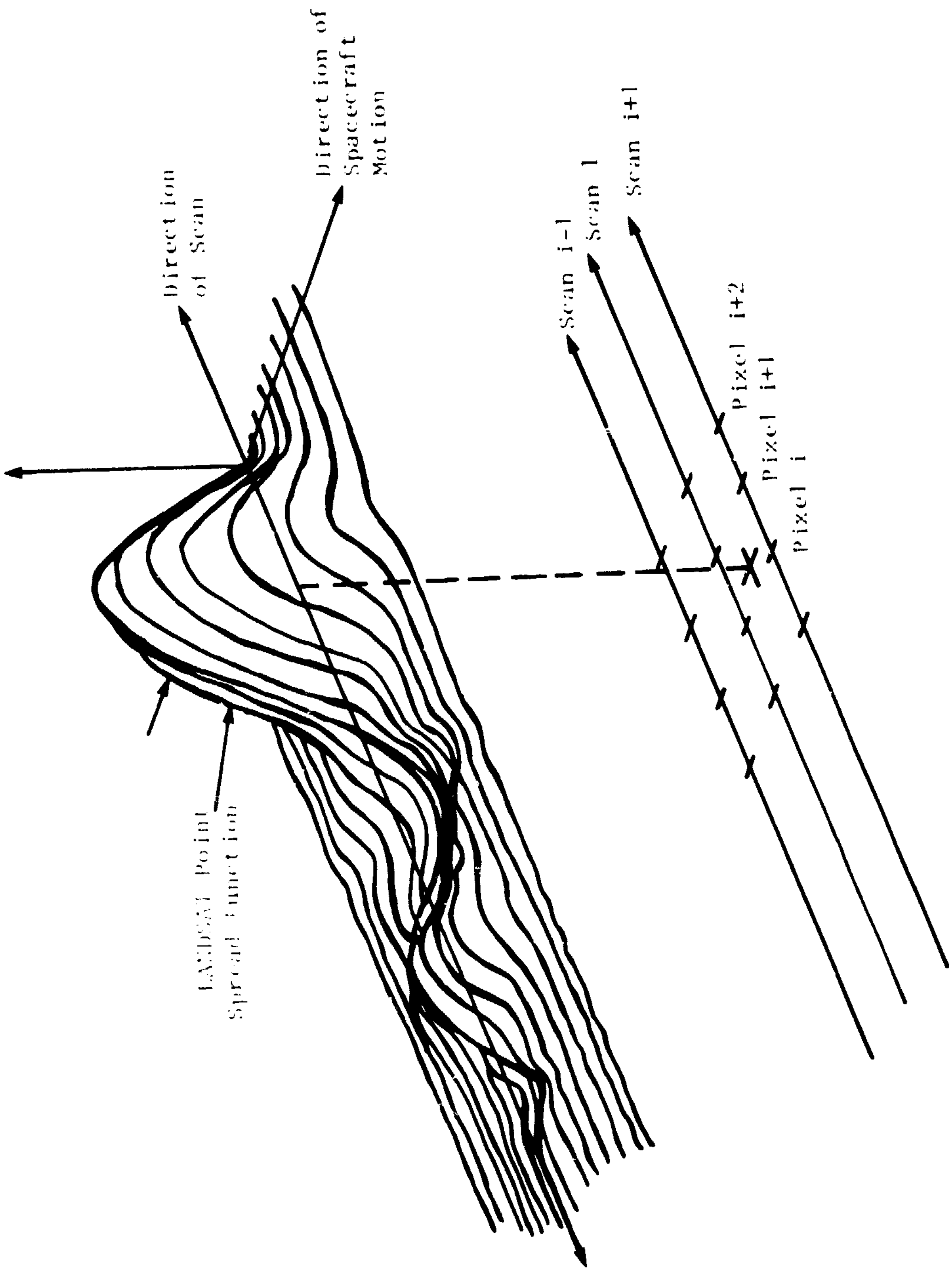
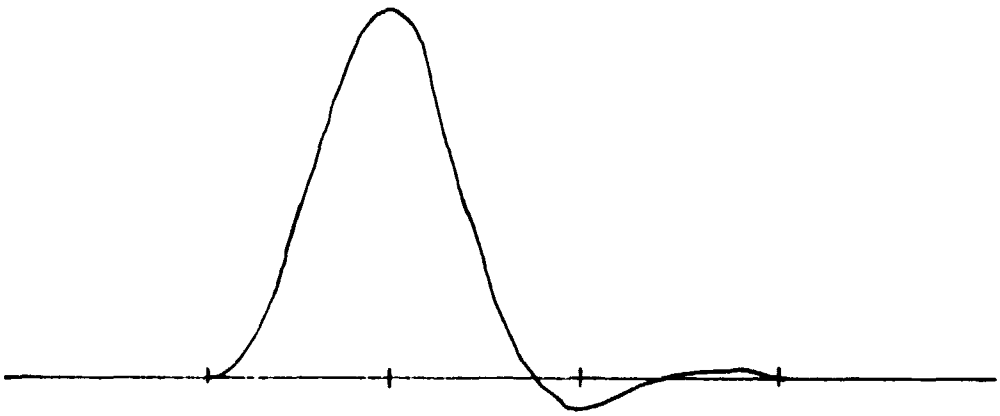
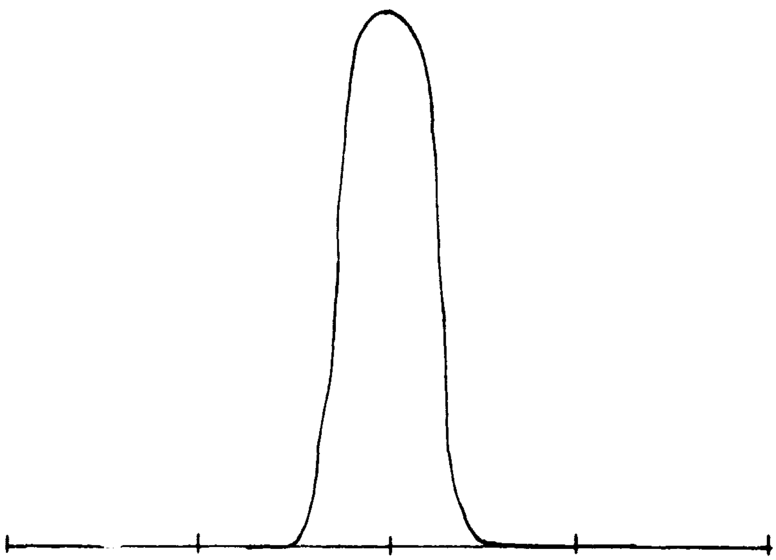


FIGURE 2.24. LANDSAT MSS POINT SPREAD FUNCTION



Landsat Along Scan Line Point Spread Function



Landsat Down Track Point Spread Function

FIGURE 2.25. PROJECTION OF MSS POINT SPREAD FUNCTION
ALONG AND DOWN TRACK

in a counterclockwise direction. It is important that there be no gaps in adjacent fields and non-nil intersections can cause unexpected results. We assume that all fields are simply connected, but more general sets could be incorporated into the model easily.

A two-dimensional grid of points is assigned polygon identification. The point (x,y) is assigned to the first polygon whose winding number is positive. The polygon search begins with the polygon which contained the previous pixel. If only translation misregistration is to be simulated then this pixel-to-field assignment only has to be performed once. If more general misregistration is to be simulated then the points $\{U_i, V_i\}$ can be replaced by $\{H_t(U_i, V_i)\}$ where H_t is the warping transform for time t. Examples of H_t are

$$H_t(U,V) = \left(\sum_{q=0}^{5} \sum_{j=0}^{q} a_{qj} U^j V^{q-j}, \quad \sum_{q=0}^{5} \sum_{j=0}^{q} b_{qj} U^j V^{q-j} \right) \tag{11}$$

and

$$H_t(Z) = A_t(Z - Z_t) + Z_t$$

where

$$z = u + v_i, \; Z_t = U_t + V_t i, \; \text{and} \; A_t = R_t e^{i\theta_t} \tag{12}$$

Functions of the form (11) are often used to correct geometric distortions in Landsat data. Regression methods are often used to estimate the coefficients a_{qj}'s and b_{qj}'s. Since there are 21 terms in each coordinate of (11) there should be somewhat more than 21 control points used in the estimation, if estimates of all coefficients are desired. Stepwise regression methods tend to get good results with 5-9 control points. Functions of the form (12) represent a rotation of θ_t and a scaling by R_t about (U_t, V_t).

Crop Response as a Function of Time and Field. The crop for point (x,y) on the ground at time t is

$$q(x,y,t) = U_k P_{ck}(t-Tk) + \varepsilon_{txy}$$

where

- k is the field containing (x,y).
- U_k is the scale factor for field k,
- C_k is the crop growing in field k,
- T_k is the time of planting,
- $P_c(.)$ is the Greenness/Brightness response of crop c as a function of time since planting, and
- ε_{txy} is the within-field variance.

The polygon specific parameters U_k, C_k and T_k are saved in a file until all acquisitions are generated. U_k and T_k are viewed as random variables such that $E\{U_k\} = 1$ and the distribution of T_k is obtained from a crop calendar specific to the region being simulated. Empirical profiles were incorporated for grain, sunflowers, corn, soybeans, and three types of grass/pasture/hay. New profiles can be added or old ones modified easily.

Presently the within-field error term is used only to add texture to the pixels contained in a given field. Data which would support an accurate estimation of the covariance matrix of ε_{txy} do not exist. The reason is that ground-truth polygons often contain more than one field with the same ground truth code, while field-finding algorithms are constrainted to construct field-like regions with small within-field variances.

The Convolution. The convolution of the sensor's point spread function blurs the image by adding correlations between nearby pixels. The sensor's response at point (x,y) and at time t is

$$f(x,y,t) = \iint g(x-r,y-s,t)p(r,s)drds.$$

We use two different levels of approximations of f(x,y,t):

$$f_1(x,y,t) = \sum_{i=-16}^{48} \sum_{j=-16}^{16} g(x - \tfrac{i}{16}, y - \tfrac{j}{16}, t)p_1(\tfrac{i}{16}, \tfrac{j}{16}) \qquad (13)$$

where

$$p_1(\tfrac{i}{16}, \tfrac{j}{16}) = \frac{p(\tfrac{i}{16}, \tfrac{j}{16})}{\sum_{r=-16}^{48} \sum_{s=-16}^{16} p(\tfrac{i}{16}, \tfrac{i}{16})} ;$$

and

$$f_2(x,y,t) = \sum_{i=-4}^{16} \sum_{j=-4}^{4} g(x - \tfrac{i}{4}, y - \tfrac{j}{4},t)p_2(\tfrac{i}{4}, \tfrac{j}{4}) \qquad (14)$$

where

$$p_2(\tfrac{i}{4}, \tfrac{j}{4}) = \frac{p(\tfrac{i}{4}, \tfrac{j}{4})}{\sum_{r=-4}^{16} \sum_{s=-4}^{4} p(\tfrac{r}{4}, \tfrac{s}{4})} .$$

2.6.3.4 An Example

To illustrate the capabilities of the model, the field pattern from the southwest quarter of Segment 844, during the year 1978, was digitized in polygonal form. Crops were assigned to the fields at random. The crop probability and planting date distributions in Table 2.10 were used. The field scale factor was generated randomly from the uniform (.95,1.05) distribution for each field. Figure 2.26 gives a plot of the field pattern used in this simulation. This region was represented by a 256x256 subpixel grid. Each pixel was defined to be a 4x4 subpixel region. The crop signatures were generated at the subpixel level; thus, within-pixel

TABLE 2.10. PARAMETERS USED IN GENERATING THE SIMULATION

Crop	P	Tk Distribution
Grain	.10	N(105.10)
Pasture V1	.05	N(105.10)
Pasture V2	.05	N(105.10)
Pasture V4	.10	N(105.10)
Sunflower	.10	N(138.10)
Corn	.25	N(148.10)
Soybeans	.25	N(156.10)
Flax	.10	N(105.10)

mixtures were in multiples of 1/16. The field identification of each point in the subpixel grid was obtained from the polygons. A 64x64 simulated image was produced for the following dates: 160, 169, 178, 187, 196, 205, 214, 223, 232, 241, 250, 259, 268, and 286 with no misregistration.

Figure 2.27 gives a Greenness/Brightness scatterplot for Date 178. The spring crops are for the most part greening down from their peak value of Greenness, while the summer crops are just starting to green up.

Figure 2.28 gives the scatterplot for Date 205. The spring crops by then have almost all dropped below a Greenness value of 10 and the summer crops are approaching their peak Greenness values. Corn and soybeans have not separated yet. There also are many mixed spring/summer crop pixels which take on the whole range of values between the high Greenness values of the summer crops and the low Greenness values of the spring crops.

Figure 2.29 gives the scatterplot for Date 223. Corn and soybeans are at their point of maximum separation. The random planting dates,

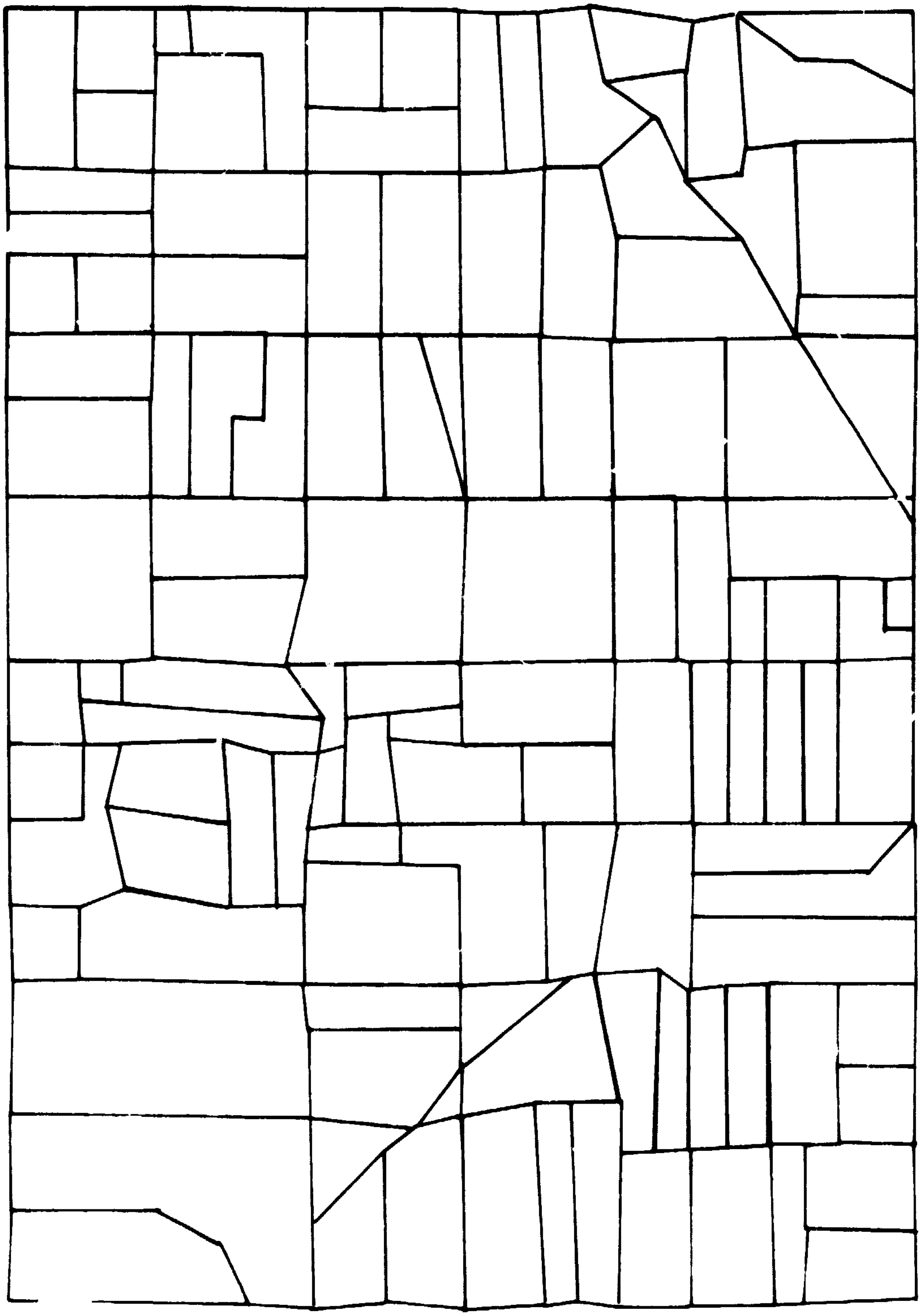


FIGURE 2.26. FIELD PATTERN FROM SEGMENT 844, 1978

ORIGINAL PAGE IS
OF POOR QUALITY

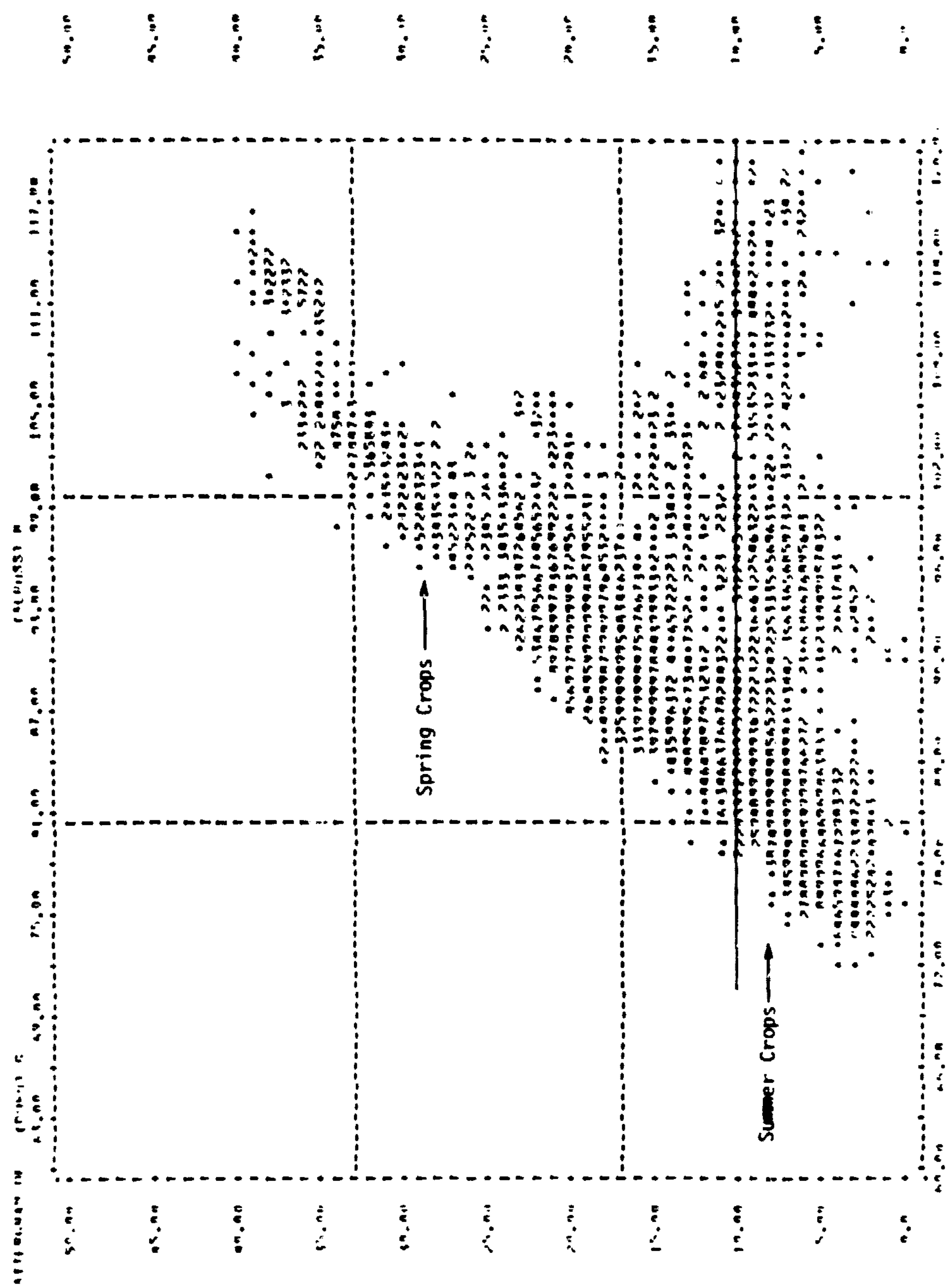


FIGURE 2.27. BRIGHTNESS VS. GREENNESS FOR DAY OF YEAR 178

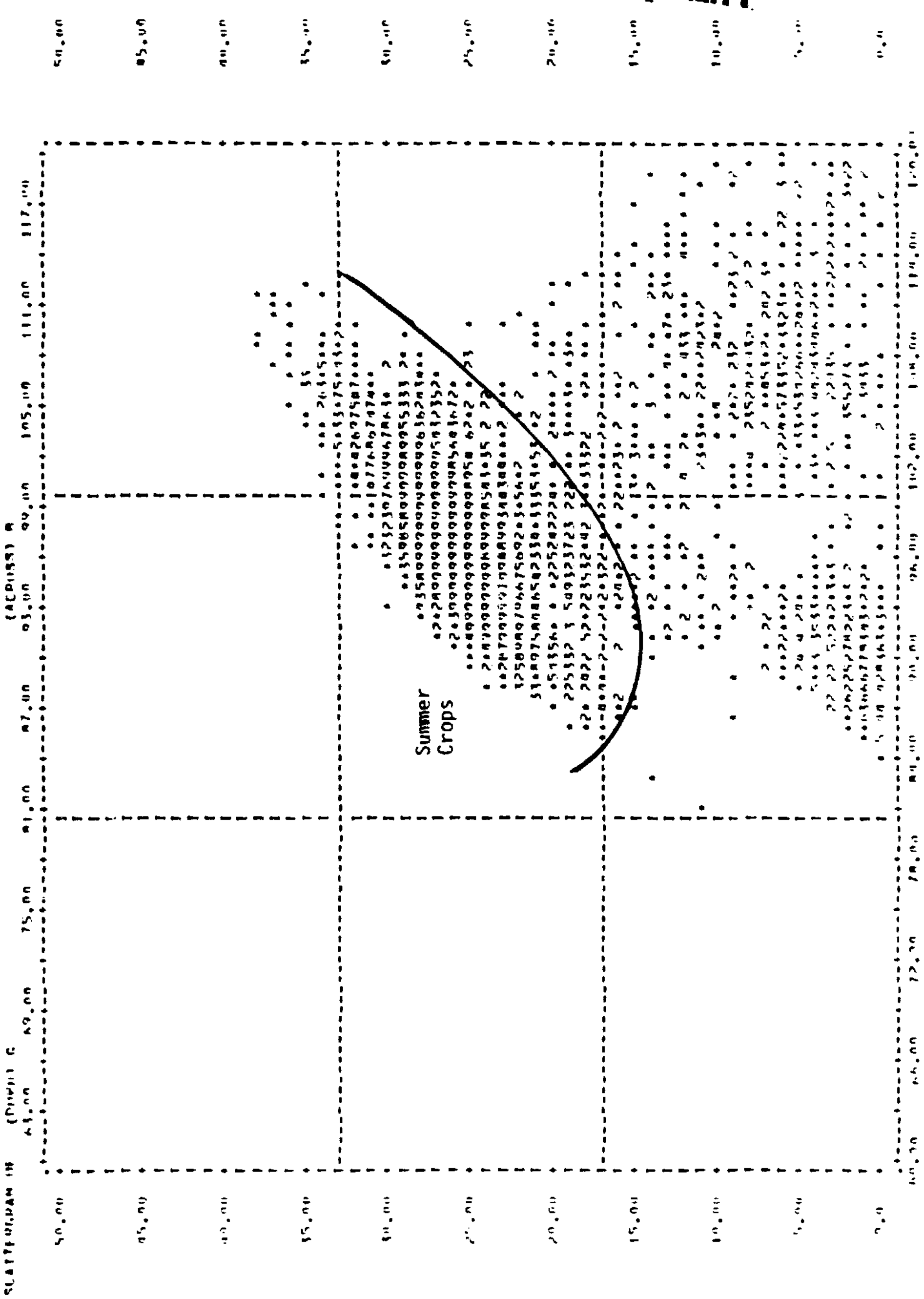


FIGURE 2.28. BRIGHTNESS VS. GREENNESS FOR DAY OF YEAR 205

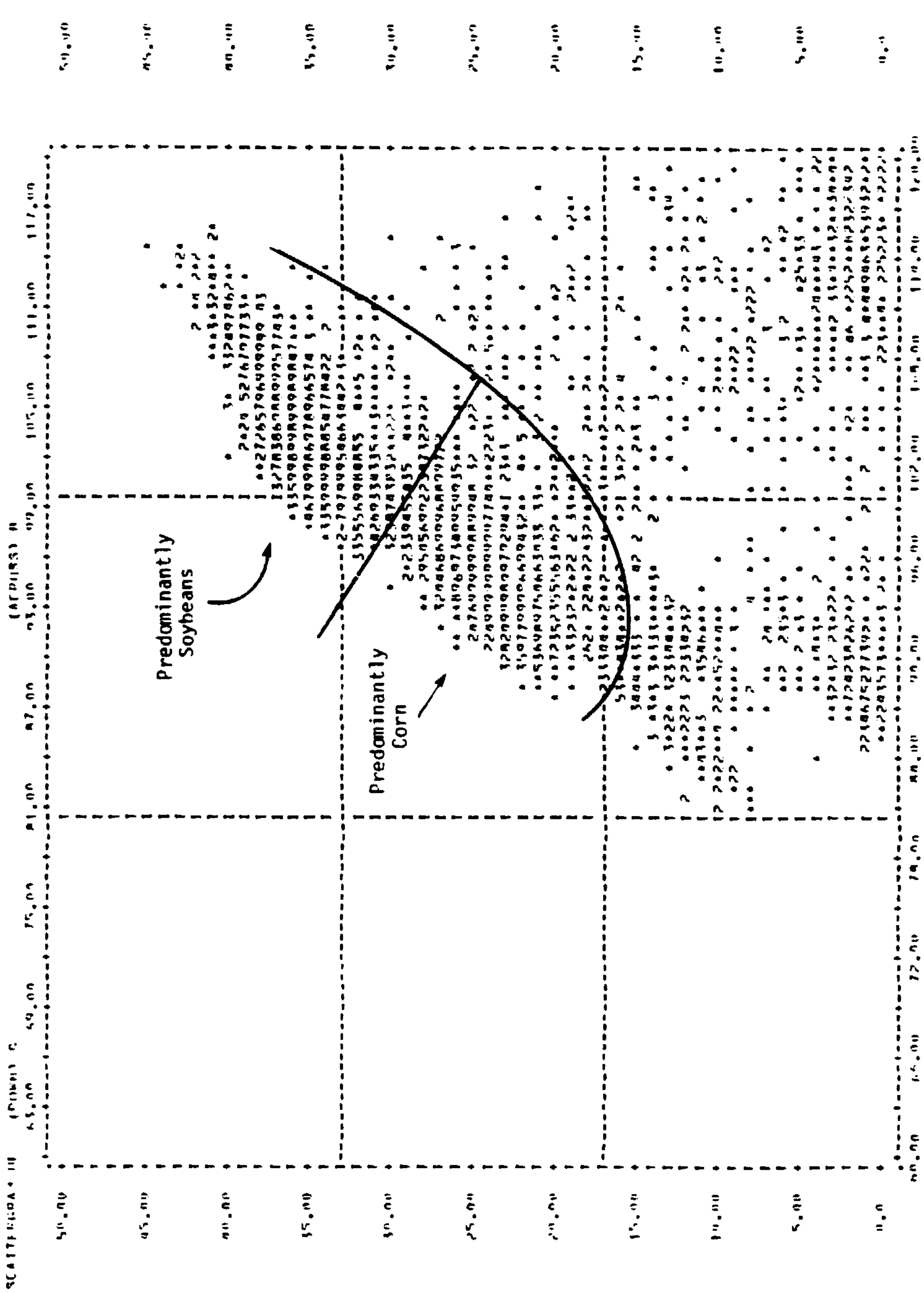


FIGURE 2.29. BRIGHTNESS VS. GREENNESS FOR DAY OF YEAR 223

scale factors, and mixed corn/soybeans pixels blur the spectral boundary between the two crops. A body of early summer crop pixels, mostly sunflowers, are greening down ahead of the main body of summer crops. Mixed spring/summer crop pixels still are evident.

2.6.3.5 Summary

The present understanding of several components in the Landsat signal-generation process allows the simulation of Landsat data.

The simulation described in this section allows for:

(1) Mixed pixels,
(2) Field geometry,
(3) Landsat point spread function,
(4) Crop development spectral profiles, and
(5) Variation in planting dates.

The simulation has been used in small field research. Other applications include the simulation of other sensors, the test of new procedures, and the study of new crop mixes and field patterns,

2.7 SMALL GRAINS LABELING TECHNIQUES

Research and development of automated labeling techniques for small grains were conducted primarily during 1980 and concluded during the first half of the current contract year. Two reports were written, one to describe the development procedure and initial test results [43] and the other to document the computer programs that were written and adapted to JSC computer facilities [44].

The work was a continuation of prior research in which a machine procedure was developed to discriminate between spring wheat and barley, given that the targets under consideration were spring small grains [13]. The objective here was to develop an automated technique for making the initial identifications of those spring small grains. Both labeling techniques exploit the temporal-spectral characteristics available from spatially registered multidate Landsat data. This technique was not intended to be the final and best use of profile technology, but rather a first generation technique, a demonstration of concepts, that can be used to more fully understand profiles and their uses, and thereby to develop improved labeling techniques.

2.7.1 DEVELOPMENT AND EVALUATION OF AN AUTOMATIC LABELING TECHNIQUE FOR SPRING SMALL GRAINS

Crop acreage estimates made using Landsat invariably require association of a crop label or labels with some sampling entity (e.g., pixel, field, cluster, etc.). The accuracy with which this association is made clearly has a substantial impact on the accuracy of the acreage estimates produced. In the Large Area Crop Inventory Experiment (LACIE), the labeling step, which was carried out through manual analysis of imagery and associated information, was found to be both time-consuming and a source of considerable error. An obvious candidate for improving both the objectivity and the timeliness of labeling decisions is automation of much of the labeling process.

The technique described in Reference 43 and summarized here was a response to the need for a faster, more accurate, and more objective labeling procedure. Human analysts are utilized only to set up the system and provide contextual information which can be used to adjust the labeling procedure to local conditions; the labeling decisions themselves are left to the machine. A problem addressed is that Landsat observations are fairly widely spaced and discrete samples in time of the generally continuous spectral development patterns of crops. To counter this, we developed "profile" techniques to characterize the sampled patterns and adjust for planting date differences and, to a degree, normalize stress effects among fields of a given crop [13,45].

The central element in the procedure is a group of profile sets representing spectral development of a number of crops in the domain described by Tasseled-Cap Greenness and Brightness. These profile sets were developed using spectral data from fields of known crop type, sampled from the U.S. Northern Great Plains over three growing seasons. They serve as reference standards to which each unknown sampling entity is compared.

For each profile set, a series of comparisons is carried out. First, a temporal shift is determined which maximizes the cross-correlation of the data points to the Greenness profile. This provides an estimate of the date of spectral emergence, and indirectly of the start of the growing season of the target field. The temporal shift estimate also provides a means of normalizing the planting dates of fields of a single crop type, andthereby minimizes one major source of spectral confusion.

After estimating and applying the temporal shift, a multiplicative scale factor is computed, again using the Greenness profile. This scale factor is applied to normalize the magnitude of the Greenness development profile which is strongly influenced, within a single crop type, by the percentage of ground covered by green vegetation (which is itself influenced by such factors as planting density, fertilization and moisture availability).

With both adjustments made,a goodness-of-fit of the data to the Greenness profile is computed, and similarly, using the Greenness profile temporal shift, a fit or correlation of the Brightness data to the Brightness profile is computed.

The shift, Greenness fit, and Brightness correlation are used to compute a probability associated with the crop represented by the profile set and the sampling entity, and this combined probability serves as the basis for labeling decisions. In a different application of this procedure, one might use different or additional features to compute the requisite probabilities.

2.7.2. TEST RESULTS AND EVALUATION SUMMARY

The small grains labeling technique was applied to 38 5x6-nautical-mile sample segments spanning three growing seasons. The labeling technique was run on ground-truth identified small grains targets in a number of different configurations, with various combinations of profile sets, test statistic weightings, and probability thresholds.

Although acquisition requirements for the procedure were not severe (three vegetated acquisitions), only 57% of the targets (spectral-spatial clusters called "blobs") in both the development and testing data sets (64 total segments) met the acquisition requirements for labeling. However, most sample segments were either "labelable" or not; 31 of the 64 had more than 80% of their blobs labeled, while 16 of the segments had less than 75% labeled.

Grain labeling accuracies reached 86%, but large errors of commission occurred at this level. Overall accuracies reached 74%. Major causes of errors were Grasses and Flax (and the Grass and Flax profiles), and overall results were substantially improved when the profiles of these two crops were omitted from the profile set.

Several improvements were supported by the test results. A mechanism by which pasture blobs could be detected prior to application of the grain labeler would remove the largest source of erroneous labels. Improvement in the Brightness profiles, and in our understanding of Tasseled-Cap Brightness as it relates to crop characteristics, would also be beneficial.

The test results also point up some larger issues with regard to crop identification using Landsat. The low percentage of labelable blobs, using acquisition requirements similar to those observed by others [46,47], strongly suggests the need for more frequent coverage, and places a premium on development of techniques which can extract the maximum information possible from a limited set of observations.

Second, the relatively frequent occurrence of abnormal spectral patterns for blobs of a known crop type raises questions related to profile matching techniques. While a range of variability is expected and accomodated in the techniques described in this section, extreme deviations cannot be accomodated. We suggest that such patterns are, in the vast majority of cases, the result of catastrophic events which reduce, or eliminate any yield from the field in question (or ground truth error). Thus, while profile matching techniques may be less appropriate for detection of all fields of a given crop, they should serve well in detecting yield producing fields.

2.7.3 SOFTWARE DOCUMENTATION

Reference [44] describes and documents the computer software necessary to perform two research labeling procedures, the small grains labeling procedure described in the preceding sections and they previously developed procedure for discriminating between spring wheat and barley. The subroutines were designed to operate on three computer systems in an environments developed for use on the AgRISTARS program and built around the ERIM/UCB corn and soybeans baseline classification procedure. These facilities are located at ERIM (actually the University of Michigan), Purdue/LARS and NASA/JSC (EODLS).

2.7.4 GROUND TRUTH SUMMARIES FOR U.S. AREAS

A summary of crop proportions in digitized ground truth data was prepared under this contract for all 5x6-mile segments inventoried (and digitized) for AgRISTARS in agricultural areas of the United States during the years 1976-1979 [48]. The complete set of ground truth data was collected by ground truth enumerators from the U.S. Department of Agriculture. The enumerators recorded crop type and condition and field boundaries on base maps. The resulting ground truth records were digitized by LEMSCO and by ERIM.

These complete ground truth records were used by ERIM to prepare summary data. Fifty-four year-independent crop categories were established and further consolidated into a concise summary of major crop types and groups present in each segment. The occurrences of special categories and situations arealso noted, such as percent of scene in special fields, percent strip farmed and percent abandoned. The proportions were based on a systematic, 20% sample produced by processing one line in five of the original (2x3) sub-pixel data. These summaries should be useful in screening and selecting segments for analysis and conducting evaluations of developed procedures.

2.8 SUPPORTING RESEARCH CONCLUSIONS AND RECOMMENDATIONS

Substantial progress was made along two major lines of research for supporting crop inventory systems utilizing Landsat data. These addressed, respectively, sampling and estimation technology and measurement technology, the latter dealing with the extraction of agrophysically meaningful features from Landsat data for use by the former in crop inventory estimation and assessment.

A prime emphasis of the sampling and estimation research was on techniques capable of providing estimates throughout the growing season, particularly early in the season. Crop estimation was characterized as being a composite process beginning primarily with prediction and becoming more dependent on actual measurement as the season progresses. An approach, that was developed and thought to be original, was to merge early, but current-season Landsat-derived information with prior season inputs of a conventional crop acreage prediction model. The resulting Landsat-augmented crop acreage response model (CARM) showed potential for early season estimates with improved accuracy. Also, the model was applied to a regional area rather than the usual national-level use of the conventional CARM models. We also explored ways in which knowledge of cropping practices at the regional and local levels could be used on a field-by-field basis to improve the quality and accuracy of information extractable from Landsat; included were techniques that could use multiyear information such as on year-to-year crop rotations. Finally, a segment-level Bayesian estimation approach was formulated to incorporate the key elements identified for through-the-season estimation.

Multisegment research examined approaches for increasing sampling efficiency and reducing measurement cost without sacrificing accuracy. Signature extension, regression, and bin methods were studied and an experiment using the bin method was carried out before the scope of these activities was reduced.

The remaining activity under estimation technology research was the organization and conduct of a field trip to Argentina to acquire crop identification data for over 600 fields in 14 segments located in three major agricultural provinces. This trip was arranged on short notice to provide an initial Landsat data set with ground truth information. Possibilities for continued and expanded ground data collection activities in South America were explored and draft plans were generated.

Under measurement technology research, studies were made of crop temporal-spectral profile characteristics, three simulation models were developed, and a previously started small-grain labeling procedure was completed. Field measurement reflectance data for corn and soybeans were analyzed, with emphasis on relating temporal-spectral (Greenness) profile features and characteristics to crop development stages and the effects of farm management variables such as planting date and fertilization.

The first simulation modeling activity interfaced a meteorologically driven wheat growth model with a vegetation canopy reflectance model to provide a capability to simulate the observable crop characteristics as a function of time and environment. The second modeling activity extended a uniform canopy reflectance model to include row effects. The final model was able to simulate both the spatial and spectral characteristics of agricultural scenes in order that mixed and boundary pixel effects can be analyzed. Effects of the Landsat spatial point-spread function and varied planting dates also were included.

Several recommendations are made on the basis of the conducted research and experience of the investigators.

(1) Recommendation re Sampling and Estimation Technology

(a) That research be continued on the Landsat augmentation of conventional crop acreage response models.

(b) That development be continued of techniques, that blend prediction and measurement capabilities and incorporate agronomic information at the field level, taking advantage of multiyear data where available; for long range development, we specifically recommend investigation of knowledge engineering systems tailored to this application.

(c) That research into multisegment approaches be conducted to improve inventory system efficiency and that it be closely linked to through-the-season requirements and techniques.

(d) That plans for ground data collection in Argentina and/or Brazil be further developed and carried out to provide basic information essential to the full development of Landsat-based inventory techniques for that region.

(2) Recommendations re Measurement Technology

(a) That Brightness profile variables from crops be investigated in addition to Greenness variables and that the study be extended from reflectance data to Landsat data.

(b) That the Seed-to-Satellite model be upgraded to incorporate the revised Ritchie wheat growth model and that extension to other crops, such as corn and soybeans, be pursued.

(c) That the row effects extension of the canopy reflectance model be verified by comparison with empirical data.

(d) That the existing models be used to further investigate small-fields effects in Landsat data from agricultural scenes and its impact on estimation accuracy.

3
INVENTORY TECHNOLOGY DEVELOPMENT PROGRESS AND RESULTS

Activities in support of the AgRISTARS Inventory Technology Development Project (ITD), formerly Foreign Commodity Production Forecasting, have revolved about developing Landsat-based crop inventory system component technology that is appropriate for eventual application in a foreign context, specifically for corn and soybeans in Argentina and Brazil. Activities reported in this section represented a joint effort involving ERIM and The Space Sciences Laboratory of the University of California at Berkeley (UCB), with test and evaluation support from Lockheed Engineering and Management Services Company, Inc. (LEMSCO).

3.1 APPROACH AND TASK STRUCTURE

The approach pursued in support of ITD in AgRISTARS has involved overlapping phases as is illustrated in Figure 3.1. In the initial phase, effort has been placed in the application and evaluation of technology based on Landsat MSS using, as in LACIE, segment sampling for wide area estimates of crop acreage and production in the U.S. where developmental data is readily available. The next stage would focus on the development of alternative techniques to establish a base of technology that could be comparatively evaluated and adapted to the foreign application and be supportive of an end-to-end inventory technology for Argentina and Brazil. This would then be evaluated in a controlled experimental environment to determine the technologies' feasibility for the foreign context.

Section 3 describes efforts conducted in the first two phases of the program to develop crop inventory technology for Argentina and Brazil. Efforts have been structured into two tasks in addressing the

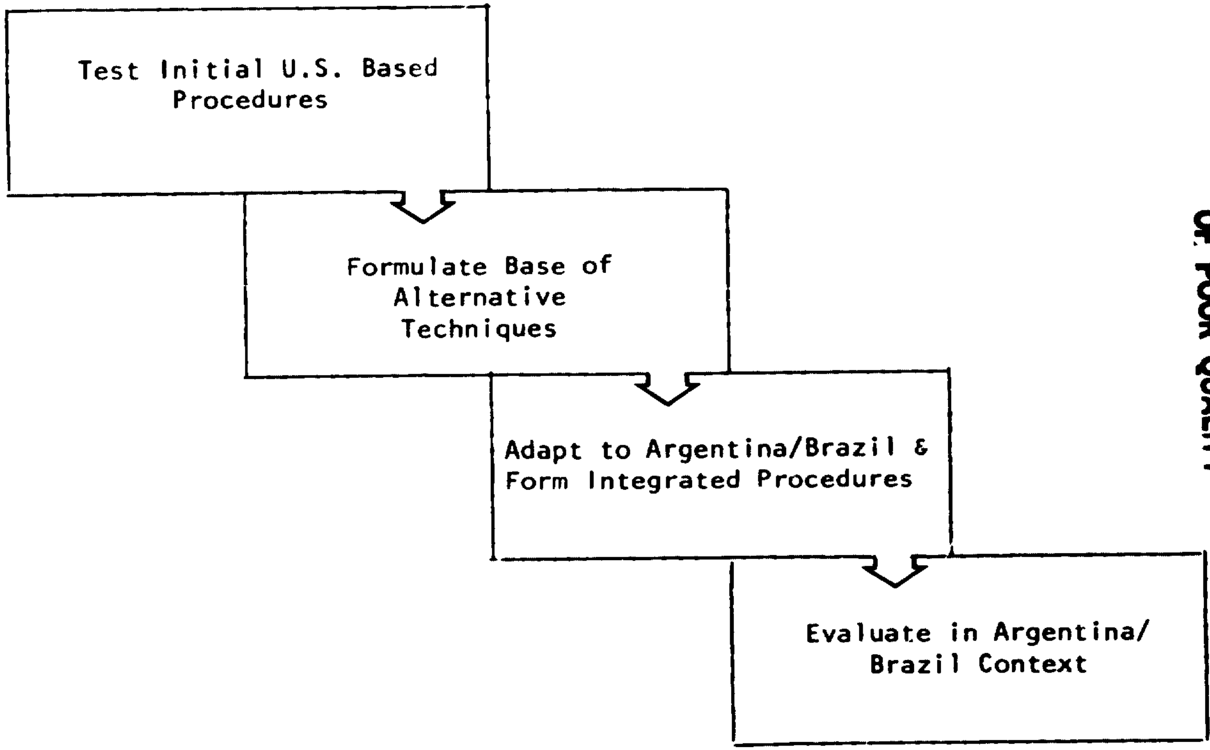


FIGURE 3.1. TECHNICAL PHASES FOR ARGENTINA/BRAZIL CROP INVENTORY TECHNOLOGY

overall objective of using remote sensing as a tool to inventory and assess corn and soybeans in Argentina and Brazil.

The first task, entitled "Experiments", tested and evaluated systems of technology components for crop inventory (i.e., procedures) under controlled and documented conditions. This task focused on evaluating the technology, formed into procedures, with respect to accuracy, objectivity, efficiency, timeliness and applicability to foreign conditions.

Reported in Section 3.2 is the development and evaluation of a procedure designated the Baseline Corn and Soybean Area Estimation Procedure. The technique was (and modifications continue to be) rigorously evaluated under configuration controlled conditions. The experiment discussed in Section 3.2 is referred to as the U.S. Corn and Soybean Pilot Experiment and was a joint LEMSCO/Consortium activity. In addition, a description is provided of the software system called STARS designed for purposes of configuration controlled procedure testing.

The second task is entitled "Technology Development, Evaluation and Integration". The major objectives of this task are to obtain, adapt (modify), or develop technology components (as opposed to end-to-end procedures) for assessing crop status, to evaluate the components for applicability to the problem, and to select and integrate appropriate components into end-to-end procedures for more formal evaluation.

Five areas of study are presented in Section 3.3 pursuant of the objectives of the second task. First, an examination of potential discriminating features of corn and soybeans with respect to key confusion crops present in Argentina was undertaken. Since procedures are developed under the constraint that ground training data would not be used, it is critical to determine the level of discriminating information directly derivable from Landsat based on prior understanding of crop attributes. This activity was largely the responsibility of UCB. Secondly, a technique based on the use of parametric models of MSS spectral features was examined with respect to its feasibility in establishing crop features derived from the multitemporal Landsat data that relate to crop agronomic attributes, for example, the length of a growth cycle. Thirdly, a study was carried out to assess methods that establish the basic sampling unit within a segment. Automatic techniques for definition of field-like targets were of central interest. Fourthly, an analysis of a double sampling technique to aggregate segment estimates to a regional level was undertaken. In this analysis the feasibility of joining two types of estimates, an inexpensive and less accurate technique, with a more expensive and accurate technique, was found to reduce the variance of estimates for given cost constraints. Finally, an effort carried out to prepare an initial foreign ground data set collected in Argentina (see Section 2.4) is described.

3.2 U.S. BASED CORN AND SOYBEAN AREA ESTIMATION PROCEDURE DEVELOPMENT AND TESTING

The Corn and Soybean Consortium, with ERIM assigned the lead technical role, was given the responsibility of developing a baseline corn and soybean area estimation procedure which uses Landsat data without ground observed training data. This procedure was designated the Baseline Procedure because it was intended to serve as the standard against which all future modifications of the procedure, and new procedures, would be judged and thereby provide a benchmark against which progress can be assessed. Twofold design specifications of this procedure required first that it consist of a modular framework within which individual component technologies could be developed, compared, substituted and evaluated, and secondly that the procedure could be carried out by analysts that were not necessarily expert. The procedure which resulted, called C/S-1, was developed by ERIM and UCB and delivered to JSC for evaluation in a major test conducted by LEMSCO.

This test, known as the U.S. Corn and Soybean Pilot Experiment, was structured in two phases. The first phase, conducted from January to April 1981, consisted of 39 segment processings of Landsat MSS data from the U.S. Central Corn Belt; 30 of these were 1978 data and 9 were 1979 data. The test involved 3 teams of 2 analysts each. A balanced, incomplete design was used, resulting in each segment being processed twice, but not necessarily by the same two analyst teams. It was intended that these processings be evaluated in time to allow modification of the procedure, if necessary, prior to proceeding with the second phase of the experiment. The second phase was scheduled to be completed in FY1982, and is to include testing of the aggregation procedures, which would produce regional estimates as well as the segment estimation procedure. To allow aggregation, approximately 50 segments of 1980 data are to be processed in each of Iowa, Indiana, and Illinois.

Results of the Phase 1 test of the experiment indicated the presence of bias in C/S-1 in excess of 10% relative to the true. This led to a decision to study the procedure in greater depth to provide guidance for efforts aimed at (1) reducing the observed bias, and (2) improving the efficiency of the procedure. This study consisted of component and subcomponent performance evaluations of C/S-1 performed by LEMSCO and ERIM, respectively, to identify those parts of the procedure which held the most promise for modification, resulting in improved accuracy and efficiency. Implementation by ERIM and UCB of the modifications recommended by this study resulted in the augmented baseline procedure, C/S-1A. Initial tests performed by ERIM indicate that C/S-1A represents an improvement over C/S-1 in both accuracy and efficiency. Further testing of the procedure is to be performed in FY1982 in the second phase of the pilot.

Development and implementation of machine procedures for C/S-1 and C/S-1A was performed by ERIM using the Software Technology for Aerospace Remote Sensing system (STARS). This system was developed by ERIM to provide a controlled environment for procedure implementation as well as providing the user and data interfaces necessary for smooth operation of the procedure in a production mode. This latter capability was demonstrated in the U.S. Pilot experiment, in which both C/S-1 and C/S-1A operated within STARS.

The following sections provide a more detailed description of the history, technical specifications, and evaluation of the baseline corn and soybean area estimation procedure and STARS.

3.2.1 BACKGROUND

The Baseline Procedure represents the integration of three earlier component technologies: (1) Procedure M [13], (2) the corn/soybean classification logic [49], and (3) the Delta Function Stratification (DFS) [50]. Procedure M (for multicrop) was developed at ERIM in parallel with development of Procedure 1 in LACIE [12]. Procedure 1 was developed by the Earth Observations Division of NASA/JSC in 1976-77 and supported producing LACIE generic wheat estimates. It was the forerunner of the Baseline Procedure from the standpoint of being the first "proceduralized" approach to large area crop inventory in foreign areas using Landsat. Proceduralized means employing a well-defined methodology which can be objectively applied over large areas. Procedure 1 also broke new ground by relying on a statistical design to generate crop proportion estimates, as opposed to more typical pixel classification techniques.

The switchover from classification technology to strategies employing stratified areal estimation statistical designs was justified on the grounds that the latter techniques are theoretically unbiased, while classification technologies are not. Furthermore, the component technologies necessary to support a statistical approach were now in existence and tested sufficiently to provide the confidence that such a procedure could be practically implemented.

Since ground observed training data was not used, the sample labeling logic used in Procedure 1 relied on analyst interpreters making decisions about the identity of areas located under dots (pixels). These sample dots were located systematically throughout a segment of Landsat MSS data (5x6 miles). The system is described as an "expert" labeling system, because the analysts did not have to follow a well-defined decision logic to reach an identity for the sample but, rather, only had to stay within general guidelines and exercise their own judgment.

Roughly in parallel with the development of Procedure 1, a similar procedure called Procedure M was developed at ERIM in 1977-78. Procedure M was designed to reduce labeling errors by using a different method of selecting the samples that the analyst was required to label. Studies had shown that a major source of labeling error in Procedure 1 was the problem of not being able to correctly identify boundary pixels

(pixels located on the edge of fields). Procedure M reduced this problem by using an algorithm called BLOB to find field-like samples and then restricting labeling to blob interiors, on the presumption that they were spectrally and spatially (in terms of ground truth) pure. In a related change, the systematic selection of the samples to be labeled was also dropped, in favor of a stratified random selection of blobs, where the stratification is based on the spectral similarity of the blobs. The method of sample selection tested to show a substantial reduction in the variance of the estimate over that of Procedure 1.

The development of Procedure M resulted in a general proceduralized approach that could be used to produce estimates for a variety of crops. To make it applicable for producing corn and soybean estimates, a decision logic capable of identifying samples of these crops was also required. An initial logic was available as a result of work done by Lockheed in 1979. Their original goal was to test whether or not a well-defined decision logic could produce consistent classification results as accurate as those generated by an "expert" system. The results of this work showed promise in achieving objectivity. In 1980 the initial corn/soybean logic was substantially revised and augmented by UCB for incorporation in the Baseline Procedure.

The other key component required to complete the Baseline Procedure is the Delta Function Stratification (DFS) technology. DFS is a way of introducing crop calendar data into the procedure in a useful and consistent fashion. The development of DFS began at UCB in 1978, continued in 1979, and was integrated into the procedure in 1980. A side benefit of DFS is that it also provides a method of obtaining first cut estimates of the proportion crops, other than corn and soybeans, and other land use categories in the segment early in the procedure, and without the need to actually classify the data into crop types.

In 1980, when all of these component technologies were successfully integrated, the Baseline Procedure, or C/S-1, was born [51].

It is a procedure unique in the fashion in which a convergence of evidence produced by different subcomponents feed each other and result in a statistically trackable estimate of the crop proportions in a segment.

3.2.2 BASELINE TECHNOLOGY

3.2.2.1 Introduction

The U.S. Baseline corn and soybean segment classification procedure is a methodology for estimating the corn and soybean acreage in Landsat segments selected from the U.S. Corn Belt (Illinois, Indiana, and Iowa).

It is designed to produce near-harvest crop proportion estimates within segments for corn and soybeans using multitemporal Landsat data. The estimates are produced by an integrated Analyst/machine procedure. The procedure is initiated with the Analyst screening the Landsat data for quality, selecting acquisitions for analysis, and participating in stratification of the scene. The machine then digitially preprocesses the Landsat data to remove external effects, completes the stratification of the scene, and samples the data proportional to the size of the strata. The Analyst then labels these samples as to crop type using an objective decision logic.

Assignment of crop type labels follows a "convergence-of-evidence" approach. That is, a progressive accumulation of information contributes to the selection of a particular crop label. Multi-date Landsat data are required since phenological crop development patterns which manifest themselves as changes in Landsat reflectance over time are the key to crop separability. The samples, consisting of field-like labeling targets called blobs, are objectively labeled by an Analyst according to crop type, specifically "corn", "soybean" or "other". Analysts label blobs according to an objective, well-defined decision logic with the aid of spectral plots and statistics provided by the machine, keeping in mind the influences of local meteorological conditions and cropping practices.

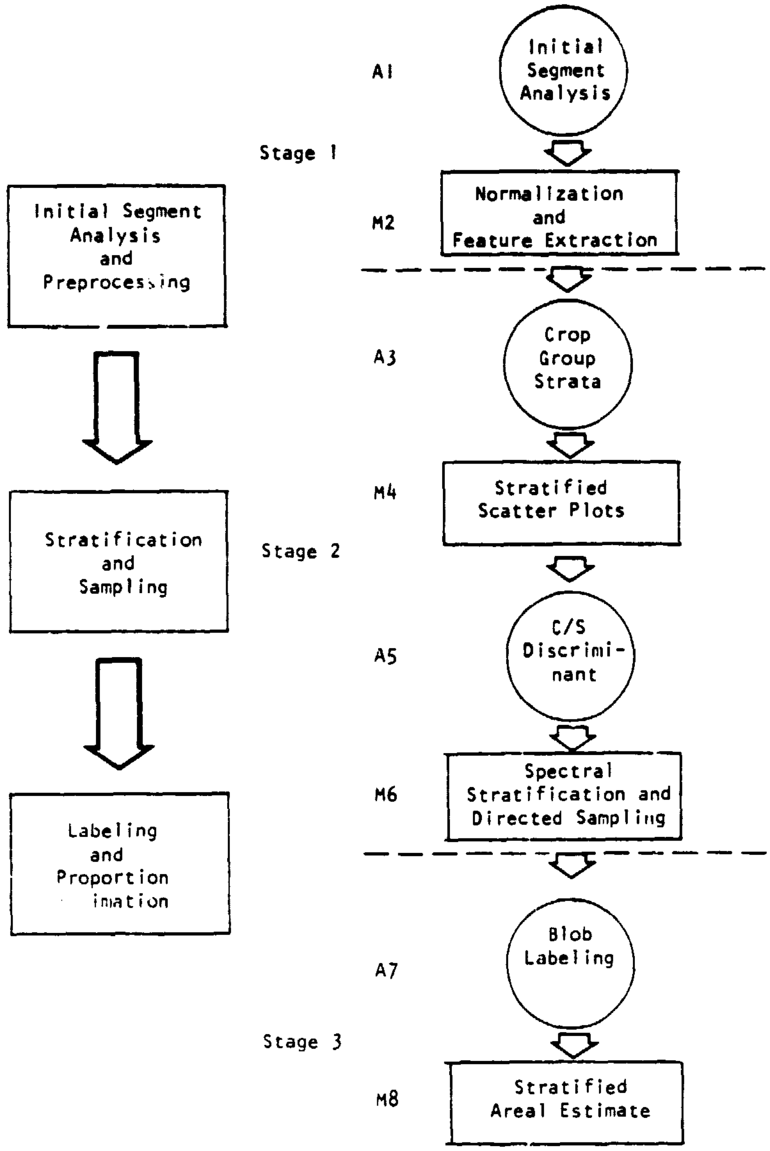
The machine then combines the labeled samples into a final segment wide proportion estimate of the crops observed.

The regional aggregation of segment-level area estimates produced in this manner and the formation of production estimates are functions outside the scope of this classification procedure.

3.2.2.2 Summary Description of C/S-1 Procedure

The flow of the specific activities which make up the U.S. Corn/Soybean Baseline Procedure (C/S-1) is characterized by an integrated, mutually supportive, Analyst/machine effort. The machine performs routine data manipulation functions, supports the Analyst's activities through the production of aids, maintains the data base, and insures statistical objectivity in the estimation process. The Analyst is responsible for data quality assurance through acquisition screening and selection, data verification and adjustment such as in biowindow boundary placement, and data analysis through crop group stratification and target labeling.

The Baseline Procedure can be functionally divided into three major stages as illustrated in Figure 3.2(a). These three stages are (1) segment familiarization and preprocessing, (2) stratification and sampling, and (3) labeling and proportion estimation. The purpose of the first stage is to extract information from both pertinent collateral data and from the Landsat segment image to provide a foundation for the labeling and estimation activities. The second stage, stratification and sampling, results in the identification of targets for labeling and the development of analysis aids that will be used in the blob labeling process. The final stage involves the labeling of a sample of blobs and the aggregation of those samples to a segment-wide proportion estimate.



(a) (b)

FIGURE 3.2. OVERVIEW OF U.S. BASELINE C/S PROCEDURE: (a) MAJOR FUNCTIONS; (b) ANALYST AND MACHINE SEQUENTIAL PROCESSING STEPS

As stated earlier, analyst and machine interact in this procedure. Thus, within each stage it is possible to further subdivide the procedure on the basis of whether an activity is primarily an analyst activity or machine activity. Subdividing the procedure this way has resulted in breaking it down into eight basic steps. These steps are shown in Figure 3.2(b). The number of each step is preceded by an "A" or an "M", indicating whether it is primarily an analyst or machine function, respectively.

A description of the activities that make up each of these steps is presented next.

STAGE 1: SEGMENT FAMILIARIZATION AND PREPROCESSING

Step A1: Initial Segment Analysis

This step is an analyst function and consists of four separate activities:

Segment Familiarization. If an analyst is not familiar with the environmental and cultural characteristics of a region in which a segment is located, the analyst should study the materials supplied in (1) the analyst information manual, and (2) the segment analysis packet.

Data Screening. Through the use of standard imagery products (PFC 1 and PFC 3), acquisitions are visually screened and those with excessive cloud cover, heavy haze and bad data are deleted. This function is designed to eliminate unusable acquisitions from further consideration.

Crop Calendar Analysis. Crop calendars are used to identify the expected phenological patterns for different crops and define bio-windows for those crops in the geographical area where the segment is located. This requires the use of the best available phenological crop calendar. The Analyst compares the normal phenological crop calendar for the area to the apparent spectral development of a crop by associating each acquisition with a crop growth stage. If there are differences between them, then the normal phenological crop calendar is adjusted by the Analyst to conform to the crop development pattern observed for the year in which the Landsat data was collected.

Acquisition Selection. A total of up to ten acquisitions may be processed. Based on inputs from the crop calendar analysis and acquisition priority listings, up to seven of these acquisitions are chosen for Temporal Pattern Class (TPC) extraction. These acquisition selections are identified to the computer for machine processing.

Step M2: Normalization and Preprocessing

This step is a machine function and consists of two separate activity sequences:

Normalization. Normalization of spectral data is a process designed to adjust for effects of haze, varying sun angle and sensor calibration, and to screen out clouds and other unusable data. The purpose of this activity is to reduce the effect in the Landsat data of phenomena that are external to, or bear no information with respect to, agricultural factors that are of interest. The goal is to provide the Analyst with products that are consistent between dates with respect to the conditions under which the scene is observed, and thus minimize segment-to-segment variations in signal that are not actually due to development of the crops (See Figure 3.3).

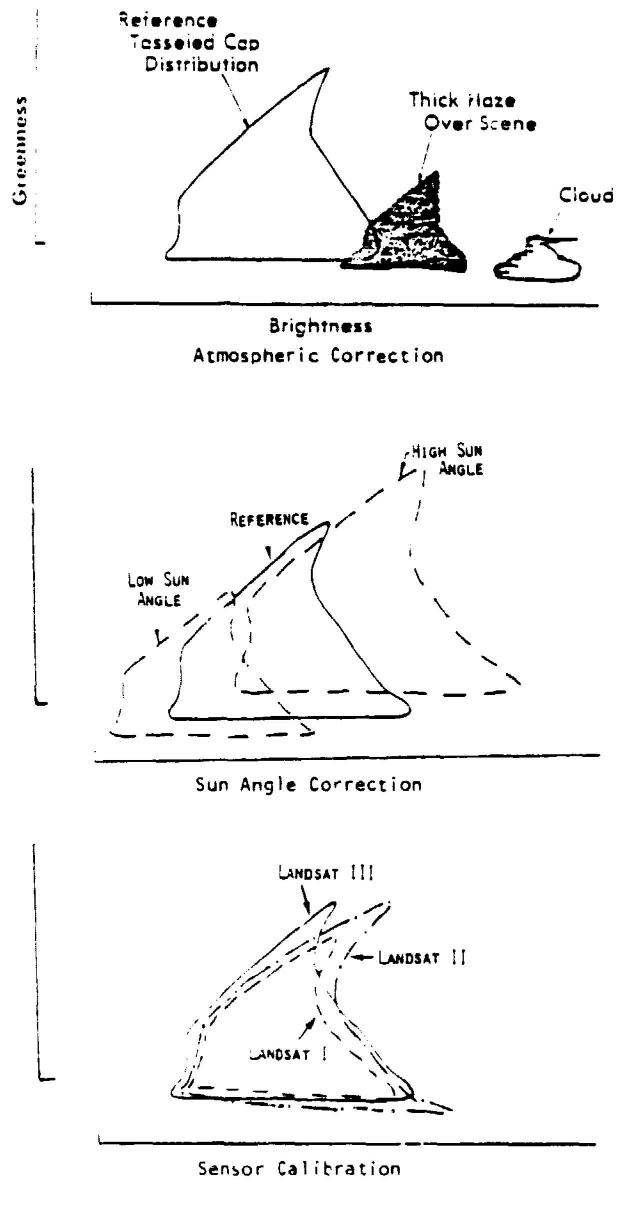


FIGURE 3.3. MSS NORMALIZATION REQUIREMENTS

Spectral/Temporal Feature Extraction. Using normalized spectral/temporal data, features are extracted in this activity sequence that facilitate analysis of agronomic conditions. Specifically, the Tasseled-Cap transformation is computed and a Greenness measure defined as the "Greenness Above Bare Soil" (GRABS) is derived. These features are eventually used for crop discrimination in this procedure. A benefit of this step is that the dimensionality of the data is reduced by a factor of two.

A related activity is the extraction of a Temporal Pattern Class (TPC) for each pixel. A TPC describes the pattern of vegetation development observed for a pixel over the course of the growing season with regard to the number of acquisitions available and the Crop Group Biowindows in which they occur. Thus, each crop group considered in the crop calendar analysis has an expected TPC based on the acquisition history of the segment relative to its idealized phenological development. The result of this activity is a report which summarizes the TPC patterns observed for the segment.

STAGE 2: STRATIFICATION AND SAMPLING

Step A3: Crop Group Stratification

Using information derived from crop calendar analysis and the TPC report generated in Stage 1, the Analyst stratifies the TPCs into major crop groups based on expected patterns for summer crops, small grains, permanent vegetation, and non-vegetated areas. Crop group stratification is used both by the machine in producing the stratified area estimate, and by the Analyst to facilitate the analysis process associated with blob labeling. Of immediate concern is the fact that the summer crop stratum is used to produce a spectral aid, a GRABS vs. Brightness scatterplot.

Step M4: Stratified Scatterplots

Scatterplots of GRABS vs. Brightness are generated for each acquisition using pixels assigned to the pure summer crop stratum or significantly large alternate summer crop subclasses. These plots show the progression of the vegetation phenology of this stratum in the potential crop separation window. The initial use of these stratified scatterplots will be to verify the boundaries of the Separation Window. Only those acquisitions showing a distinct separation in the distribution of points along the 'Green Arm' are to be considered separation acquisitions (See Figure 3.4).

Step A5: Corn/Soybean Discriminant

Using the GRABS vs. Brightness scatterplot of pixels in the summer crop stratum for each available acquisition, the Analyst determines when the best separability between corn and soybean distributions is achieved. Examining crop development along the "Green Arm" the Analyst looks for soybeans to cluster at higher GRABS values than corn. A boundary is placed between these distributions and perpendicular to the Green Arm for each acquisition exhibiting separability. This boundary and associated limiters will be used in preliminary labeling of blob targets as corn or soybeans. At this point the analyst also identifies a subset of acquisitions that are used in defining field-like targets (blobs) (See Figure 3.4).

Step M6: Blobbing, Blob Clustering and Sampling

Blobbing (Target Definition). Field-like targets called blobs are defined. These targets are intended to correspond to farmers' fields and provide candidate labeling targets. Ideally, each target is composed of a single crop type. The machine clusters pixels on the basis of their spectral characteristics and spatial

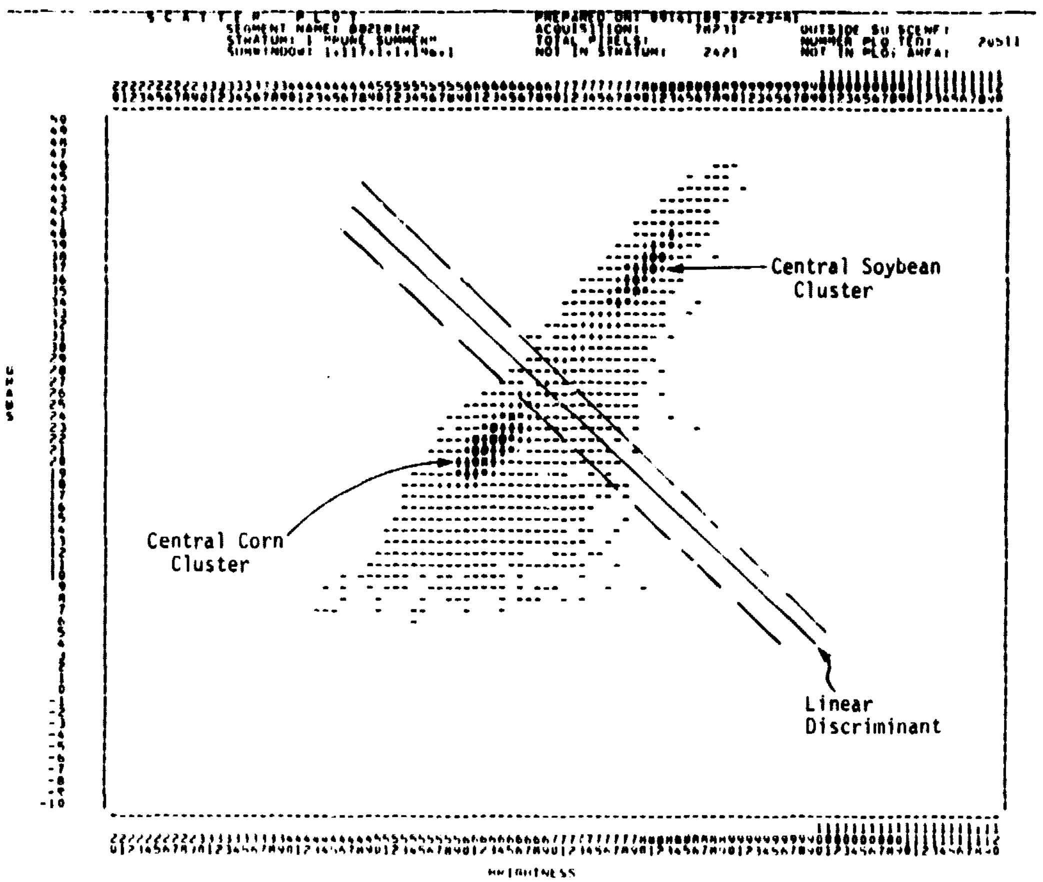


FIGURE 3.4. LINEAR DISCRIMINANT BOUNDARY PLACED BY ANALYST TO SEPARATE CORN AND SOYBEAN CLUSTERS

position. Pixels grouped in a single blob must be spectrally similar and spatially contiguous. Once the blobs are formed they are separated into two groups according to their size. The first group, called "big blobs", consists of all blobs that have at least one pixel in their interior (i.e., one pixel left when a one pixel boundary is stripped off the blob). The second group, or "little blobs", has no interior. Only big blobs are candidate labeling targets. This segregation is carried out in order to isolate mixture pixels and very small fields which prove to be poor labeling targets. Each blob, big or little, is assigned to crop group strata according to the vegetative temporal pattern of their spectral means. This is done by the machine based on the temporal pattern class assignments previously defined by the Analyst.

Blob Clustering. Since it is too time-consuming to label all big blobs, it is desirable to produce a sample of blobs for labeling that would best represent the entire population. In order to realize a gain in sampling efficiency, big blobs are grouped into smaller strata within each crop group. An unsupervised clustering algorithm is used to group the blobs into spectrally homogeneous strata that ideally are homogeneous with respect to crop type, as well.

Sampling. Once strata are formed, a specified number of blobs are selected for labeling. The sample is allocated proportional to the size, in pixels, of each stratum. Since blobs are of different sizes, the Midzuno technique [13] is used to select a sample that is an unbiased representation of each stratum. Once the sample is selected, a number of labeling aids are produced for the Analyst including GRABS vs. Time and GRABS vs. Brightness plots, a PFC overlay identifying the blobs to be labeled, and other diagnostic statistics.

STAGE 3: BLOB LABELING AND PROPORTION ESTIMATION

Step A7: Blob Labeling

Using aids produced by the machine, the Analyst follows a well-defined decision logic to label each sampled blob according to its major crop group (see Figure 3.5). The crop group stratification assignment is used as an initial indicator of crop group. This assignment is refined using additional available information. The resultant label will be either "Summer Crop" or "Non-Summer Crop".

If supported by the segment acquisition history, the Analyst will also label each blob sampled according to its crop type, in particular "corn", "soybean", or "other". Again the Analyst makes use of a well-defined decision logic (See Figure 3.6). Since this procedure was designed for the Corn Belt where corn and soybeans are dominant, other summer crops are not discriminated. In addition to crop labels, the Analyst assigns a confidence to the label to indicate an expectation regarding the accuracy of the label. These labels are provided to the machine for the final estimate of crop area proportions.

Step M8: Estimation

Stratified Area Estimate. A weighted aggregation of the labels of the sampled blobs in each spectral stratum results in an estimate of summer crop area, or, if information is sufficient for crop type labeling, corn and soybean area, for each stratum. An estimate is then produced for each crop group stratum by a simple weighted aggregation of the spectral stratum estimates.

Segment Proportion Estimates. Each crop group stratum was previously assigned an estimate of summer crop area, or, corn and soybean area, according to a sample of big blobs. The segment

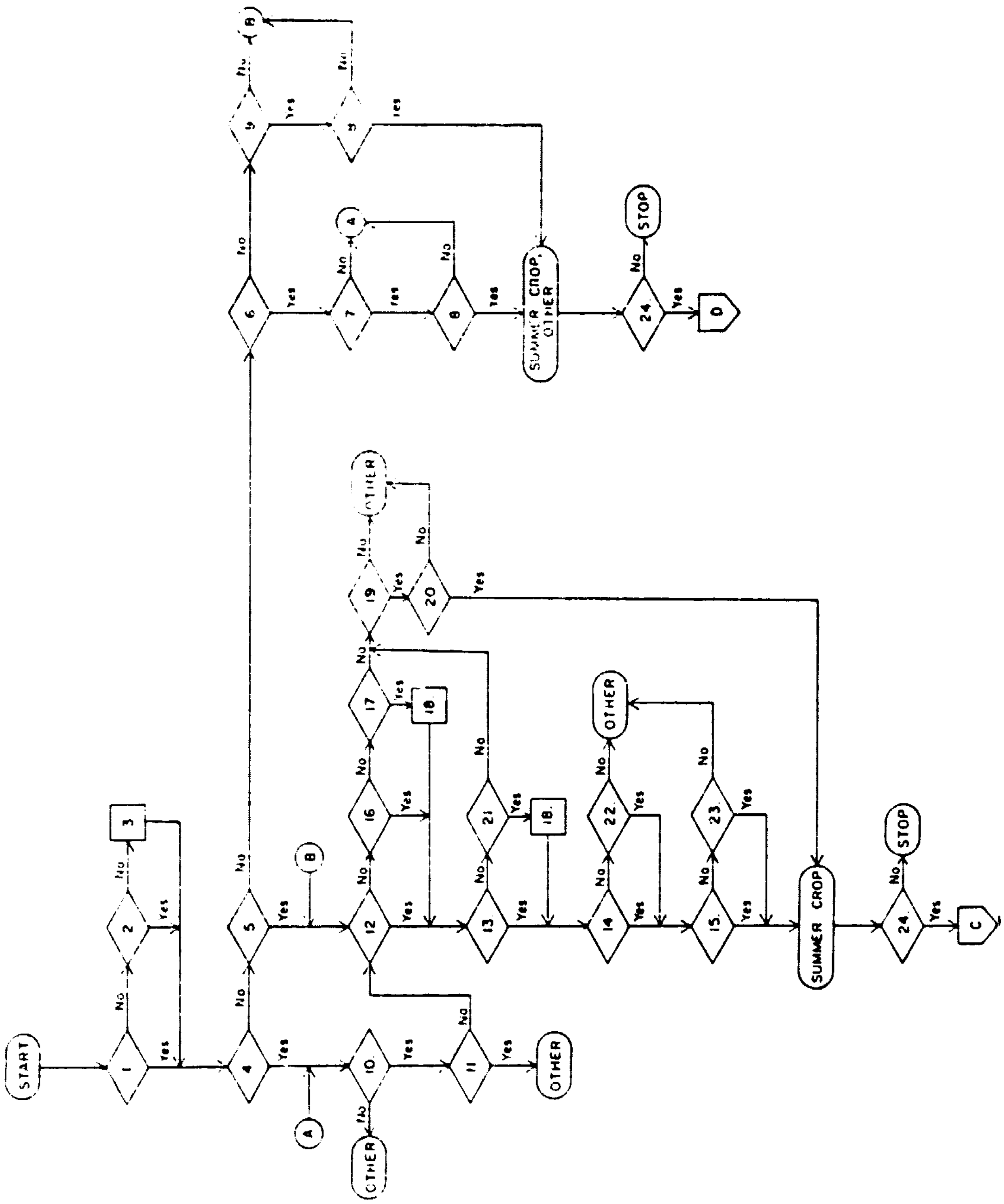


FIGURE 3.5. CROP GROUP DECISION LOGIC

ORIGINAL PAGE IS
OF POOR QUALITY

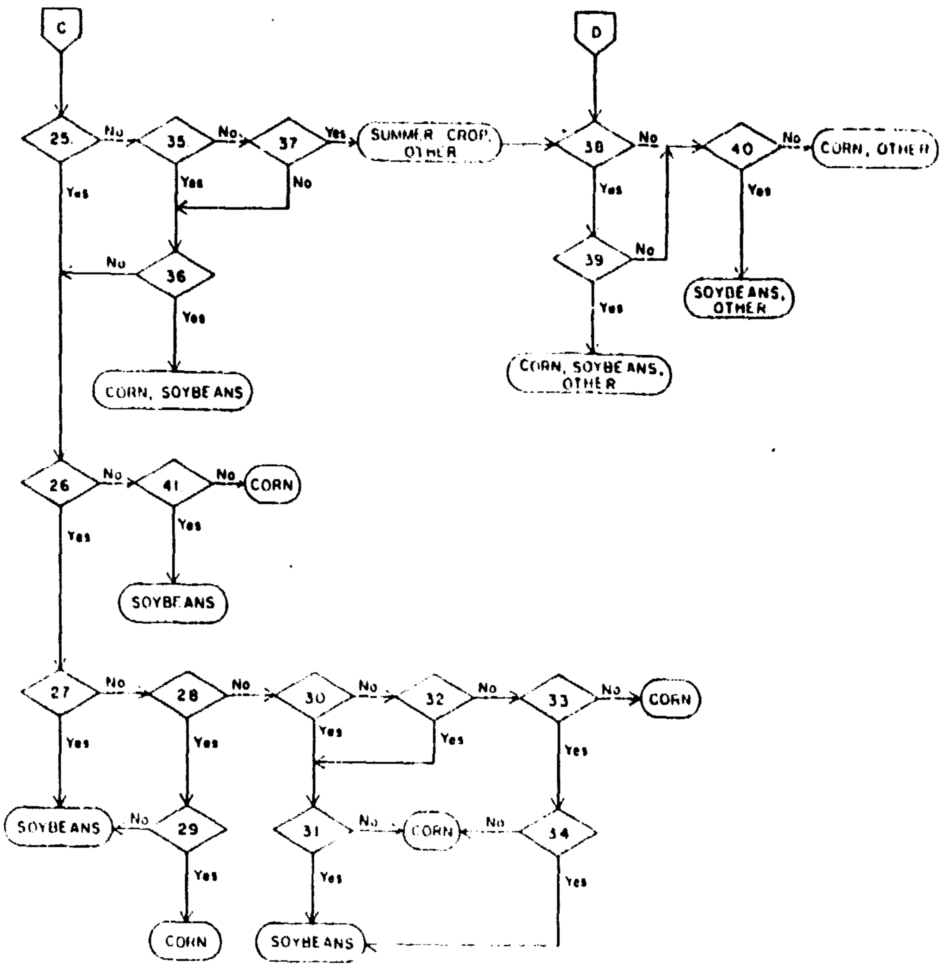


FIGURE 3.6. CROP TYPE DECISION LOGIC

area estimate is produced by extending the crop group stratum estimates to the in-stratum unsampled (little) blobs, and then aggregating the overall stratum estimates. In this process, the weights used are formed from the total number of pixels in each blob. Figure 3.7 graphically illustrates the estimation process.

3.2.2.3 Evaluation of C/S-1 Procedure

Overall Results

Results from the 46 segment processings performed in Phase 1 of the Pilot indicated that while the estimates for summer crops as an aggregate were within 1.5% relative to the true (see Figures 3.8(a) and 3.8(b)), corn was significantly overestimated and soybeans were underestimated by a similar amount. Table 3.1 identifies statistical measures used.

To eliminate the effect analyst labels might cause on the final segment estimates, the blobs were given actual labels from digitized ground truth. With these labels the estimates illustrated in Figures 3.9(a) and 3.9(b) were produced. While these results are a substantial improvement over the analyst-produced results, especially in terms of variance, significant error still remained, indicating that the errors were both machine and analyst induced.

Detailed Analysis

In order to gain insight into the sources of these errors as soon as possible, the initial 11 segments processed were selected for in-depth analysis. As the pilot processings progressed, four additional segments were included in the study. As it eventually turned out, the particular 15 segments analyzed exhibited poorer soybean estimates than the ensemble of processing, mainly because of the unusual conditions encountered

ORIGINAL PAGE IS OF POOR QUALITY

P = PROPORTION
A = AREA (OR NO. PIXELS)

STRATA ACREAGES | CLUSTER P_{11} | SAMPLED (fields) | UNSAMPLED (mixed pixels)

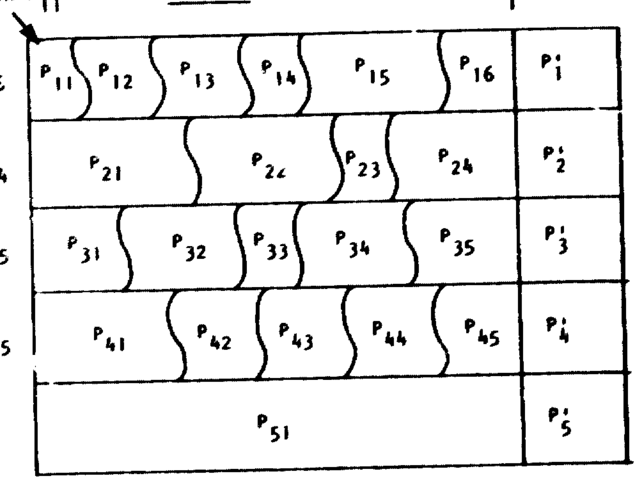
$P_1A_1 = P_{11}A_{11} + P_{12}A_{12} + \dots + P_{16}A_{16}$

$P_2A_2 = P_{21}A_{21} + P_{22}A_{22} + \dots + P_{24}A_{24}$

$P_3A_3 = P_{31}A_{31} + P_{32}A_{32} + \dots + P_{35}A_{35}$

$P_4A_4 = P_{41}A_{41} + P_{42}A_{42} + \dots + P_{45}A_{45}$

$P_5A_5 = P_{51}A_{51}$



$P_1' = P_1$ Summer

$P_2' = P_2$ Permanent Vegetation

$P_3' = P_3$ Grain

$P_4' = P_4$ Non-Vegetation

$P_5' = P_5$ Unknown

SEGMENT PROPORTION

$$P = \frac{P_1(A_1 + A_1') + P_2(A_2 + A_2') + P_3(A_3 + A_3') + P_4(A_4 + A_4') + P_4(A_4 + A_4')}{A_1 + A_1' + A_2 + A_2' + A_3 + A_3' + A_4 + A_4' + A_5 + A_5'}$$

FIGURE 3.7. C/S-1 CORN/SOYBEANS ESTIMATION

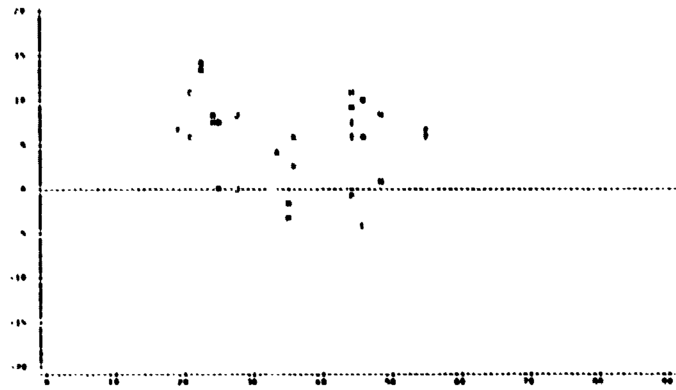
ORIGINAL PAGE IS
OF POOR QUALITY

SUMMARY STATISTICS

Segment Proportion Estimation Error Vs. Ground Truth Proportions

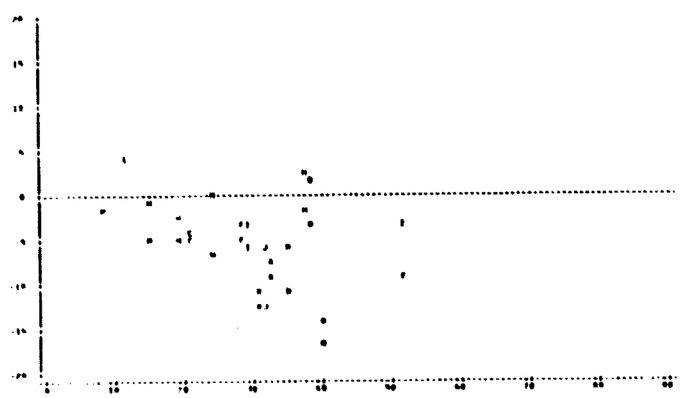
$\bar{e}$	5.3
S_e	4.8
M.A.E.	6.2
R.M.E.	15.0
$\bar{p}$	35.2
n	30

CORN



$\bar{e}$	-5.5
S_e	4.9
M.A.E.	6.1
R.M.E.	-18.7
$\bar{p}$	29.7
n	30

SOYBEANS



$\bar{e}$	-0.9
S_e	6.5
M.A.E.	4.9
R.M.E.	-1.5
$\bar{p}$	63.8
n	36

SUMMER CROPS

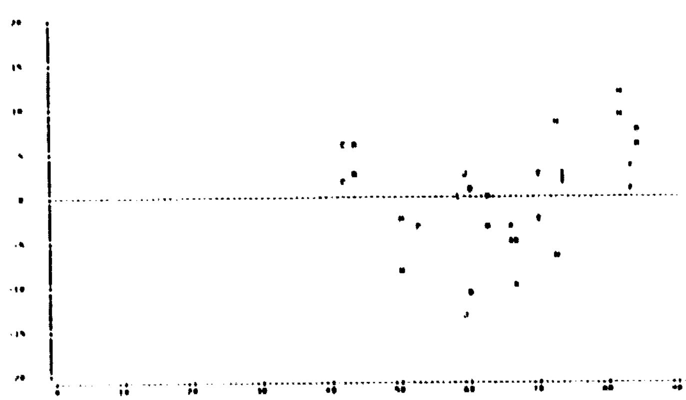


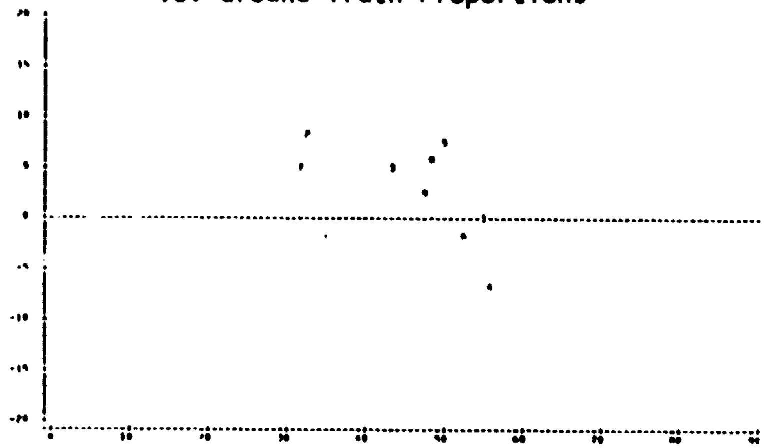
FIGURE 3.8(a). PERFORMANCE OF C/S-1 IN CROP YEAR 1978 USING ANALYST LABELS

SUMMARY STATISTICS

Segment Proportion Estimation Error vs. Ground Truth Proportions

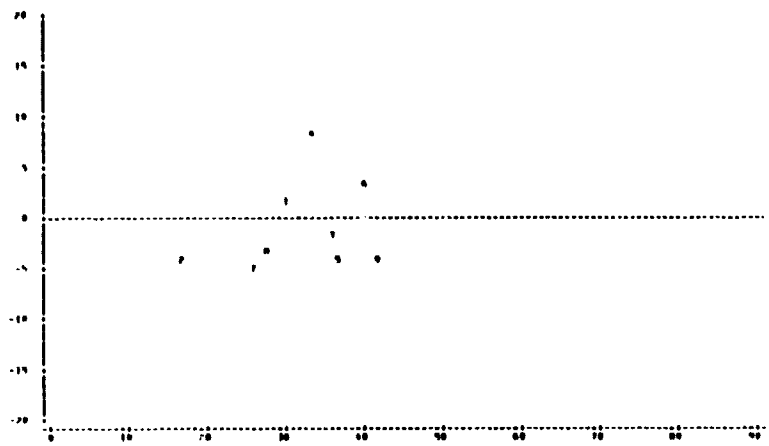
$\bar{e}$	2.8
S_e	4.8
M.A.E.	4.7
R.M.E.	
$\bar{p}$	46.3
n	9

CORN



$\bar{e}$	-1.0
S_e	4.5
M.A.E.	3.9
R.M.E.	-3.0
$\bar{p}$	32.0
n	9

SOYBEANS



$\bar{e}$	1.4
S_e	2.0
M.A.E.	2.1
R.M.E.	1.8
$\bar{p}$	77.9
n	10

SUMMER CROPS

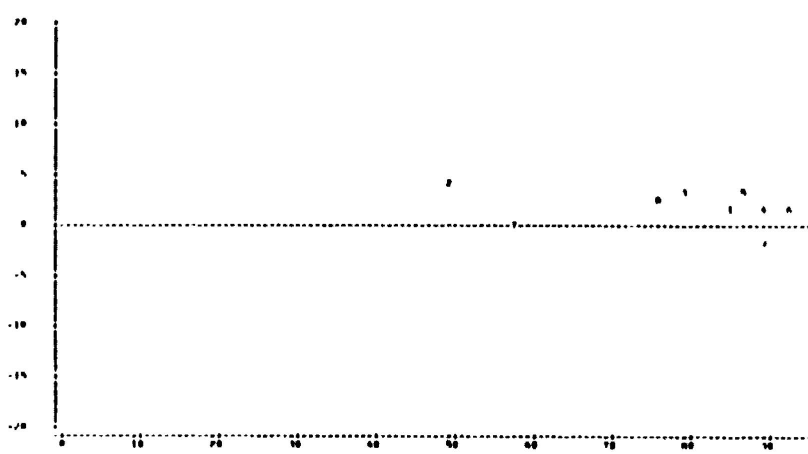


FIGURE 3.8(b). PERFORMANCE OF C/S-1 IN CROP YEAR 1979 USING ANALYST LABELS

TABLE 3.1. STANDARD STATISTICAL MEASURES OF AREA PROPORTION ESTIMATION PERFORMANCE FOR n SEGMENT PROCESSINGS

MEAN ERROR ($\bar{e}$): $\sum_{i=1}^{n} e_i/n = \bar{\hat{P}} - \bar{P}$

STANDARD DEVIATION OF ERROR (s_e): $\left[\sum_{i=1}^{n} (e_i - \bar{e})^2/(n - 1)\right]^{1/2}$

MEAN ABSOLUTE ERROR (M.A.E.): $\sum_{i=1}^{n} |e_i|/n$

RELATIVE MEAN ERROR (R.M.E.): $\bar{e}/\bar{P}$

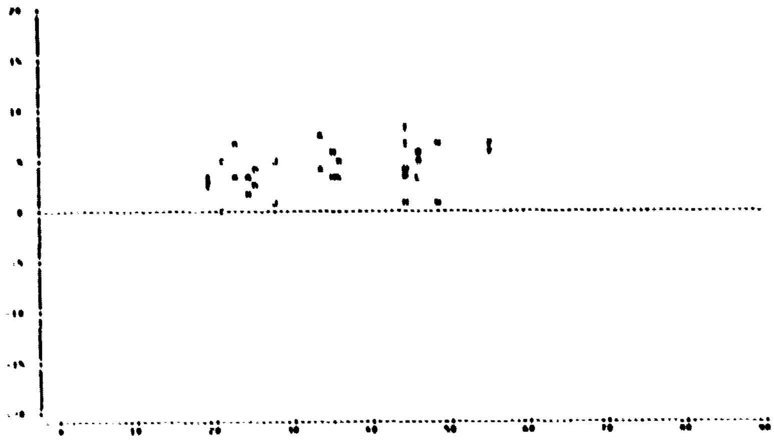
GROUND TRUTH PROPORTION FOR iTH SEGMENT:	P_i
ESTIMATED PROPORTION FOR iTH SEGMENT:	$\hat{P}_i$
ERROR FOR iTH SEGMENT:	$e_i = \hat{P}_i - P_i$
ABSOLUTE ERROR FOR iTH SEGMENT:	$\lvert e_i \rvert$
MEAN GROUND TRUTH PROPORTION:	$\bar{P} = \sum_{i=1}^{n} P_i/n$
MEAN ESTIMATED PROPORTION:	$\bar{\hat{P}} = \sum_{i=1}^{n} \hat{P}_i/n$

SUMMARY STATISTICS

Segment Proportion Estimation Error Vs. Ground Truth Proportions

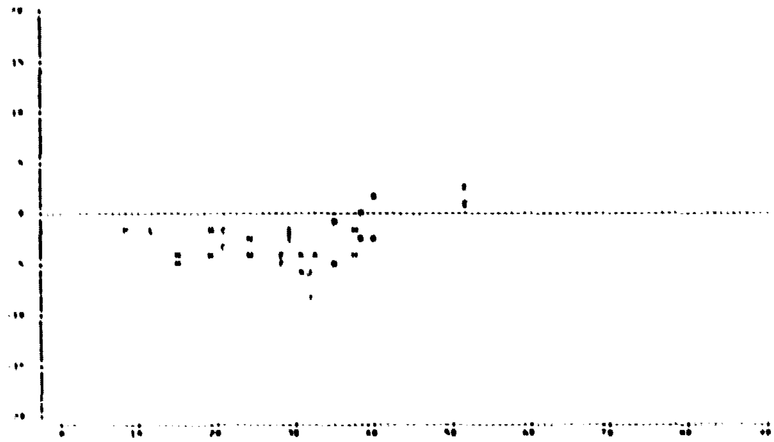
$\bar{e}$	3.9
S_e	2.1
M.A.E.	3.9
R.M.E.	11.1
$\bar{p}$	35.2
n	30

CORN



$\bar{e}$	3.0
S_e	2.3
M.A.E.	3.2
R.M.E.	-10.1
$\bar{p}$	29.7
n	30

SOYBEANS



$\bar{e}$	0.6
S_e	3.0
M.A.E.	2.6
R.M.E.	0.9
$\bar{p}$	63.8
n	36

SUMMER CROPS

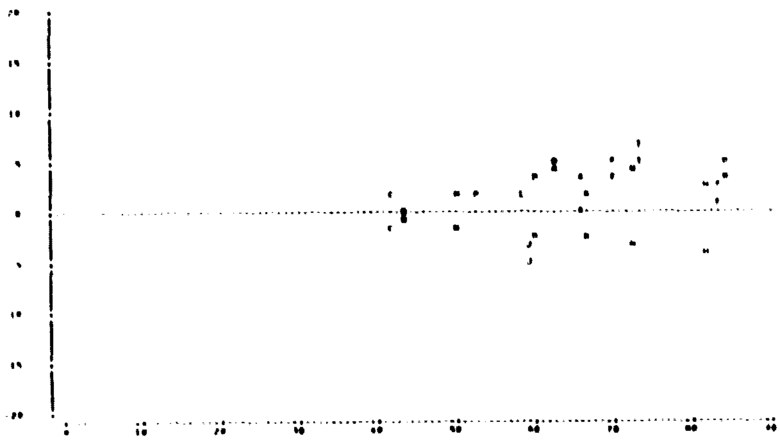


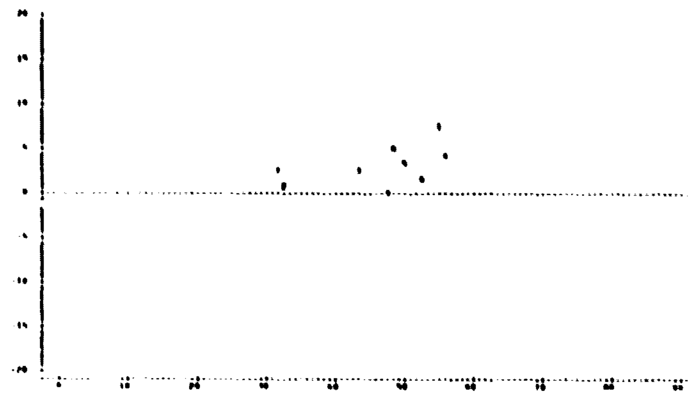
FIGURE 3.9(a). PERFORMANCE OF C/S-1 IN CROP YEAR 1978 USING GROUND TRUTH LABELS

SUMMARY STATISTICS

Segment Proportion Estimation Error Vs. Ground Truth Proportions

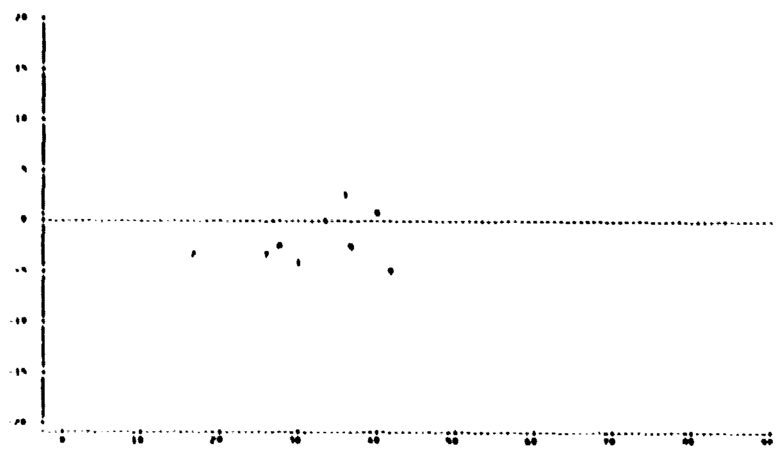
$\bar{e}$	3.2
S_e	2.2
M.A.E.	3.2
R.M.E.	6.9
$\bar{p}$	46.3
n	9

CORN



$\bar{e}$	-1.9
S_e	2.5
M.A.E.	2.7
R.M.E.	-6.1
$\bar{p}$	32.0
n	9

SOYBEANS



$\bar{e}$	1.6
S_e	3.3
M.A.E.	3.2
R.M.E.	2.1
$\bar{p}$	77.9
n	10

SUMMER CROPS

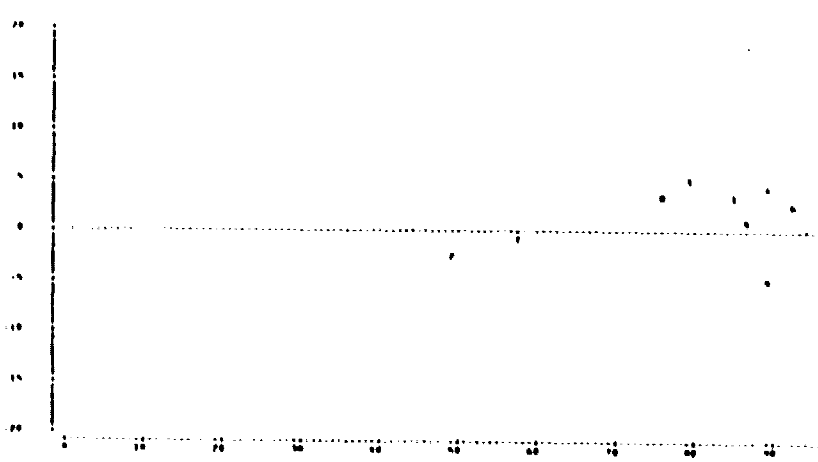


FIGURE 3.9(b). PERFORMANCE OF C/S-1 IN CROP YEAR 1979 USING GROUND TRUTH LABELS

in these segments with regard to acquisition histories, environmental conditions during the growing season, and an atypical case of double cropping of soybeans that was encountered. Nevertheless, the general performance characteristics of the procedure observed in the analysis of these 15 segments represent the same trends later observed in the later analysis by LEMSCO of the full set of 46 segment processings, and so formed a reasonable basis for investigating the source of errors associated with the C/S-1 procedure.

The approach ERIM adopted to investigate the sources of error can be described as a series of subcomponent level evaluations of C/S-1. This approach was selected because it made it possible to isolate (1) how much error is built into the automated machine side of the system, versus that contributed by labeling, and (2) even more specifically how much error is contributed by each step of the estimation process carried out by the machine. The effects of labeling error were removed by substituting ground truth information for the labels normally furnished by the analyst. Thus, machine functions were analyzed in the absence of other error sources, and observed deviations between the machine's crop proportion estimates for the segment and the true crop proportion estimates, as computed from ground truth, could be attributed to deficiencies in the estimation procedure.

Thus, tests of each major step in the flow of estimation-related activities were conducted. These tests, it was hoped, would show the amount of error introduced into the final crop proportion estimate due to the error contribution of each step of the procedure. This resulted in the evaluation of the following six strategies of the estimation procedure:

1) The effect of using only the big blobs (those with interior pixels) of the segment and their boundaries to produce the estimate;

2) The effect of using only the interiors of the big blobs of the segment to produce the estimate;

3) The effect of using only certain allowable mixture proportions when describing the composition of the interiors of mixed blobs;

4) All of the above conditions applied to only a sample of the big blobs;

5) All of the above conditions, with analyst labels substituted for ground truth;

6) All of the above, plus the effect of adding in the little blobs, which constitute an unsampled stratum.

Behind each of these strategies there is an assumption. So, by comparing the actual crop proportion estimates of a segment with those produced using these strategies we have a way of testing the following assumptions:

1) That the pixels contained in big blobs and their boundaries are a representative sample of all the pixels in the segment;

2) That the interior pixels of a blob are representative of the entire blob;

3) That the proportion of crop types found in mixed blobs can be accurately measured using a system that allows designating mixture in terms of halves and thirds of a blob;

4) That a sample of available big blobs, produced by the sampling procedure used, yields an unbiased estimate of the crop proportion found in all big blobs;

5) That analyst labels are accurate;

6) That the little blobs (the unsampled stratum) have the same crop proportions as the big blobs in the same crop group, and taking advantage of this adjusts for crop error due to assumption #1.

Figure 3.10 illustrates the results of this study with a plot of the cumulative estimation error which results at the end of each step. From the plot the following can be seen:

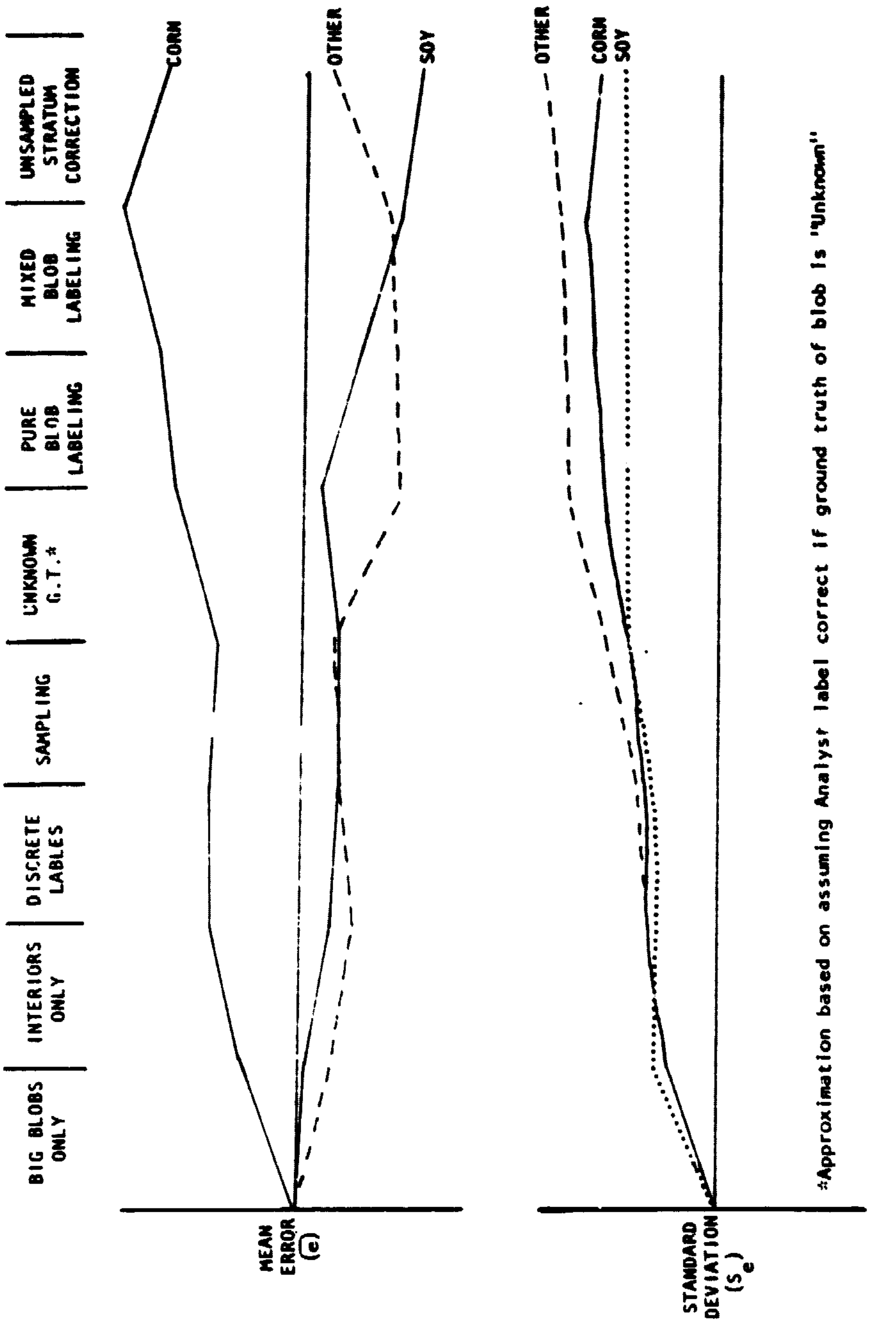
1) The big blobs alone are not representative of the entire segment. Further analysis indicated that typical corn blobs were bigger than typical soybean and non-summer blobs, and that the little blobs and smaller big blobs were predominantly non-corn. While the available evidence indicates that actual corn field are bigger than actual soybean and non-summer fields on the average, the difference in blob size for different crops may be a phenomenon associated witn the manner in which the BLOB algorithm works. Analysis of BLOB in these terms is discussed in Section 3.3.3.

2) Extending the label of the blob interior to the blob boundary is not an unbiased assignment. In particular, it was determined that the boundaries of corn blobs were "dirtier" than the boundaries of non-corn blobs. This is due to the central position corn occupies spectrally between non-summer crops and soybeans; and to the fact that the BLOB algorithm grows a field-like target until a variance threshold is exceeded. The spectral position of corn will tend to make mixed signatures look like corn (non-summer + soybeans will be too green for non-summer, not green enough for soybeans; corn + soybeans will look like green corn or weak soybeans), and the lower variance observed in corn blobs will tend to make them grow excessively.

3) Forcing a blob label to be quantized into fractional parts of 1/3 or more had no significant effect on the results of the procedure.

4) The Midzuno sampling was unbiased in implementation as it is by theorem, and added variance to the estimate, as expected.

Up to this point in the analysis, all labeling of blobs had been done using digitized ground truth, with non-inventoried pixels being



*Approximation based on assuming Analyst label correct if ground truth of blob is "Unknown"

FIGURE 3.10. PROGRESSIVE CONTRIBUTION TO ERROR BY SUBCOMPONENT

designated as "other". Additional analysis required analyst labels, which posed a problem when all or part of the blobs in question was "non-inventoried". To assess the uncertainty introduced into the estimate by this "unknown" ground truth, the following approach was taken:

An estimate was produced using ground truth labels for the sampled blobs, with "non-inventoried" pixels counted as "other". The estimate was recalculated, this time using the analyst label for each blob which was 50% or more "non-inventoried", in effect assuming that the analyst labels for those blobs were correct. This estimate was then the base with which to compose estimates using analyst labels exclusively.

5) The introduction of analyst labels into the procedure added significant bias, particularly with respect to corn and soybeans. To gain additional insight into the nature of the labeling errors, the blobs were divided into the set of all blobs with at least 5/6 of their interior pixels of the same crop class, and the set containing all other blobs. These strata were designated "pure" and "mixed" blobs, respectively.

Analysis indicated that although the 80% of the sampled blobs which were "pure" were labeled with good accuracy (96% for corn, 88% for soybeans and 92% for non-summer crops), the 20% of the blobs which were mixed contributed 50% to 70% of the final error caused by labeling.

Two basic factors contributed to the poor labeling performance on mixed blobs: 1) too many mixed blobs were being created by the BLOB algorithm, and 2) the analysts were only detecting approximately 10% of the mixed targets. It appeared that this problem resulted primarily from non-optimal acquisition selections and to a lesser degree, to inherent limitations on the separability of the crops using Landsat MSS.

6) The correction for the unsampled stratum performed as desired in correcting summer/non-summer bias introduced by sampling big blob

interiors only, but was inadequate in dealing with crop type (corn/soybean) corrections. This appeared to be primarily the result of assignment of little blobs to crop group strata, which did not allow for corn/soybean discrimination. It was also observed that this bias correction step was the major contributor to the variance of the estimates.

The above analysis led to the following major conclusions:

1) The labeling targets defined by C/S-1 were of unsatisfactory quality. In particular, too many impure blobs were being formed, and the analysts were not able to detect these blobs as mixed.

2) The correction for the bias introduced by sampling only from big blobs was inadequate with respect to crop type, but performed as desired in eliminating crop group bias.

On the basis of these findings, a set of modifications to C/S-1 was proposed which it was felt would remedy the most serious of the procedure's deficiencies. These modifications and the procedure resulting from their implementation are described in the following section.

3.2.3 AUGMENTED BASELINE PROCEDURE (C/S-1A)

3.2.3.1 Description of Procedure

The augmented Baseline Corn and Soybean Procedure, C/S-1A, was developed in response to weaknesses observed in Procedure C/S-1 as detailed in the previous section. Technical specification of C/S-1A is provided in Appendix 1 and procedures are documented in [52]. The major areas targeted for development were the unsampled stratum bias correction, and target definition and labeling. Additional modifications aimed at increasing the consistency and efficiency of the procedure were also implemented, although they were not the primary focus of the development effort. The basic structure of the procedure remained unchanged, with most of the modifications being a continuation of development along the original philosophical lines.

Development directed at reducing the mixed blob labeling problem proceeded along two lines: (1) modifications which decreased the number of mixed blobs, and (2) modifications which improved the accuracy with which the remaining mixed blobs were labeled. To reduce the number of mixed blobs created, the blob acquisition guidelines were clarified (the importance of an acquisition in the corn/soybean separation window to reduce corn/soybean mixtures was emphasized, and the use of an acquisition prior to summer crop emergence to reduce summer crop/other mixtures was recommended). Additionally, the decision rule in BLOB was modified to apply acquisition-by-acquisition thresholds, as well as a threshold based on averages over all acquisitions.

To improve the detection and labeling of mixed blobs, a machine procedure for automatic detection of potentially mixed blobs was developed, and the labeling logic was modified to label those blobs flagged as potentially mixed on a pixel-by-pixel basis.

Another important modification was to automate those parts of the decision logic that were completely objective. This resulted in a segment-specific set of reference crop profiles for the analyst to use as references, as well as a decreased number of blobs that the analyst had to label. This modification allowed the machine to label approximately 50% of the blobs with a high level of confidence (about 95% accuracy).

It was observed in the analysis of the C/S-1 test results that the nature of the bias problem associated with the unsampled stratum was primarily a corn/soybean problem, as opposed to a summer crop/other problem. Thus the bias correction step was modified so that little blobs were assigned to a stratum within a DFS, instead of assigning them to the DFS alone. The rationale behind this modification was that the sub-DFS strata allowed crop type stratification while DFS is simple a crop group stratification.

An additional modification aimed primarily at decreasing the time required to run the procedure was the automation of the assignment of TPC's to DFS. During the first phase of the pilot experiment it was found that this essentially rote step was one of the most tedious and error prone activities performed by the analyst. The automation was performed by developing a machine procedure which precisely followed the objective, well defined logic which the analysts had employed.

A summary of the modifications to C/S-1 and the observed problems which motivated the modifications is given in Table 3.2. Appendix 1 provides a detailed specification of the subcomponents comprising Procedure C/S-1A.

3.2.3.2 Evaluation of C/S-1A

Evaluation of the C/S-1 subcomponents that were modified for use in C/S-1A was performed by ERIM to determine the performance improvement which could be expected. Three major tests were performed. They were: target definition, automatic labeling, and the unsampled stratum correction. Due to resource constraints, it was not possible to use all 39 segment processings for each test, and so some tests were performed using only a subset of these 39. These subcomponent evaluations indicate C/S-1A possesses a potential for improvement in segment proportion estimates over C/S-1. End-to-end performance will be determined during Phase 2 of the Pilot, which will be initiated in FY1982. Descriptions of these tests and the results follow.

TABLE 3.2. C/S-1A MODIFICATIONS FROM C/S-1

Weakness in C/S-1	Modification	Evaluated
Labeling Performance		
1. Inconsistent labeling of pure targets	1. Machine labels "classic" targets, partially labels remaining targets	X
2. Misdetection of crops with two vegetative phases	2. Labeling logic refined, examples given	
3. Few mixed targets detected	3. Machine identifies potentially mixed blobs	X
4. Poor labeling performance on mixed blobs	4. Label selected pixels from mixed blob, not blob mean	
5. DFS assignment tedious and error prone	5. Automated DFS	X
Machine Performance		
1. Target definition	1. Reduce number of mixed blobs 1) Improved acquisition selection 2) Modified Blob algorithm	X
2. Biased treatment unsampled stratum	2. Assign little blobs to clusters	X

Target Definition

To determine the improvement, if any, in target definition realized by a modified BLOB rule and clarified acquisition selection guidelines, blobs were produced two different ways and compared. In one case, the original C/S-1 BLOB rule was applied to acquisitions selected during Phase 1 of the Pilot; in the second case, the C/S-1A BLOB rule was applied to acquisitions selected using the C/S-1A acquisition selection guidelines. Both sets of blobs were analyzed in terms of interior purity and the proportion of the scene covered by each of the blob interiors, blob edges, and little blobs. The results of this study are presented in Table 3.3.

From these results, we can conclude that the modifications in C/S-1A have had the intended effect, i.e., the analysts are now presented with labeling targets of higher purity than they experienced with C/S-1. As a consequence, however, the size of the unsampled stratum (little blobs) has increased significantly, placing even greater importance on the proper treatment of this stratum.

Automatic Labeler

The automatic labeling subcomponent was evaluated in a test conducted on blobs created by C/S-1 during Phase 1 of the Pilot. The automatic labeler labeled those 60% of the targets that were >5/6 pure, achieving 96% accuracy for crop type and 98% accuracy for crop group.

Because the automatic labeler requires a corn/soybean discriminant defined in terms of maximum GRABS vs. Brightness, the discriminants identified in the C/S-1 processings were not usable due to being acquisition specific. As a compromise, a standard discriminant value of 64.0 was used for all processings in this test. This value has been shown to provide good results over a large number of segments in the past (See Table 3.4).

TABLE 3.3. TARGET DEFINITION SUBCOMPONENT

	Original C/S-1 (conducted on 5 segments)	New Acquisitions, New Blobbing Rule
Blob Interior Purity	87.2%	93.6%
% of Scene one:		
• Big blob interiors	36.0%	24.4%
• Big blob edges	52.0%	46.8%
• Little blobs	12.0%	27.8%

TABLE 3.4. C/S-1A AUTOMATIC LABELER PERFORMANCE

Crop Type (39 processings, 1534 blobs labeled)

		GT		
		C	S	O
Labeler	C	96.0		
	S		95.7	
	O			98.1

		GT	
		Summer	Other
Labeler	Summer	97.6	
	Other		98.1

Crop Group* (7 processings, 242 blobs labeled)

		GT		
		C	S	O
Labeler	C	89.4		
	S		82.3	
	O			95.7

		GT	
		Summer	Other
	Summer	96.0	
	Other		95.7

*The C/S-1 procedure did not allow processing to crop type for some segments. Use of the maximum GRABS vs. Brightness discriminant default allows crop type estimates for these segments.

Unsampled Stratum Correction

The modified unsampled stratum bias correction subcomponent was evaluated on 39 Phase 1 Pilot processings by comparing the performance of the C/S-1 and C/S-1A bias correction subcomponent on a set of identical blobs. To prevent contamination of this test by errors in analyst labels for these blobs, labels derived from ground truth were used. These labels were produced by LEMSCO from digitized ground truth.

A comparison of the results of this test is presented in Table 3.5. This comparison indicates that the modification tested produced the desired effect, i.e., the bias remaining after the C/S-1A correction is performed is approximately half that observed when the C/S-1 bias correction procedure is used with identical labels.

3.2.4 STARS

3.2.4.1 Introduction

The Software Technology for Aerospace Remote Sensing system (STARS) was developed by ERIM to fulfill a need for a standardized, controlled environment within which development, testing, processing, and evaluation of image processing procedures could take place.

This system has been successfully used to develop three crop area estimation procedures, support major experiments with two of these procedures, evaluate the procedures, and is being used to develop new techniques for estimating crop area using Landsat data.

3.2.4.2 STARS Design Features

Several design features of STARS make it unique. Key ones include: the manner in which individual modules are relatively independent of one another and of the host operating system, the data management capabilities of STARS, and the status tracking features. These will be discussed in greater detail in the following paragraphs.

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 3.5. COMPARISON OF RESULTS USING GROUND TRUTH LABELS

Year		C/S-1			C/S-1A		
		Corn	Soybean	Summer Crop	Corn	Soybean	Summer Crop
1978	$\bar{e}$	4.12	-2.91	0.91	2.19	-0.96	0.85
	S_e	2.13	2.33	2.95	2.47	2.49	3.23
	n	30	30	36	30	30	36
		Corn	Soybean	Summer Crop	Corn	Soybean	Summer Crop
1979	$\bar{e}$	3.15	-1.94	1.50	1.90	-1.22	1.21
	S_e	2.18	2.53	3.27	2.22	2.93	3.73
	n	9	9	10	9	9	10

Due to the fact that software may often be developed on one computer facility with one set of operating conditions, then transferred for use on a different facility with different conditions, there exists a need for the software to be independent of the conditions, such as the operating system, within which it performs. If this independence is not achieved, extensive modifications may be required to the individual modules to allow the transfer to occur, which in turn would require additional testing to verify that modified code.

To achieve this independence from the underlying operating system, a set of system primitives, called System Interface Routines (SIRs) was developed. These primitives are implemented for each system for which STARS is intended to be used. Given these primitives and a compatible compiler, software which interacts with the system only through the SIR's can be transferred from one system to another with no modifications. The functions provided by the SIR's include I/O operations (Create, Open, Close, or Destroy files; Read, Write, Delete records; obtain access to non-file device); memory management (Get space, Free space); and other necessary functions (Get current time/date, query if Batch or Interactive, error handling). In every implementation of the SIR's, the interface with the calling program is unchanged.

In addition to this independence from the operating system, the independence of each application module from all others was required to facilitate testing of individual modules as well as to simplify the substitution of one module for another. To meet this need, all data is passed to/from the application modules via parameter lists, and only a limited number of specialized application modules are permitted to use the SIR's.

For each application (e.g., merge data, produce maps, etc.) an overall controlling program, called a scenario, directs the operation of the individual application modules. This scenario controls the sequence of execution of the application modules and provides all the data management for those modules.

The data management capabilities are provided through a set of primitives available only to the scenario. These primitives access a simplified data base called Collateral Holding And Retrieval Library for Information Extraction (CHARLIE). CHARLIE is composed of a collection of entities, each of which is a FORTRAN-like variable, i.e., scalar, vector, multi-dimensioned array. Each has a descriptor containing the variable name (up to 40 characters), size, shape (dimensions) and mode (Real, Integer, Logical, Complex, Character). The data primitives provide the capability to create an entity in virtual memory, give it initial value, change its shape (e.g., from dimensions of 1, 1, 1 to 3, 4, 117), save it in permanent storage, and retrieve an entity from permanent storage. With these primitives, the burden of data base access is constrained to the scenarios, with the application modules viewing the data as standard FORTRAN variables.

To insure repeatability of results, it is necessary to know which version of each software module was used in the run. It is also useful during development to know what events have occurred up to a given point. To serve this need, STARS has a status tracking capability which records the entry and exit of each application module and scenario, each data base access, any errors detected and major I/O events, such as transferring a file from disk to tape or destroying a file. This log, which is maintained automatically, contains information describing the time of the event and the version of the module.

3.2.4.3 Image Processing on STARS

A primary use for STARS is image processing. With that as a design consideration, two major requirements were identified: images must be processed efficiently, and the system must be adaptable to the various formats images are stored in.

The image processing efficiency results from the "assembly line" processing capability in STARS. In this mode, an image is read, scan line by scan line, and each line (or group of lines) is processed by one or more application routines before the final transformed scan line is saved. This method of image processing minimizes I/O operations, reading each line of the image only once.

The images which are processed may be found in any one of several formats. However, all application modules must share a common view of all images to allow the "assembly line" processing to occur. Therefore, images are viewed by STARS as existing in two forms: Internal and External. The Internal form is the view all application modules have of the image. It is a standardized, one scan line at a time image format. The External image form encompasses all possible formats an image may be stored in External to the STARS environment (e.g., Universal, EROS, etc.). It is the job of Format Service Routines (FSRs) to convert images between Internal and External image formats. With the appropriate FSRs, any external image format may be handled by STARS without modification of application modules.

3.2.4.4 Production Processing in STARS

For STARS to be used for processing in a production environment several criteria must be met. The integrity of the data must be insured, management must have access to processing status, the user interface must be simple, and management must have the ability to allocate storage facilities (disk and tape) as needed.

To maintain the integrity of the data, that data generated by each user of the system is kept physically separate from data belonging to other users. Additionally, the user has no need to know precisely where the data is stored, and in fact the actual names of the data files are hidden from the user. This is all accomplished through a file directory which is maintained by the SIRs. This directory provides the translation between the logical file name, which the application module uses, and the physical location of the data.

The quantity of data generated in many image processing applications is enormous. To maintain all this data in disk files would impose excessive disk requirements on the system, as well as tying up the disk storage with files which may be used very infrequently. Storage of data on tapes is an obvious alternative, but tape access is relatively slow, and processing multiple images simultaneously would require several tape drives - leading to long waits for an available drive. Even more untenable is the possibility that these several images exist on the same tape, and scan line by scan line processing then becomes nearly impossible.

The solution to this problem is the use of a mixture of tape and disk storage. Data is stored on disk as it is generated, then transferred to tape if it is expected to be inactive for a long time. Prior to using data, the scenario insures the data is on disk, transferring it from tape to disk if necessary. The mechanism used by the scenario to affect these transfers is a simple command, and the scenarios may easily be modified to change the decision of what gets transferred where and when. For example, if disk storage is scarce, the decision could be made to transfer all data to tape immediately after it is generated, then back to disk each time it is needed.

In a production environment where throughput is important, it is essential that the interface between the user and the system be simple, both to minimize errors and rework and to minimize training time for new users. To this end the scenario/command language concept was developed.

The user invokes the scenario through a simple command, then each input is requested with a prompt and verified before being finalized. After all user inputs are received, the user is given a final chance to abort the processing or continue. This approach allows most entry errors to be corrected without needing to await the results of processing. User inputs are requested only for those data which the machine cannot otherwise obtain (e.g., from CHARLIE), minimizing the quantity of user inputs.

To insure the integrity of the overall experiments, management must have access to status information. A management query capability exists in STARS which allows information describing processing status, error conditions, disk and tape status, and intermediate and final processing results to be extracted and placed in a report. The query system also provides a limited "Help" facility which describes the capabilities of the system and the commands necessary to utilize those features.

3.2.4.5 Research and Development in STARS

In its applications to date, STARS has been used primarily as a production processing environment. Another intended use of STARS is in a research or development mode. Although many of the needs of a research user are identical to those of a production user, there are requirements which are in conflict.

The primary difference between the researcher and production user is one of data access. Where the production user wants to process a data set only once, wants measures which prevent modification of that data set, wants use of that data set restricted to himself, and wants a fixed set of simple commands, the researcher may want to process the same data set multiple times with different parameters or modules, and several researchers may want to share a common data set.

To permit this duality, two avenues for development have been established. The primary one is the concept of workspace management, wherein each user maintains data in a separate workspace, but data may be easily transferred from one workspace to another, and a workspace may be shared by multiple users under proper conditions. The second concept is the use of a command language or scenario processor to replace the current scenario modules. This command language would allow a scenario to be easily built by the user, providing much more flexibility than currently exists while retaining the capability for simple, pre-defined commands.

Although these concepts are still under development, STARS has already proven to be useful for research and development of new procedures. The modular construction demanded of application modules makes modification of existing modules simpler and minimizes debugging time.

3.2.4.6 Summary

STARS was designed to provide a controlled environment for image processing procedure development and processing. Software for an area estimation procedure (C/S-1) and its subsequent modifications (C/S-1A) were developed by ERIM and exercised in major experiments at NASA/JSC. A number of additional applications are also available. This process of development and testing provided an excellent basis for the evaluation of the design concepts behind STARS.

The volume of the code developed was considerable - more than 30,000 lines of FORTRAN. The productivity achieved in producing this code was good, and the procedures were transferred from the system on which they were initially developed (the University of Michigan's Amdahl 470/8 using the MTS operating system) to a second system for shakedown testing and user training (CMS on the LARS IBM 3031), and finally to the user's system (CMS on the NASA/JSC EODLS AS/3000). The initial transfer (MTS to LARS) required the rewriting of the SIRs, which comprise less than 10% of the total code. The final transfer (LARS to EODLS) required no modifications. In no case did the application modules or data management routines require any modification.

The procedures were run at JSC by persons who had limited prior computer experience and received minimal training. These users reported STARS to be a smooth running, easy to use system. The management of data and permanent storage was totally transparent to these users.

In the evaluation of these procedures, extraction of both intermediate and final results was greatly simplified through the use of CHARLIE. Additional evaluation capabilities, such as the processing of ground data, were readily developed.

STARS has been shown to successfully meet all of its original design goals, but development of the system should not stop here. Effort should continue in the development of workspace and command language capabilities, and the use of STARS for additional applications should be pursued.

3.3 RESEARCH ON TECHNOLOGY ADAPTATION TO ARGENTINA

In this section we consider five research topics that address known technical needs for Argentina crop inventory. These are introduced in the succeeding paragraphs.

First, work on ground cover classes that are likely to be spectrally similar to corn and soybeans was carried out, both to identify those classes and to begin to study growth and spectral characteristics that may serve to distinguish the cover classes from corn or soybeans. This work was principally carried out at UCB [1].

A second topic examines spectral-temporal features derivable by profile-fitting methods to identify corn/soybean/other discrimination information presented in several types of profile features. This effort is aimed at extracting crop-related information that is not sensitive to extraneous factors such as data acquisition date and thereby working toward procedures that are automatic in that they do not rely on a human analyst.

Due to the growing need to reduce the cost of making crop estimates a third topic presents a double sampling method of combining inexpensively obtained segment level crop estimates with ones that are more expensive and accurate to produce a required estimate. This discussion identifies how targets can be allocated to the crop estimation methods so that estimation error is minimized subject to a fixed total cost. This method is carried out to illustrate a minimum error solution based on one set of cost and budget assumptions. However, the most significant aspect of this work is the method used to set up and solve this type of optimization problem.

Another research area is aimed at improving the targets that are selected for identification in crop inventory procedures. This study consists of attempts to quantify the performance of such targets, identify sources of error (especially bias) that these targets may introduce into an overall procedure, and improve the quality of the system components that form or select potential targets.

The final topic presented in this section summarizes preparation of familiar ground truth products from the data collected by the consortium Argentina mission previously discussed in Section 2.4.2. The methods used for this preparation are emphasized, and a description of the available data is presented. This activity has produced a data base that is available through NASA/JSC for further work in adapting or developing crop estimation technology for Argentina.

3.3.1 CONFUSION CROP RESEARCH

In order to carry out accurate inventory of corn and soybeans in Argentina, it is necessary to deal with inventory conditions present in Argentina that are not present in the U.S., to which inventory techniques have been primarily tuned. One of these conditions is the presence of crops other than corn and soybeans that have the same growing season and other characteristics as corn or soybeans. The ability to understand and distinguish these confusion crops is a key issue in the effort to develop an estimation procedure in Argentina.

Principle Argentina confusion crops are sorghum, sunflowers, and peanuts. Secondary confusion crops are cotton and rice. The regions in which these crops are grown are shown by Figure 3.11. The work described herein deals with sorghum and sunflower confusion crops only. To date, no conclusive keys to eliminating these crops from the confusion category have been found, although several insights have been gained.

Corn and Soybean Features

The current inventory technologies are based on several discriminating features related to the spectral and temporal development of

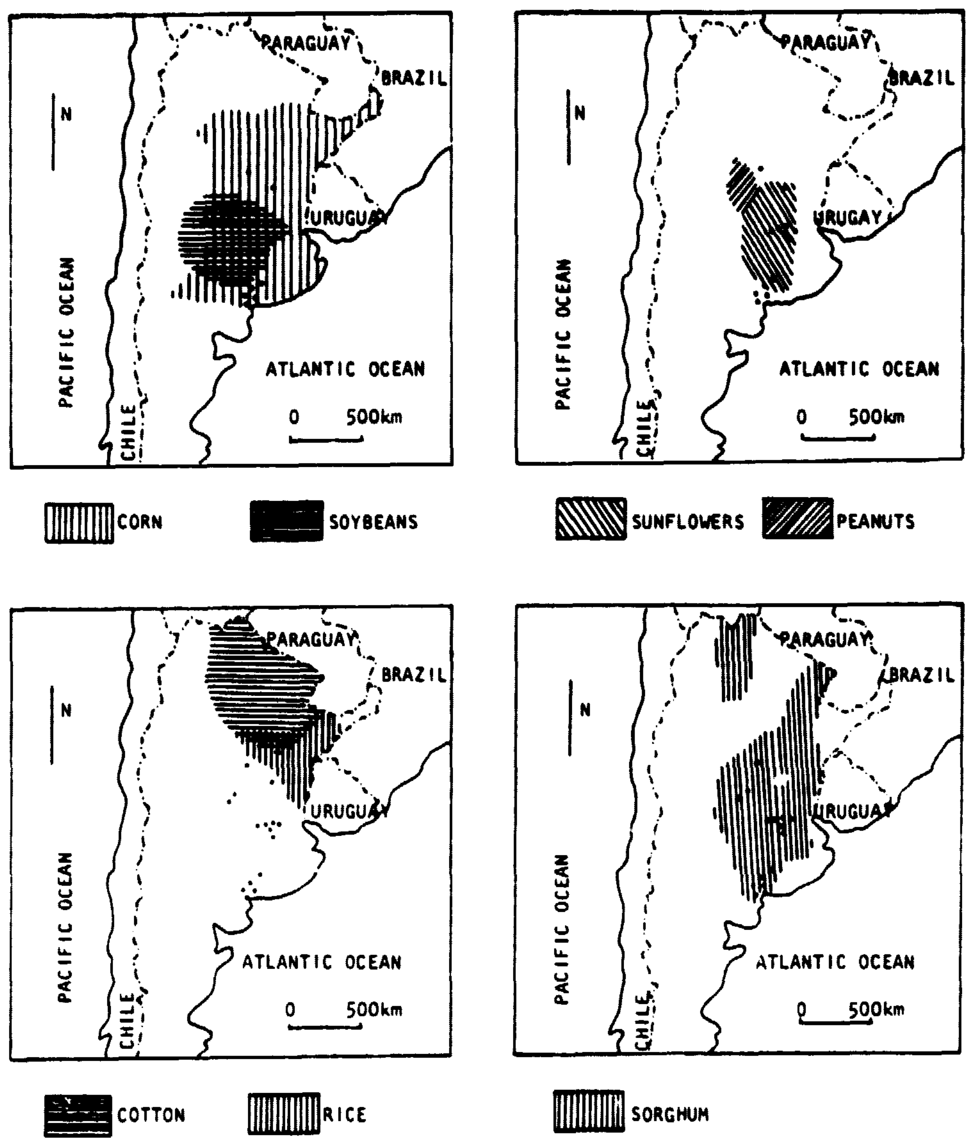


FIGURE 3.11. DISTRIBUTION OF ARGENTINA CONFUSION CROPS

corn and soybeans in the U.S. Corn Belt. The principle feature used is a Landsat green vegetation measure (GRABS) that tracks the growth of the crops. Figure 3.12 illustrates that GRABS values taken throughout the growing season track the early growth and ripening of grains, the lengthy continuously green vegetation in pasture, and the relatively late greening up of summer crops such as corn and soybeans.

Discrimination between corn and soybeans is based on several features. Soybeans are often planted slightly later than corn, and therefore reach their highest GRABS values later than corn. Discrimination is still possible without this temporal difference, however, since soybeans generally have both higher GRABS and Brightness values than corn. Soybeans also often have a greater variability in GRABS and Brightness values (Figure 3.13).

Discussion of Confusion Crops

Intensive study of sorghum spectral characteristics have revealed how closely sorghum parallels corn in both spectral and temporal development (Figure 3.14). Sorghum appears to be slightly later in spectral green-up than corn, and rarely much later. The maximum GRABS are similar, with sorghum occasionally being greener. For any given GRABS value, sorghum tends to have slightly higher brightness values than corn, especially when the corn is irrigated. Occasionally, irrigated corn is greener than the sorghum, although the sorghum remains brighter.

Generally, soybeans achieve higher GRABS values than sorghum, and higher Brightness values when the GRABS values are much higher. When the GRABS values of the two crops are similar, the Brightness values also coincide (Figure 3.14).

Sunflowers are generally greener than corn, less green than soybeans, and brighter than either (Figure 3.15). Temporally, sunflowers are similar to both corn and soybeans, but are much more variable than either. Two types of sunflower spectral patterns have emerged in the

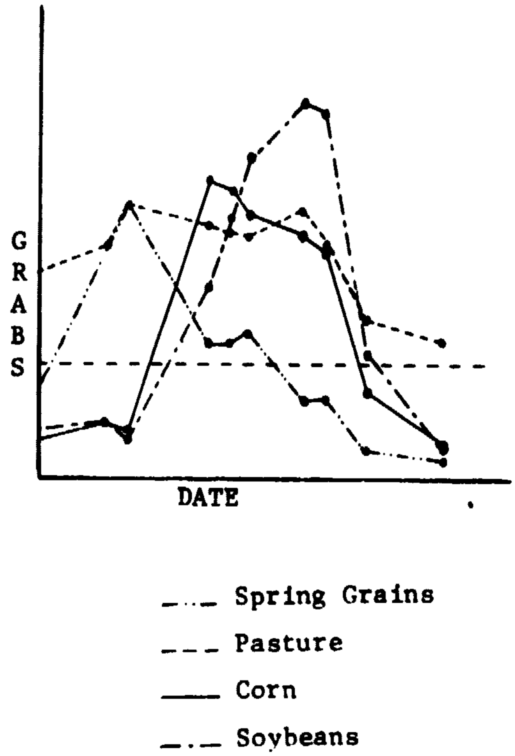


FIGURE 3.12. SEPARATION OF SUMMER CROP FROM OTHER CROPS BASED ON A GREENNESS MEASURE

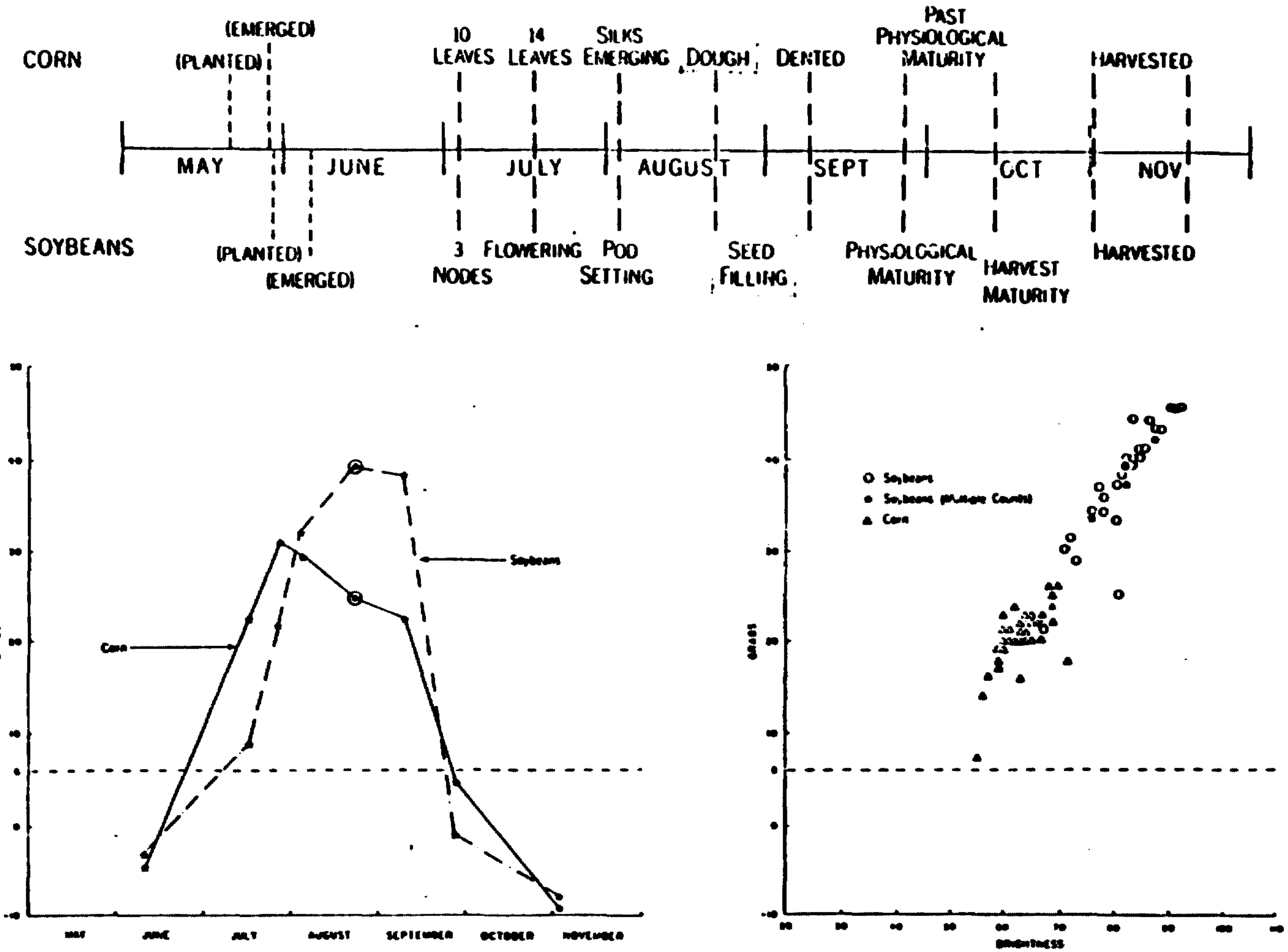


FIGURE 3.13. SEPARATING CORN FROM SOYBEANS BASED ON SPECTRAL AMPLITUDES (MAXIMUM GREENNESS OR GRABS), CENTRAL U.S. CORN BELT

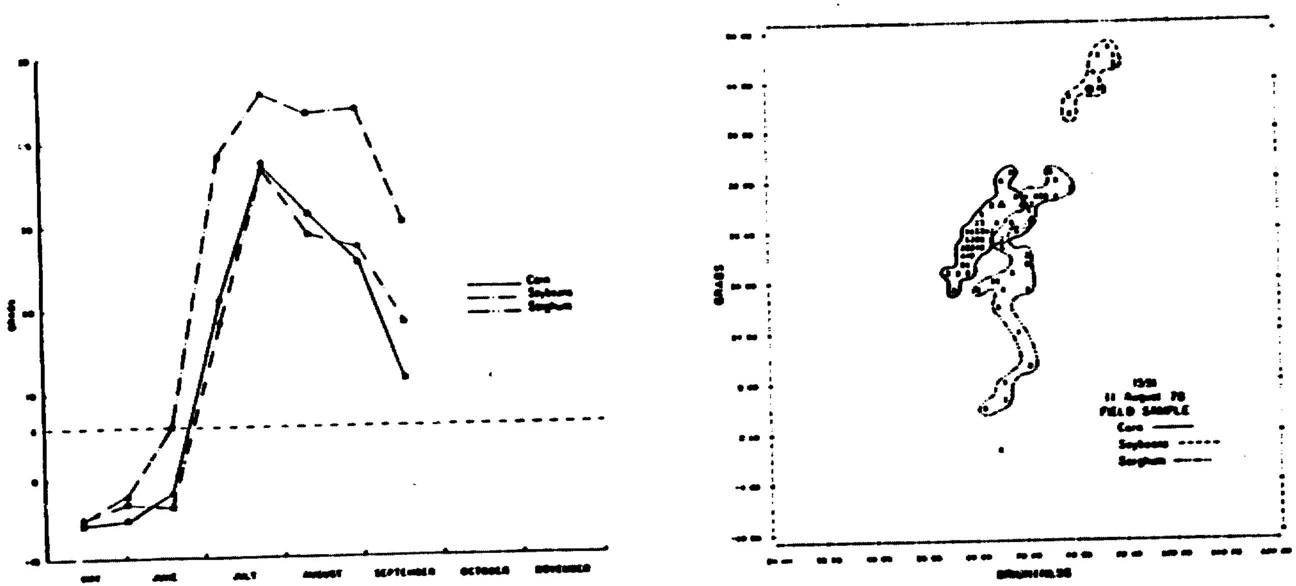


FIGURE 3.14. SORGHUM AS A POTENTIAL CONFUSION CROP RELATIVE TO CORN, SEGMENT 1591, WEBSTER, NEBRASKA

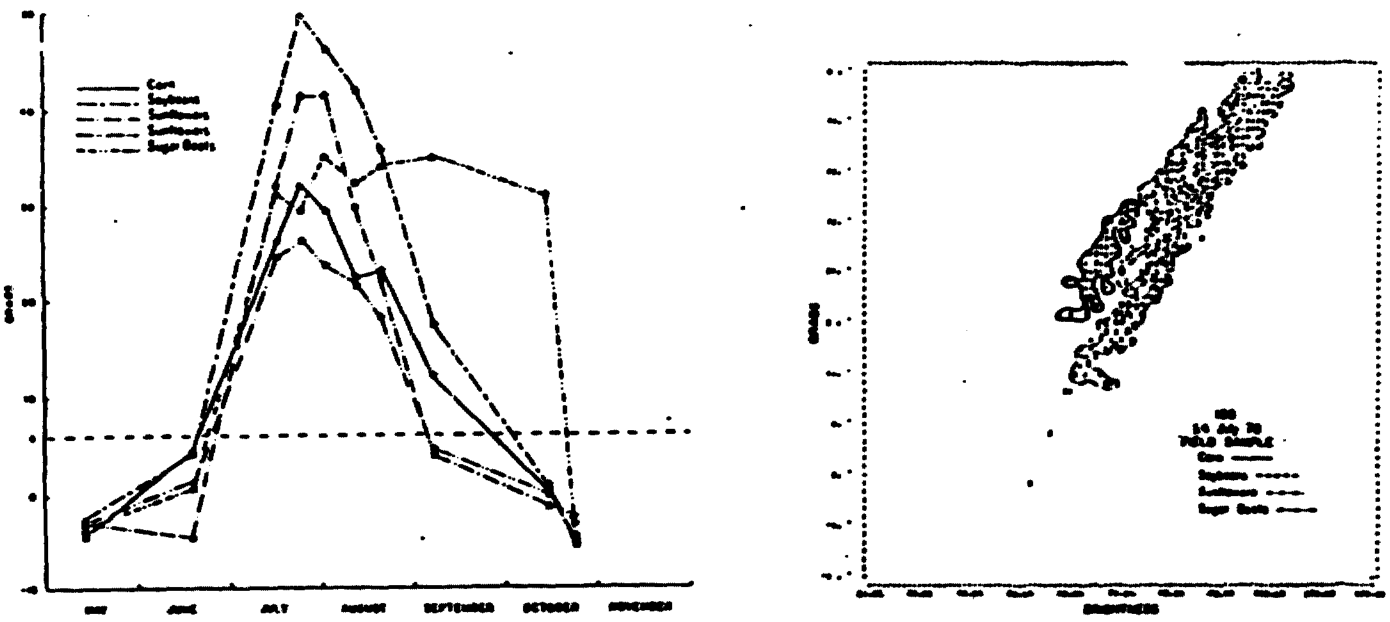


FIGURE 3.15. SUNFLOWERS AS A POTENTIAL CONFUSION CROP RELATIVE TO CORN AND SOYBEANS, SEGMENT 185, TRAVERSE, MINNESOTA

study, referred to as higher-green and lower-green sunflowers. The lower-green sunflowers cause confusion with corn and the higher-green sunflowers cause confusion with soybeans. Maximum spectral separation between sunflowers and corn seems to occur at different times of the growing season from sunflower/soybean spectral separation.

Separation between corn and sunflowers is complicated by the variability of sunflowers. Sunflowers are usually, but not always, greener than corn. They are usually brighter for the same GRABS value, exhibiting a "parallel green arm" effect. This parallel green arm is only visible, however, at certain times in the growing season. Some segments display a spectral progression through the year as follows: (a) sunflowers brighter with GRABS similar; (b) corn and sunflowers similar; (c) sunflowers greener; then (d) corn brighter with GRABS similar at maturity and harvest.

The parallel green arm effect has also been observed at certain times of the year between soybeans and sunflowers, and on plots of maximum GRABS vs. Brightness; with sunflowers tending to be brighter for a given GRABS value (Figure 3.16). Temporally, soybeans tend to develop later than sunflowers. Spectrally, the green canopy of soybeans tends to be of longer duration than that of sunflowers.

The above insights provide a basis for further study into the problem of confusion crops rather than conclusive keys to crop differentiation. Many of these insights are based on distributions visible only in a research environment, and are not as yet useful as analysis tools in a crop inventory procedure.

3.3.2 AREA ESTIMATION USING PROFILE-DERIVED FEATURES

The estimation of corn and soybean acreage from remotely sensed data is a complex process. Considerable effort has been invested in attempting to automate the process as much as possible. While many

ORIGINAL PAGE IS
OF POOR QUALITY

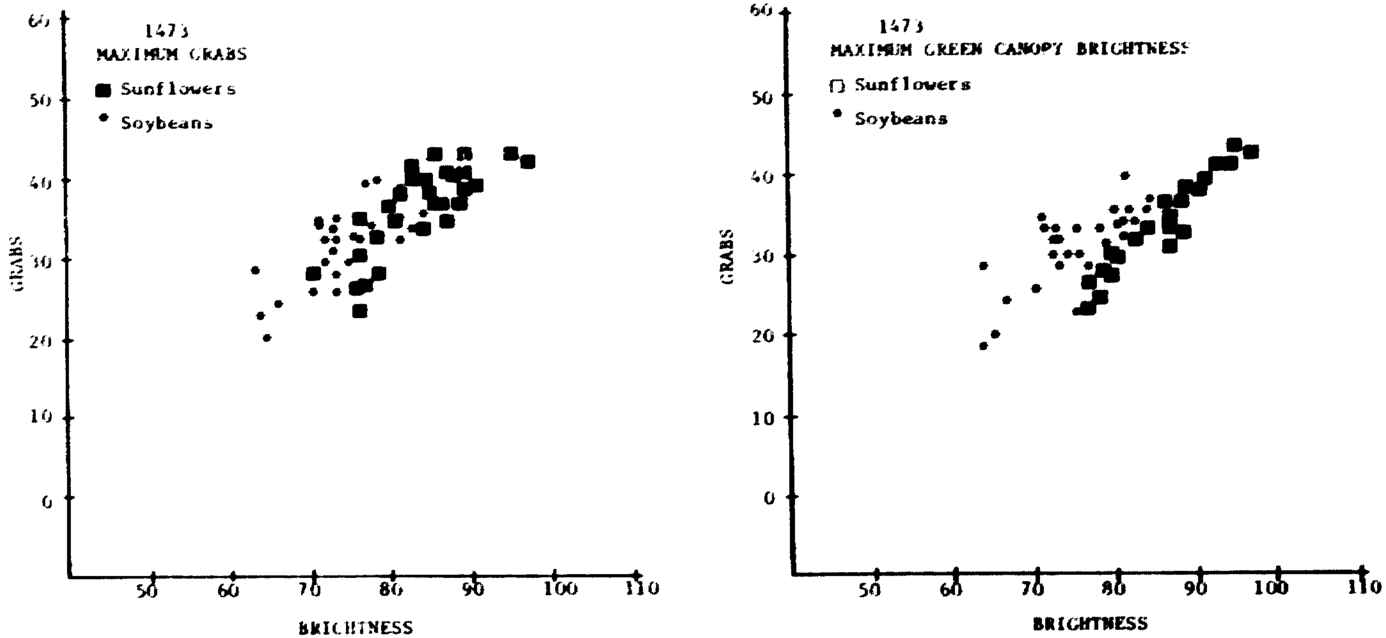


FIGURE 3.16. SUNFLOWERS AND SOYBEANS, SEGMENT 1473
CASS, NORTH DAKOTA

phases have been successfully automated, the critical area of scene classification remains at least partially dependent on the human analyst. In an attempt to minimize the amount of analyst labor, several researchers, notably Dr. G. Badhwar of NASA [9] and Dr. W. Malila and E. Crist of ERIM [45], have developed semi-automatic classification and estimation procedures based on features derived from profile models. This section describes research conducted on such a model form and a preliminary classification/area estimation method based on profile models. It is hoped the method will eventually become an operational procedure requiring minimal analyst interaction or perhaps even be fully automatic.

The classification/estimation method has many conceptual similarities to the Badhwar procedure. Both model summer crop spectral-temporal behavior with a multi-parameter mathematical representation. Both attempt classification and estimation based on parameter values derived from fitting a model profile to data. There are, however, important differences between the two methods as will be seen. Since the Badhwar procedure is well-known, it will serves as a basis of comparison for the method described below. It must be kept in mind, however, that while the Badhwar technique is a complete procedure for area estimation, the method described below is still in the early stages of development.

3.3.2.1 Mathematical Model of Spectral-Temporal Behavior

At the core of the classification/estimation method is an analytical model form of the temporal trajectory of summer crop GRABS (Greenness Above Bare Soil). A GRABS value is a simple linear combination of Tasseled-Cap Greenness and Brightness given by

$$\text{GRABS} = 0.9962 * \text{Greenness} - 0.0872 * \text{Brightness} \qquad (15)$$

The model form is a two-piece sigmoidal profile jointed at the point of peak GRABS; the mathematical representation of the model is:

$$G(t) = \begin{cases} \dfrac{A}{1 + Q1^2 \cdot (t - DP)^2}, & t \leq DP \qquad (16a) \\ \\ \dfrac{A}{1 + Q2^2 \cdot (t - DP)^2}, & t > DP \qquad (16b) \end{cases}$$

G(t) = GRABS value at time t
DP = day of peak GRABS
A = peak GRABS value, i.e., G(DP) = A
Q1 = emergence to peak "green-up" rate parameter
Q2 = peak to harvest "green-down" rate parameter

Interpretation of Model Parameters

Figure 3.17 provides a graphical interpretation of the model parameters. As can be seen, the reciprocals of the rate parameters, Q1 and Q2, define the time intervals between peak GRABS and the half-peak point on each side. Thus, larger values of Q1 or Q2 correspond to increased rates of change of GRABS values, i.e., steeper slopes in the profile shape. The remaining two parameters of model form, the peak GRABS value and the day of peak are self-explanatory.

Comparing the four parameters of Equation 16 to the parameters of the Badhwar model reveals many similarities in the types of information provided by each. The Badhwar procedure fits a one-piece three parameter model to Tasseled-Cap Greenness vs. Time, and a quadratic fit to the ratio of Greenness to Brightness vs. Time. The first fit yields the parameters α, β and t_0, while the second produces the parameter σ. α and β describe the rates of "green-up" and "green-down", respectively, and so are analogous to Q1 and Q2. (Note from Equation 15 that GRABS

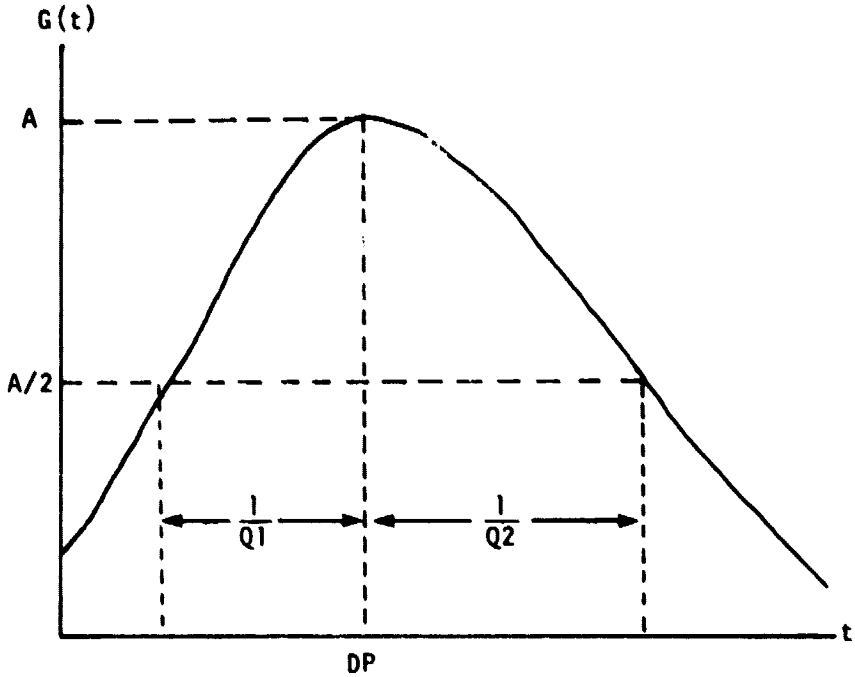


FIGURE 3.17. GRAPHICAL INTERPRETATION OF MODEL PARAMETERS

and Greenness are nearly identical quantities.) The parameter t_0 is the time of spectral emergence. Given t_0, α and β, one can calculate the time of peak Greenness and the actual peak value of the one-piece profile. The parameter σ from the quadratic fit is essentially a measure of the "width" or duration of the Greenness profile. As Figure 3.17 suggests, the same type of information is available from an appropriate combination of Q1 and Q2. Defining a quantity, SPAN, as the measure of profile width, we see that it is given by:

$$\text{SPAN} = \text{measure of profile width} = 1/Q1 + 1/Q2 \tag{17}$$

3.3.2.2 Parameter Estimation Procedure

Having established a model profile, the next step is to estimate profile parameters for various crops and crop types by fitting the model to actual data. Subsequent sections describe the results of profile fitting to 11 corn/soybean segments. The present section provides a brief discussion of the method used to fit the two-piece profile to data.

The method of fitting Equation 16 to spectral-temporal data is embodied in the program STEPFIT. The program name is descriptive of the method employed to estimate the four-parameters DP, A, Q1 and Q2. The program steps through a series of DP values, estimating the remaining three parameters at each value, in a search for the day of peak that best fits (in a least squares sense) the data.

To explain this process in greater detail, we will use the variable names appearing in STEPFIT. Equation 16 is fit to the data over an interval of DP values. The interval is defined as NDP days on each side of some center value DPØ. Thus, there are (2*NDP+1) days in the entire interval. STEPFIT uses the day of the maximum data value as the value of DPØ. The program therefore initially expects to find the "true" day of peak Greenness within NDP days of the maximum data point.

With the initial interval defined, STEPFIT sets DP equal to the first value in the interval, i.e., DPØ-NDP, calls ZXSSQ (a standard IMSL non-linear regression routine) to fit the model using that day as the peak. ZXSSQ returns, among other things, SSQ, the residual sum of squares for the final parameter estimates. This quantity is stored as a function of the corresponding value of DP. The value of DP is incremented by DPINC, usually one day, and ZXSSQ is called again with the new DP value. This process continues throughout the interval. The result is a series of SSQ values as a function of the values of DP. The value of DP with the minimum corresponding SSQ is taken as the "true" day of peak. Since the other model parameter estimates, i.e., A, Q1 and Q2, are saved with each value of DP, once the "true" day of peak is found, the optimum profile fit is already known. Figure 3.18 illustrates the above process graphically.

3.3.2.3 Profile Fitting Experiment

As mentioned previously, Equation 16 was fit to data in 11 corn/soybean segments to assess the model's usefulness for scene classification -- specifically, its ability to model corn and soybean spectral behavior. An analysis was made to determine if the profile parameters could be used to discriminate between summer crops and "other" scene features, and within the summer crop category, between corn and soybeans.

Data Base

Eleven segments located in the central Corn Belt were used in the experiment.* Each segment was processed to define quasi-fields, or "blobs". Ground truth data was available for all the blobs generated in each segment. The spectral means of the blobs were transformed into GRABS values and input to STEPFIT. The resulting profile parameters are thus characteristic of corn, soybean and other quasi-

*Segments 123, 141, 202, 205, 800, 832, 842, 852, 853, 877, 881.

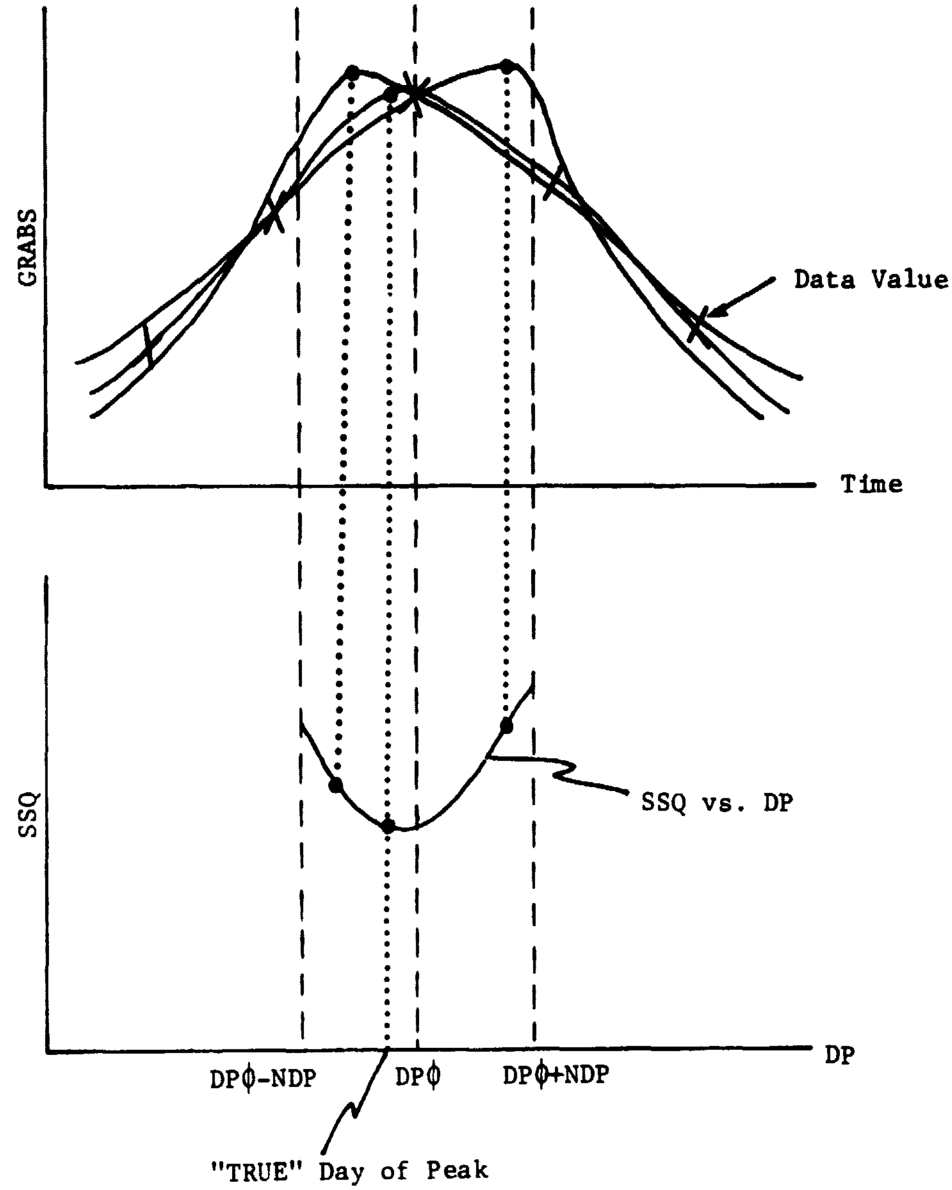


FIGURE 3.18. ILLUSTRATION OF STEPFIT METHOD OF PROFILE MODEL PARAMETER SPECIFICATION

fields. This is in contrast to the Badhwar procedure in which profile parameters are computed for individual pixels.

Two types of blobs were identified: those containing interior pixels as well as blob boundary pixels (big blobs) and those consisting of only boundary pixels (little blobs). The interior pixels of big blobs are considered to be spectrally pure, i.e., free of misregistration effects. Big and little blobs were subdivided further into those blobs whose ground truth classification exceeded 5/6 in any crop class (i.e., at least 5/6 of the blob's pixels had the same ground truth classification) and those that didn't. For the big blobs, this distinction was made by considering the ground truth classification of interior pixels only. Similarly, the spectral means of big blobs were computed solely from interior pixel spectral values.

The data base thus contained four levels of "signature purity". The first level, represented by big blobs with greater than 5/6 ground truth purity, consists of signatures contaminated by neither crop mixtures nor misregistration. The second, little blobs with greater than 5/6 ground truth purity, consists of signatures which are potentially impure due to misregistration. The third, big blobs with less than 5/6 ground truth purity, contains signatures which are impure due to crop mixtures but not misregistration. The fourth level, represented by little blobs with less than 5/6 ground truth purity, contains signatures which are impure due to both crop mixtures and potential misregistration.

Profile Fitting Results

After computing profile fits to all blobs in the 11 segments, an analysis was made to determine the efficiency of profile fitting in the four signature purity levels. "Efficiency", in this context, refers to the number of blobs that were accurately fit by Equation 16.

Accuracy or goodness-of-fit (G-O-F) is quantifiable in a number of ways. The program STEPFIT computes for each blob the following measure of goodness-of-fit,

$$\text{G-O-F} = 1 - \frac{\sum [PV(t) - D(t)]^2}{\sum [D(t) - \bar{D}]^2} \tag{18}$$

where

$PV(t)$ = computed profile value at time t

$D(t)$ = actual data value at time t

$\bar{D}$ = mean value of data values $D(t)$

and the summations are over the number of data points (acquisitions). From Equation 18 we see that G-O-F can have a maximum value of 1.00, corresponding to perfect fit, while the minimum value is theoretically unbounded.

G-O-F = 0.75 was arbitrarily chosen as the boundary between two classes of blobs: well-fit blobs (i.e., $0.75 \leq$ G-O-F - ≤ 1.00) and poorly-fit blobs (G-O-F $<$ 0.75). In addition, there exists a third class, those blobs not fit at all. This situation occurs when ZXSSQ, the non-linear regression routine used in STEPFIT, is unable to converge upon the set of profile parameters which best fit the data. This may occur for a number of reasons, but the most common is simply the inability of Equation 16 to adapt to certain spectral-temporal trajectories. This characteristic can be exploited to advantage as we shall see.

Table 3.6 summarizes the profile fitting efficiencies observed for the four classes of signature purity. In Table 3.6 , "pure" denotes greater than 5/6 ground truth purity and "impure" indicates less than that.

Tables 3.7 and 3.8 further subdivide the pure blobs into four components: corn, soybeans, vegetated non-agricultural (e.g., pasture) and unvegetated non-agricultural. These four classes comprise more

TABLE 3.6. OVERALL SUMMARY OF PROFILE FITTING EFFICIENCY

Class	No. of Blobs	% Not Fit	% Poorly Fit	% Well Fit
Pure Big Blobs	3581	17.0	12.1	70.9
Pure Little Blobs	4643	18.0	23.3	58.7
Impure Big Blobs	1459	13.1	17.2	69.7
Impure Little Blobs	4701	14.2	22.7	63.1

TABLE 3.7. BREAKDOWN OF PURE BIG BLOBS

Class	% of Blobs	% Not Fit	% Poorly Fit	% Well Fit
Corn	1134	2.9	8.2	88.9
Soy	1334	1.9	5.3	92.8
Non-Agricultural (Vegetated)	943	53.7	22.6	23.8
Non-Agricultural (Unvegetated)	170	27.1	33.5	39.4

TABLE 3.8. BREAKDOWN OF PURE LITTLE BLOBS

Class	% of Blobs	% Not Fit	% Poorly Fit	% Well Fit
Corn	808	8.8	18.6	7.26
Soy	1745	5.9	16.0	78.1
Non-Agricultural (Vegetated)	1229	31.5	29.2	39.3
Non-Agricultural (Unvegetated)	861	31.7	34.1	34.1

than 90% of all pure blobs in the 11 segments. (The remaining less than 10% were blobs for which ground truth was unknown or unavailable.)

Table 3.6 shows only small differences between the four levels of signature purity. As might be expected little blobs are fit well less often than are big blobs, however, blob purity has only a small effect on whether or not a blob is fit well. Indeed, pure blobs appear more likely to be not fit at all compared to impure blobs. This effect can be explained by considering Tables 3.7 and 3.8. When pure blobs are resolved into their four component classes, it is seen that the vast majority (80-90%) of those not fit fall into the non-agricultural category, especially vegetated non-agricultural. For example, Table 3.6 shows that 17%, or 610, of the 3581 pure big blobs were not fit. Table 3.7 shows that 53.7%, or 506, of the 943 pure big vegetated non-agricultural blobs were not fit. Thus 506 of the 610 pure big blobs not fit were vegetated non-agricultural. An additional 46 were unvegetated non-agricultural.

The reason a smaller percentage of impure big blobs were not fit may also be explained. In the 11 segments, most of the impure big blobs were mixtures of summer crops with other that was spectrally similar to summer crops. The spectral-temporal pattern of the mixture blob was therefore "summer-crop-like" in appearance. Such a blob is more likely to be fit by Equation 16 than is a blob with a purely non-summer crop appearance This is evidenced in Table 3.6 for both little and big blobs.

As seen in Tables 3.7 and 3.8, only a small fraction of pure corn and soy blobs were not fit, while a significant number of pure non-agricultural blobs were not fit. Indeed, in the 11 segments analyzed, if a pure big blob was not fit its probability of being non-summer crop was over 90%. This suggests that a reliable first order separation of summer and non-summer blobs is possible using only a single profile parameter (G-O-F) and a simple binary decision (fit or not fit). To achieve more refined Summer/Other separation of discrimination between corn and soy requires the Other profile parameters as discussed in the following section.

3.3.2.4 Classification Feature Space

A six-dimensional feature space spanned by G-O-F, SPAN (defined in Equation 17), and the four parameters of Equation 16 was analyzed to determine the potential separability of Corn, Soybean, and Other (vegetated and unvegetated non-agricultural). Only pure big blobs were considered in the analysis to ensure relative signature purity. The use of pure big blobs is analogous to the use of "pure" pixels to train an automatic classifier in the Badhwar procedure. In that procedure, "pure" pixels - those identified as being within field interiors and considered by an analyst to be pure Corn, Soy or Other - are profile fit. The resulting parameter values are used to adjust classification boundaries which are applied to the remaining pixels in the scene. Such adjustments allow the procedure some adaptability to the growing conditions in a particular region.

The analysis of the six-dimensional space used pure big blobs to define the parameter values characteristic of Corn, Soy and Other. Ideally, each class would occupy a distinct region in the feature space allowing for deterministic classification. However, in practice, this was not the case. The parameter distributions of the three classes tended to overlap to some degree. Typically, the distribution for Corn fell between Soy and Other and was overlapped by each.

Figure 3.19 is a semi-quantitative presentation of the relationships between Corn, Soy and Other in each of the feature space's six dimensions. The positions of each class are intended to correspond to the medians of their respective distributions, although the scales of each parameter are arbitrary. The figure is representative of pure

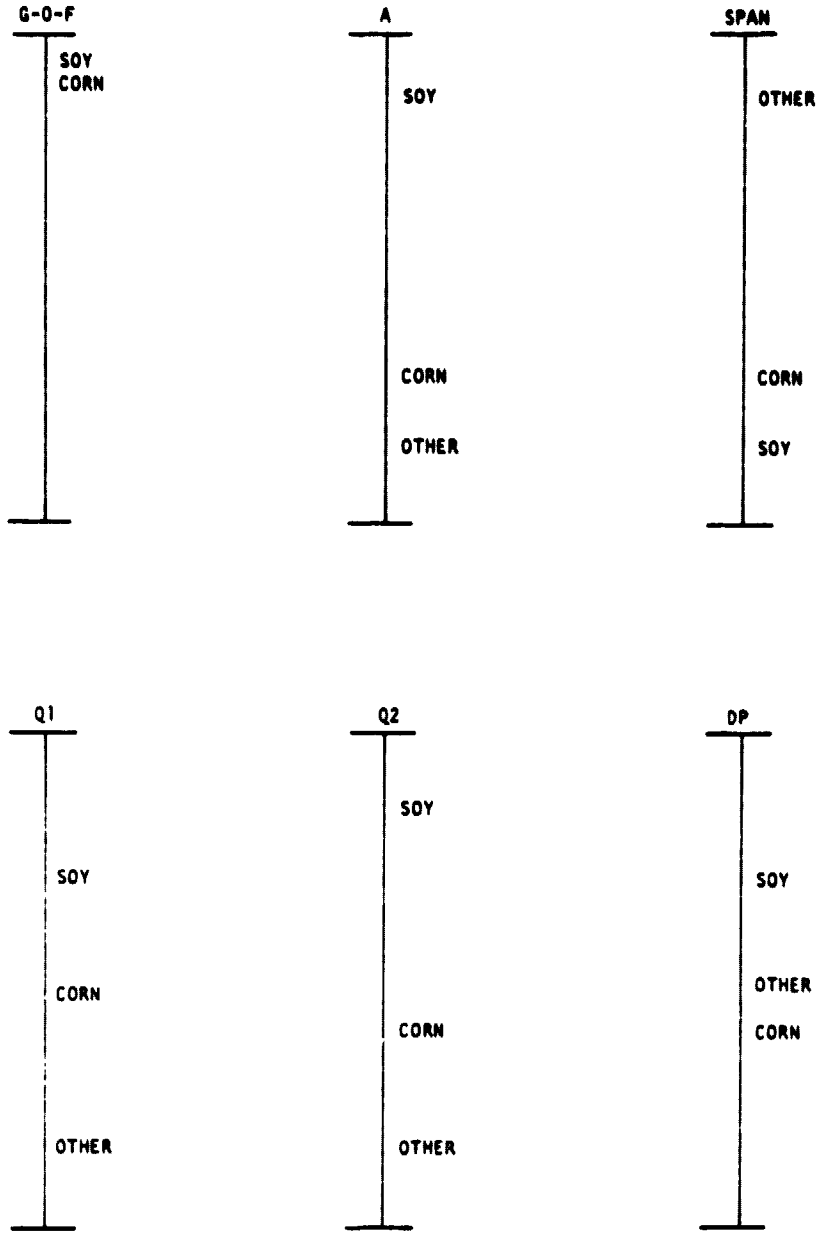


FIGURE 3.19. GENERAL RELATIONSHIPS OF CORN/SOY/OTHER IN FEATURE SPACE

big blobs that were fit. As can be seen, Other is most distinct from Corn and Soy along the dimensions G-O-F and SPAN, while Corn is most separable from Soy along the A dimension.

Figures 3.20 and 3.21 show the actual distributions observed over all 11 segments for two of the parameters, A and SPAN. Again, the distributions are for pure big blobs that were fit. Figure 3.20 shows that although Corn and Soy have relatively distinct distributions of peak GRADS, the Corn and Other distributions are completely overlapping. This illustrates the major obstacle encountered in attempting classification based on parameter values - namely, the separation of Corn from well-fit Other.

A partial solution to this problem is suggested by Figure 3.21, the distributions in the parameter SPAN. Other blobs tend to have larger SPAN values than either Corn or Soy. There is still substantial overlap between Corn and Other, but this is not as serious as it would appear for the following two reasons. The first is that the entire Other distribution of SPAN is not shown in Figure 3.21(c). Over 25% of the pure big Other blobs fit had SPAN values in excess of 250. Thus, the portion of the Other distribution overlapping the Corn distribution is less significant than it appears. The second reason is that the Other blobs making up the overlapping portion (i.e., SPAN 150) tend to have low values of G-O-F (median value = 0.50), and so could be separated from Corn based on that parameter.

Once Other is separated from summer crops, Corn and Soy pure big blobs are distinguishable using only a few parameters. Figures 3.20 and 3.21 suggest that they are fairly distinct in a plane spanned by A and SPAN. This is indeed the case as shown in Figure 3.22 where the central portion of each distribution has been outlined.

It should be emphasized that the distributions shown in Figures 3.20, 3.21 and 3.22 are composed of data from all 11 segments. The 11 segments represent a variety of growing conditions and planting dates; at least one segment contained stressed soy. The potential

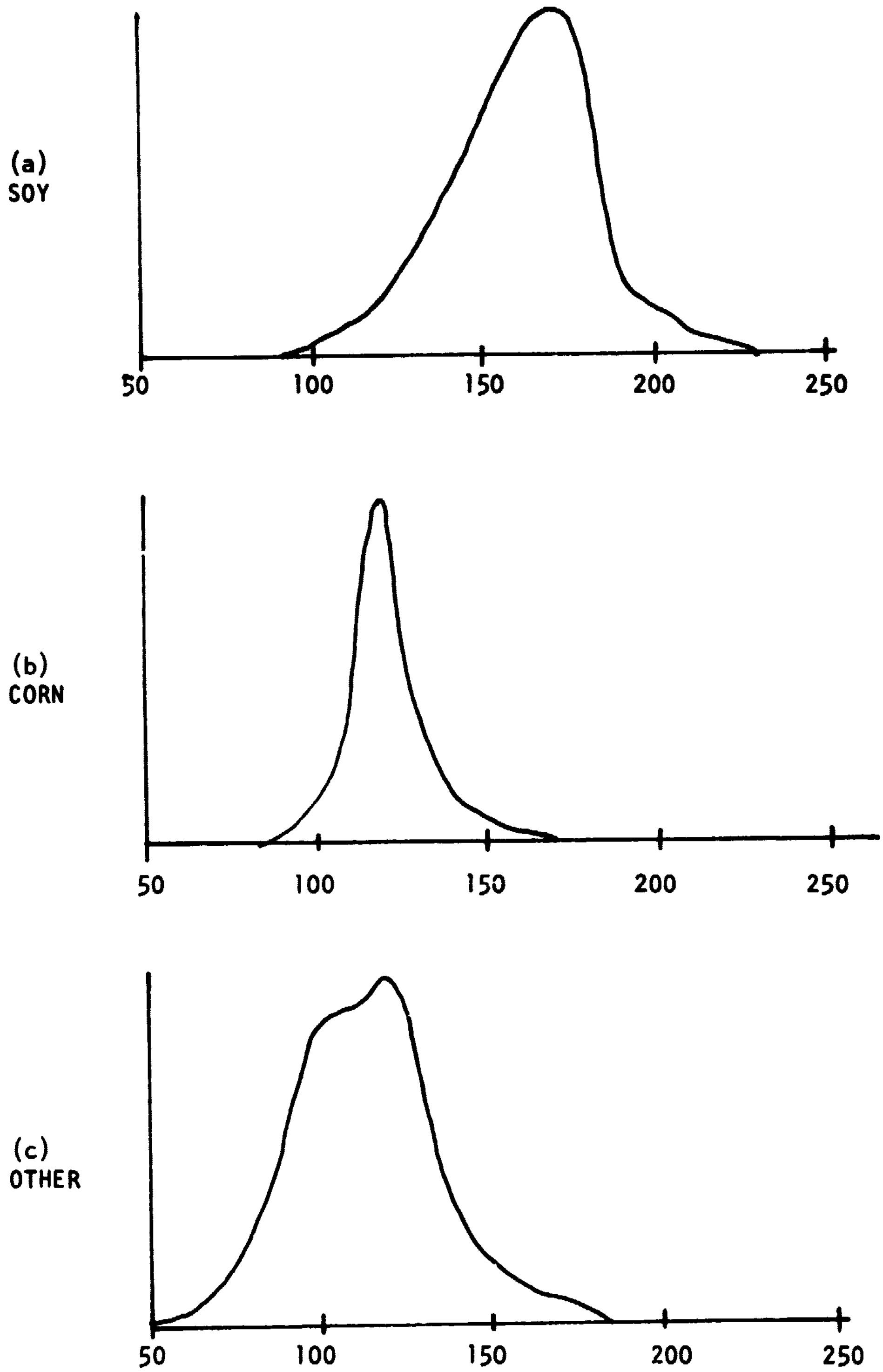


FIGURE 3.20. DISTRIBUTIONS OF A, PEAK GRABS VALUE

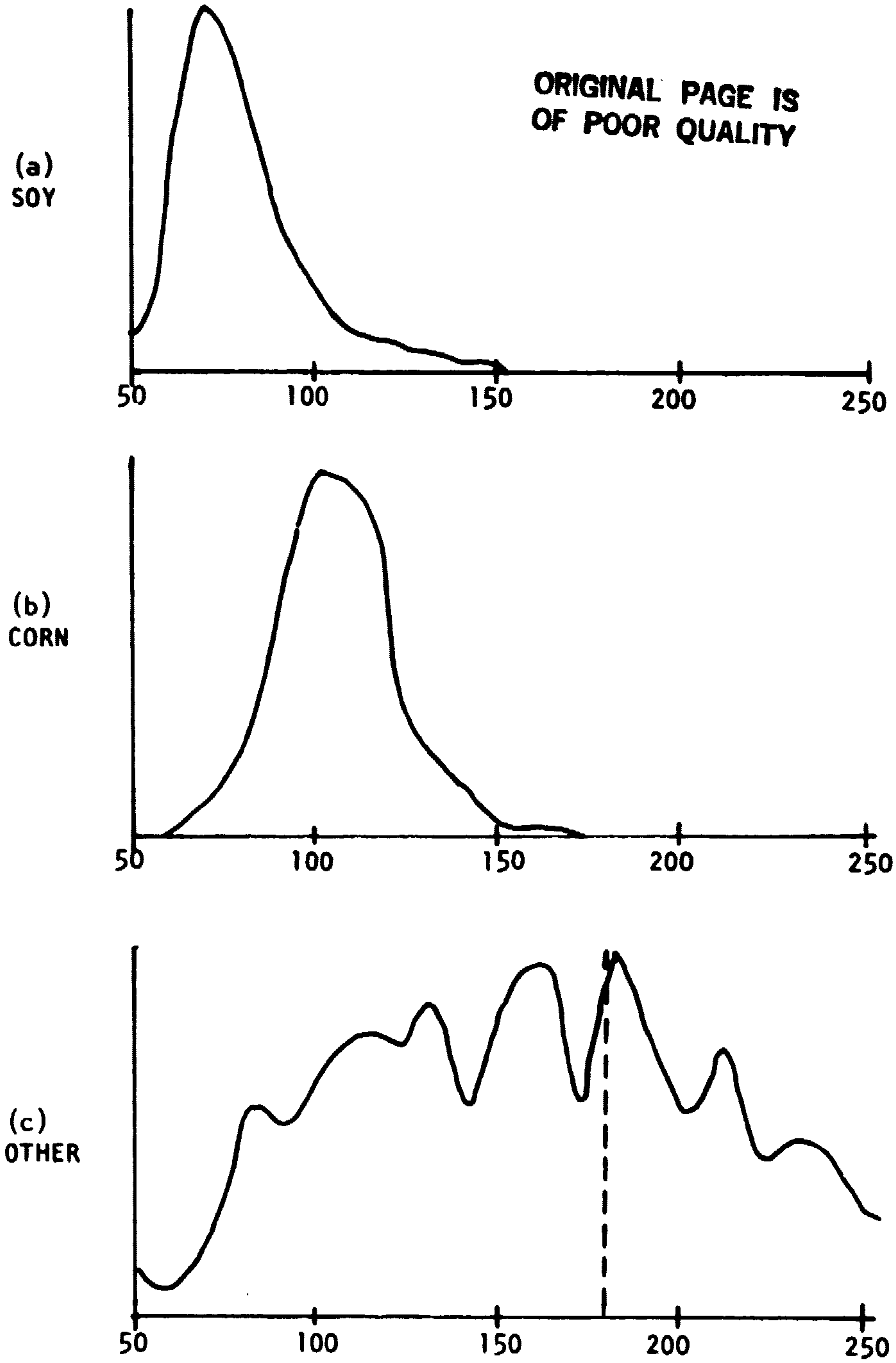


FIGURE 3.21. DISTRIBUTIONS OF SPAN, PROFILE DURATION MEASURE

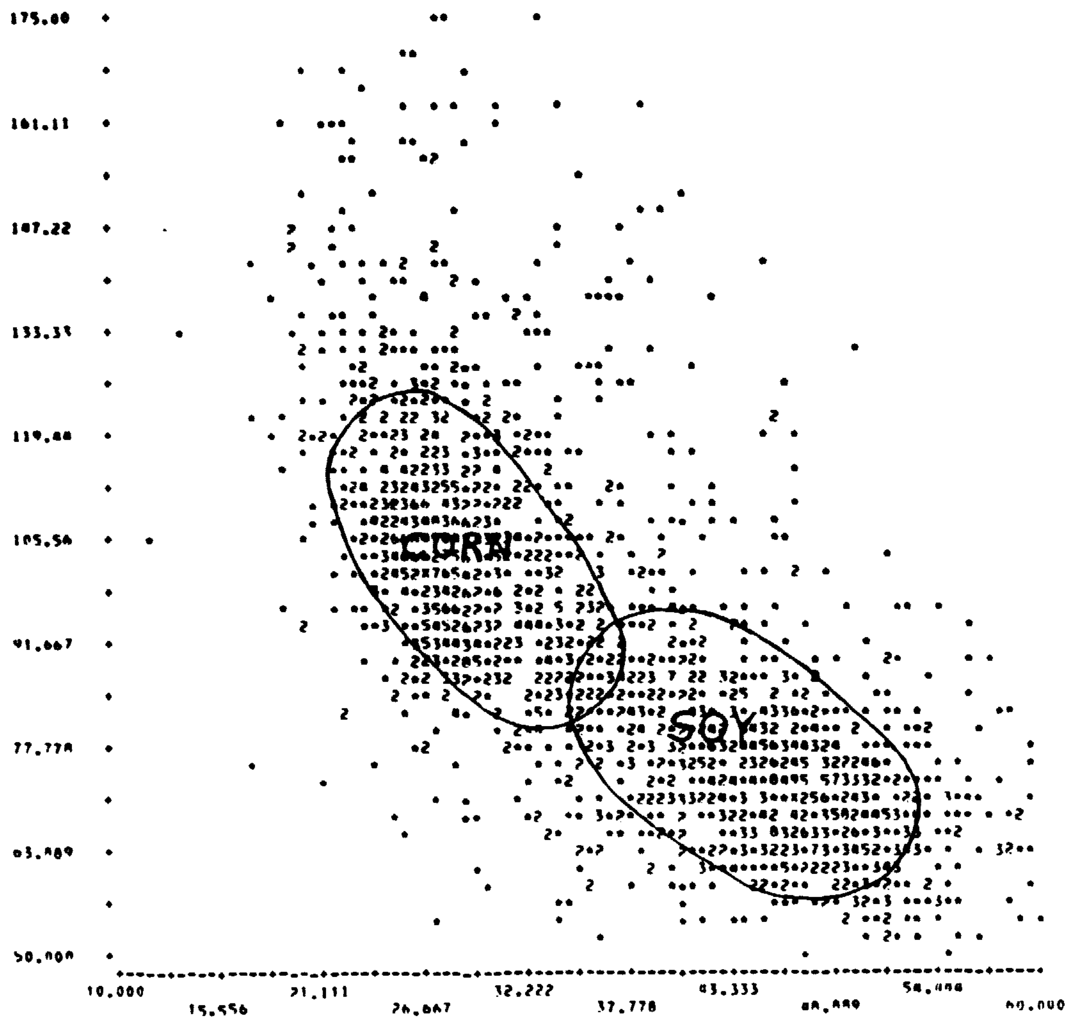


FIGURE 3.22. SPAN VS. A (PEAK GRABS VALUE)

separability of Corn, Soy and Other illustrated in these figures might be improved on a segment by segment basis. In other words, adjusting the classification decision boundaries according to the partiuclar conditions of a segment, as in the Badhwar procedure, might improve classification accuracy. However, the distributions illustrated in Figures 3.20, 3.21, and 3.22 suggest that fixed decision boundaries in the feature space could be used successfully. If this proves to be true, a fully automatic classification procedure becomes a viable concept.

3.3.2.5 Preliminary Crop Classification Experiment

A preliminary strategy for classifying pure big blobs was formulated and tested in an experiment. The basic approach used was as follows. All blobs not fit were classified as Other. This follows from the observation that over 90% of the blobs not fit were Other. The remaining blobs were separated into Summer Crop and Other based on a Stage 1 discrimination. The Summer Crop group was then resolved into Corn and Soy based on a Stage 2 discrimination. The number of pixels allocated to each class was totaled and converted into a percentage. The results were compared with the known ground truth percentages of each class.

The experiment was conducted on a segment by segment basis. The Stage 1 and Stage 2 discriminations were accomplished by applying a segment specific optimum linear discriminant to the data. The linear discriminant was calculated based on segment specific parameter distributions of the well-fit (G-O-F 0.75) pure big blobs. Thus, to a large extent, the discriminant separated the same data distributions it was "trained" upon. The key objectives of the experiment were to assess the blanket classification of not-fit blobs as Other, and the classification of poorly-fit blobs based on the parameter values of well-fit blobs.

Table 3.9 shows the results of the experiment for each of the 11 segments. Both the estimated and true percentages apply to pure big blobs only.

In all but Segment 877, the estimated percentage agrees fairly well with the true percentages. In Segment 877, for an as of yet unexplained reason, a large percentage of fit Corn was classified as Other. The error is evenly split between well-fit and poorly-fit Corn. Most of the remaining segments show a slight bias toward Other. This is to be expected due to the few not-fit Corn and Soy blobs being classified as Other. A compensating adjustment of the linear discriminant - i.e., one that biases the Stage 1 classification toward Summer Crop - could probably eliminate this bias. No definite trend was noted with respect to poorly-fit blobs. They tended to be misclassified and classified correctly with nearly equal probability, although poorly-fit Other was generally recognized as Other.

3.3.2.6 Deriving Area Estimates from Feature Space Classification

Given that crop classification based on profile parameters is possible, the next step is to generate an area estimate based on those classifications. There are several possible approaches to this problem. One would be to simply fit all blobs, classify them based on their profile parameters, and aggregate the number of pixels allocated to each class. However, this approach ignores the errors likely to arise from applying decision boundaries derived from pure big blobs to blobs which are little and/or impure. A second approach might be to classify big and little blobs independently using separate decision boundaries for each. Unfortunately, it was observed that the parameter distributions for pure little Corn, Soy and Other blobs tended to cluster together compared to pure big blobs. This makes the accurate classification of little blobs a more difficult task. A third approach might classify only big blobs, generate an area estimate for them, and then somehow extend that estimate to the little blobs, as in the C/S-1A.

TABLE 3.9. RESULTS OF CLASSIFICATION EXPERIMENT

Class	Segment 123 Estimated %	True %	Class	Segment 141 Estimated %	True %
Corn	41.5	41.4	Corn	26.7	26.7
Soy	36.2	37.8	Soy	20.5	18.6
Other	22.3	20.9	Other	52.8	54.6
	Segment 202			Segment 205	
Corn	22.9	25.7	Corn	20.7	17.2
Soy	34.3	37.0	Soy	62.8	64.0
Other	42.8	37.3	Other	16.5	18.8
	Segment 800			Segment 832	
Corn	63.5	64.9	Corn	19.8	20.2
Soy	26.4	26.0	Soy	54.8	57.8
Other	10.1	9.1	Other	25.4	22.0
	Segment 842			Segment 852	
Corn	49.0	51.3	Corn	36.4	37.3
Soy	34.8	34.8	Soy	27.2	29.9
Other	16.1	13.9	Other	36.3	32.7
	Segment 853			Segment 877	
Corn	48.9	50.8	Corn	28.2	55.2
Soy	29.6	30.8	Soy	23.5	25.9
Other	21.5	18.4	Other	48.4	19.0
	Segment 881				
Corn	47.1	47.8			
Soy	6.1	6.8			
Other	46.8	45.4			

None of the approaches outlined above adequately addresses the problem of impure or mixture blobs. While this problem is certainly not unique to profile based procedures, it is one of the most formidable obstacles to a fully automatic area estimation procedure.

There are then several areas of research in which effort is required before a complete area estimation procedure can be developed from the feature space classifications. One is determining the classification accuracies possible with little blobs. Another involves a study of mixture blobs to see if they exhibit any characteristic behavior in the feature space that would identify them as being impure. Yet another is a complete assessment of the use of fixed decision boundaries in the parameter space.

3.3.2.7 Summary and Conclusions

A summer crop spectral-temporal profile model and profile fitting procedure has been developed which accurately fits summer crop behavior and discriminates against (does not fit) non-summer crop behavior. A six-dimensional feature space based on the profile parameters was analyzed and was found to have potential for the automatic or semi-automatic classification of Corn, Soy and Other. With further research, it is felt that an automatic or semi-automatic classification/area estimation procedure could be developed from the profile techniques described above. Such a procedure would operate as an end-of-season technique and would require four well-timed acquisitions as a minimum.

3.3.3 ESTIMATING ACREAGE BY DOUBLE SAMPLING

3.3.3.1 Introduction

In crop inventory application, as in many forms of survey sampling, there may be two, nominally competing, techniques of measurement available, each with its associated per sample variance, bias, and cost. If it is necessary to choose one or the other technique, and if the techniques both have an acceptably small bias, the answer is well known: Choose the technique with smaller cost-variance product.

More often it is not necessary to choose strictly among measurement techniques. Rather, it is possible to make some of both kinds of measurements and mix the results to obtain an overall lower variance at the same total cost, even when one of the techniques, when used alone, has an unacceptable bias. Consider a low cost, biased, high variance technique and a high cost, (nearly) unbiased low variance technique whose results on the same samples are well correlated. We can view the high cost technique as a method of calibration of the low cost technique. The calibration is performed by double sampling wherein the bulk of the samples will be measured inexpensively, and a certain subset of samples are measured by both techniques. The entire set of measurements is then used to make a regression estimate which is unbiased with respect to the more expensive measurement technique and lower variance (than either technique used separately) for a given total cost. The conditions for which this is true are again given by Cochran [18]. The answer (the number of double and single samples allocated) is obtained by minimizing the variance of the estimator subject to a fixed total cost. Such situations are most likely to arise in practice if the competing techniques in question share some substantial portion of their overhead costs in common, e.g., if the more expensive technique is a more extensive or thorough application of the lower cost technique.

The USDA's Domestic Crop/Land Cover Project utilizes double sampling techniques to adjust a Landsat-based estimate over a large region by the use of an estimated regression relationship between the Landsat-based and ground survey-based estimates over a subset of the region.

The application discussed in this section centers around several Landsat-based techniques for estimating crop acreages, namely: a fictional perfect procedure, a relatively expensive analyst-intensive use of Landsat data, and a less expensive but closely related method of using Landsat data. However, the application studied in this report is of more general interest than described above in two significant ways:

a) The quantity to be estimated is multivariate, i.e., the acreages of two or more crops (in particular, corn and soybeans) simultaneously.

b) The cost constraints are more general, consisting of limitations on two or more types of resources (analysts and computers) as well as total cost.

In this more general situation one must define a suitable objective function to minimize (replacing the variance) subject to the (more elaborate) constraint set.

In the next section we describe briefly the double sampling solution algorithm and in the section following we present applications of the technique to hypothetical constraint sets.

3.3.3.2 Description of the Double Sampling Approach

The solution algorithm for the double sampling optimization problem is most completely described in Pont, Horwitz & Kauth [53], and a synopsis is given in the paragraphs below.

First, an initial determination is made as to whether double sampling would be beneficial. From [18] double sampling would be used if:

$$\frac{c}{c'} > \frac{\rho^2}{(1 - \sqrt{1 - \rho^2})^2}$$

where

c = cost of more expensive technique;

c' = cost of less expensive technique;

p = correlation (multiple correlation) between results of two estimation techniques

The greater the cost ratio, or the greater the correlation of answers from the two techniques, the more valuable double sampling becomes.

Second, once double sampling is found useful the optimum sample allocation is determined. This requires that a suitable object function be found:

$$F(n, n')$$

where

n is the number of samples allocated to the more expensive technique;

n' is the number of samples allocated to the less expensive technique.

This function could be the variance of one crop estimate, or a combination of the variances of several crop estimates. Then the problem may be formulated as finding (n, n') that minimizes F subject to a constraint set:

$$A \begin{bmatrix} n \\ n' \end{bmatrix} \leq b \quad \text{(where A is a matrix and b a vector)}$$

$$n \leq n'$$

n and n' must be positive integers

This is a nonlinear integer programming problem with linear constraints. The solution method used depends on F decreasing with n and n'. For each possible value of n, the largest possible n' within the constraints is determined, and F computed. The value of n minimizing F determines the solution point.

Finally, once the sample is taken and procedure results tabulated, the overall estimate is determined as follows. We denote the results obtained for the inexpensive procedure samples as

$$\{x_i\}_{i=1}^{n'}$$

and for the expensive procedure samples as

$$\{y_k\}_{k=1}^{n}$$

The linear relationships

$$(y - \mu_y) = B(x - \mu_x) + \varepsilon$$

where

$$\varepsilon = \text{random variable with mean } 0$$

is assumed to associate the two types of results using y_k and the subset of $\{x_i\}$ that cover the samples, the overall final estimate is

$$y_k = \bar{y} + b(\bar{x}' - \bar{x})$$

where

- $\bar{y}$ is mean of y_k
- $\bar{x}$ is mean of those x_i that cover the same samples as the y_k (n of the n' values of x)
- $\bar{x}'$ is the mean of all x_i
- b is the least squares estimate for B based on the common samples.

3.3.3.3 Example Application

In this section, two examples of double sampling are considered. First, the procedure and data used in the analysis will be described, then the example problems will be presented and finally results and comments will be given.

The examples considered are based on the C/S-1 corn and soybeans procedure discussed in Section 3.2.7. This procedure generates two types of crop estimates: (1) Stage 1 estimates that are produced early in the procedure (before analyst labeling), and (2) Stage 2 estimates which comprise the final results of the procedure. We wish to produce many Stage 1 estimates, and a smaller number of the more expensive Stage 2 estimates, in order to achieve better overall performance (e.g., lower variance) for a given cost.

The data base used in the analysis is composed of the corn and soybean segment estimates, both Stage 1 and Stage 2, that were obtained from 39 segment processings of Procedure C/S-1 carried out in early 1981 at JSC. This data base is more fully explained in Section 3.2. .

In the first example, we establish a hypothetical problem that an estimation system manager would face. Table 3.10 presents a list of the constraints, which were selected to be reasonable within the C/S-1 procedure operational environment. The question being asked is: "How many Stage 1 and how many Stage 2 samples should be processed to obtain the best overall estimate"?

In the second example, Constraints 2 and 3 are changed to 320 analyst hours and 30 computer hours.

As described in the previous section, this question is tackled by mathematically setting up the constraint space, identifying the objective function to minimize, and carrying out an integer programming algorithm to minimize the objective function subject to the constraints.

TABLE 3.10. HYPOTHETICAL CONSTRAINTS FOR CONDUCTING C/S-1 IN AN OPERATIONAL ENVIRONMENT

1. Manager has 2 weeks (ten 8-hour working days to obtain an estimate.

2. The system has five analysts at its disposal, i.e., a maximum of 400 hours.

3. The system has at its disposal a maximum of 35 hours of computer time.

4. Costs of resources for processing include:

Stage 1	2 analyst hours	.25 computer hours
Stage 2	8 analyst hours	.5 computer hours

5. The data for sufficient number of segments is available and is not counted in the cost analysis.

The constraints reduce to the following, where n' is the number of samples of less expensive (Stage 1) technique, and n is the number of samples of the more expensive (Stage 2) technique.

(1) $$\begin{bmatrix} 2 & 8 \\ 0.25 & 0.5 \end{bmatrix} \begin{bmatrix} n' \\ n \end{bmatrix} \leq \begin{bmatrix} 400 \\ 35 \end{bmatrix} \quad \text{(for Problem 1)}$$

$$\leq \begin{bmatrix} 320 \\ 30 \end{bmatrix} \quad \text{(for Problem 2)}$$

(2) $n' \geq n$ (since Stage 1 estimates always exist if a Stage 2 estimate is produced)

(3) $n > 10$ (to insure sufficient significance in the relation that is formed between the two types of estimates)

These constraints are plotted in Figures 3.23 and 3.24 for the two examples.

In a one-crop example, the object function is simply the variance of the overall crop proportion estimate. When more than one crop is involved, such as in the examples presented in this section, there are many reasonable object functions. For instance:

(1) Variance of corn estimate

(2) Variance of soybean estimate

(3) Sum of variance of each crop estimate

(4) Maximum of variance of each crop estimate

In the results presented below, each of these was included in the evaluation.

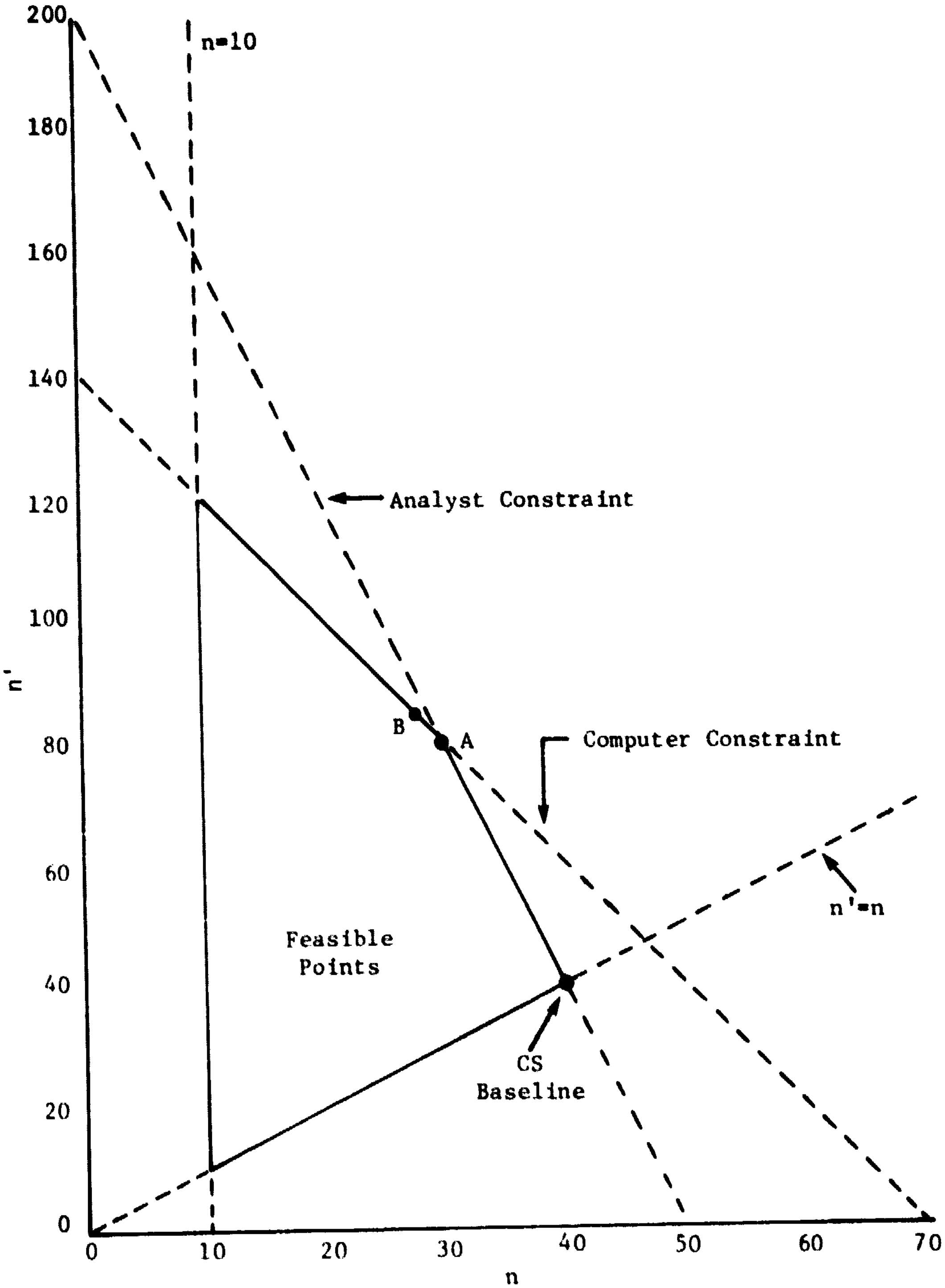


FIGURE 3.23. GEOMETRY FOR STAGE 2 WITH STAGE 1 (400 hrs analyst time, 35 computer hrs)

ORIGINAL PAGE IS
OF POOR QUALITY

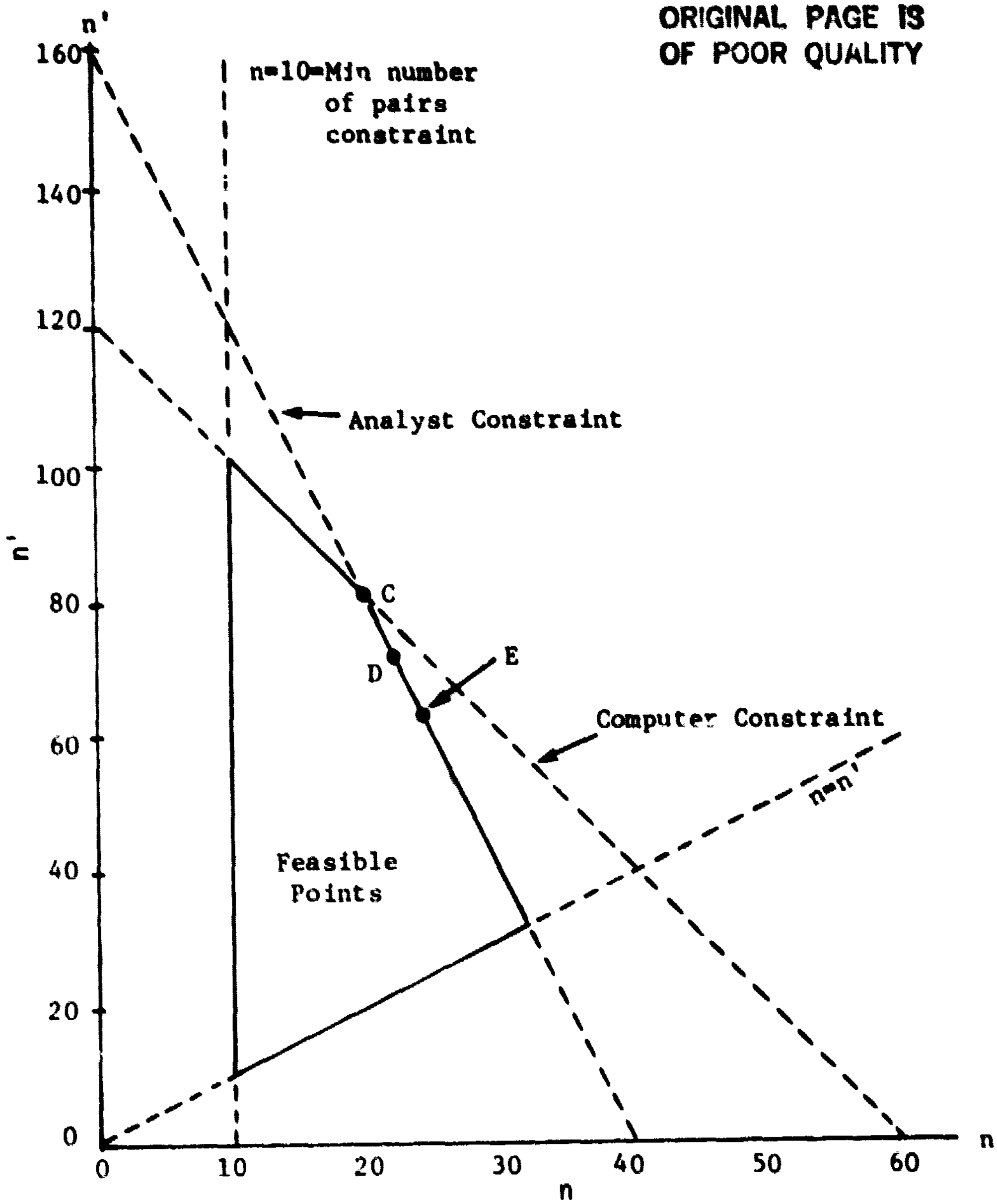


FIGURE 3.24. GEOMETRY FOR STAGE 2 WITH STAGE 1 (320 hrs analyst time, 30 computer hrs)

Denoting the Stage 2 corn and soybeans estimates as y_c and y_s, and the Stage 1 corn and soybeans estimates as x_c and x_s, the sample correlation matrix of

$$\begin{pmatrix} y_c \\ x_c \\ y_s \\ x_s \end{pmatrix}$$

was

$$\begin{bmatrix} 1.00 & .79 & .34 & .26 \\ - & 1.00 & .15 & .11 \\ - & - & 1.00 & .90 \\ - & - & - & 1.00 \end{bmatrix} .$$

The multiple R was not significantly larger than the simple correlations so only simple regression was used.

The results of the two examples are given in Table 3.11. The middle two columns represent the results of the optimized sample selection. Precision relative to baseline is a measure of improved performance resulting from the optimized choice, compared to the baseline alternative of single sampling, using the same resource constraints. The number of samples in the baseline mode is the maximum number n of Stage 2 estimates that can be afforded ($n = n'$). The column called "solution point" is the label of points in Figure 3.23 or 3.24 that represent the optimum sample selection.

In both examples, there is a clear gain in accuracy by using double sampling, and the amount of improvement is between 24 and 54%. The choice of object functions made some but relatively little difference in the results of optimization, but had a moderate effect on the measurement of relative precision.

TABLE 3.11. RESULTS OF CARRYING OUT DOUBLE SAMPLING TECHNIQUE ON TWO EXAMPLES

Example	Object Function	n' (Number of Stage 1 Samples)	n (Number of Stage 2 Samples)	Precision Relative to Baseline	Number Samples in Base-line Mode	Solution Point	Figure
1	1 (S_c^2)	80	30	1.24	40	A	3.23
	2 (S_s^2)	84	28	1.54	40	B	3.23
	3 $(S_c^2 + S_s^2)$	80	30	1.38	40	A	3.23
	4 $\max(S_c^2, S_s^2)$	80	30	1.39	40	A	3.23
2	1 S_c^2	64	24	1.24?	32	E	3.24
	2 S_s^2	80	20	1.54?	32	C	3.24
	3 $(S_c^2 + S_s^2)$	72	22	1.38?	32	D	3.24
	4 $\max(S_c^2, S_s^2)$	64	24	1.39?	32	E	3.24

3.3.4 TARGET DEFINITION ANALYSIS

As discussed in Section 3.1, target selection is an important component in crop area estimation technology. The use of quasi-fields has been emphasized in our work, but it is by no means the only workable approach. Table 3.12 lists the principal ones along with comparative attributes of each. While the existence of these alternative approaches is recognized, we have not carried out comparative evaluations of them. This section will concentrate primarily on using quasi-fields as targets for labeling.

3.3.4.1 General Remarks on Bias Characteristics Associated With Quasi-Field Definition

A key goal in defining quasi-fields is to represent true agricultural fields on the ground. If this objective is met, then quasi-field interiors are pure, and the area associated with each quasi-field is accurate. In this case, labeling of crop type of a field is more likely to be correct, and the combining of such lables, weighted by area, to form an estimate will not introduce bias.

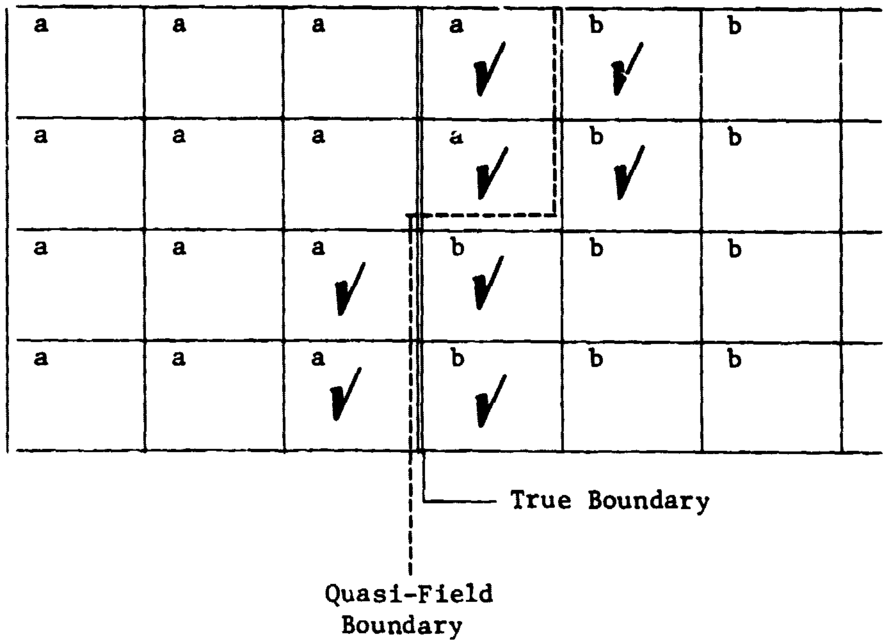
But the current quasi-field algorithms fall short of this goal. They do not perfectly locate a boundary between two distinct fields. In most cases, the algorithms successfully detect that the fields are distinct, but often there is inaccuracy in assigning pixels near the boundary to the correct field. This can introduce bias.

Figure 3.25 conceptually shows the effect of this inaccurate assignment. In the illustrated artificial region consisting of just two fields, suppose error is present in the assignment of two pixels. The results is a bias of 8% in an area estimate made over the region. Bias will be introduced over a larger region as well, when assignment error of near-boundary pixels tends to be preferential to one crop over another. This effect has been observed, and will be quantified later.

TABLE 3.12. LABELING TARGET APPROACHES

Pixels ("dots") selected from the scene

Approach	Attributes
Selected Pixels ("dots")	+ Computationally inexpensive - Mixed pixels must be handled or labeled
Selected Pixels, as Controlled by Quasi-Field Definition ("relocated dots")	- No longer inexpensive or simple • Bias characteristics same as quasi-field + Boundary pixels are identified and handled - No advantage of averaging pixels
Define Quasi-Field	- Computationally expensive + Boundary pixels are identified and handled + Target is "natural" to a human labeler + Noise reduction by averaging over pixels - Quasi-fields imperfectly represent actual fields
Select Blocks of Pixels (e.g., 3 x 3)	+ Computationally inexpensive - Mixed blocks especially hard to handle - Unnatural target + Noise reduction
Identify Spectral Distributions	- Distribution labeling requires technology, different from above, not yet perfected + Above approaches may not work in areas of very small fields • Bias characteristics more difficult to address



a, b = field number

√ = pixels marked as boundary

	Field a	Field b
True crop area	12 (50%)	12 (50%)
True interior area	8 (50%)	8 (50%)
Estimate of interior area	10 (62%)	6 (38%)
Estimate of total area	14 (58%)	10 (42%)

FIGURE 3.25. EFFECT OF INACCURATE QUASI-FIELD BOUNDARY PLACEMENT ON A CROP AREA ESTIMATE

As shown in the figure, this bias is not avoided by eliminating edge pixels. Furthermore, the bias is not avoided by using a pixel ("dot") labeling algorithm (rather than a one that labels quasi-fields) when a "dot-relocation" step is used to move pixels from the boundaries to the nearest quasi-field. This can most easily be seen by trying a 100% sample of dots on the region in Figure 3.25 and relocating the edge dots.

Evidence is presented in what follows that the situation just described hypothetically is in fact characteristic of presently used quasi-field algorithms.

3.3.4.2 Evaluation of BLOB as a Subcomponent of an Area Estimation Procedure

This section presents a detailed evaluation of the quasi-field algorithm BLOB [54] as a component of an area estimation procedure such as the one described in Section 3.2. This evaluation provides comparative information before and after two modifications in the use of BLOB that were made when the procedure was updated from C/S-1 to C/S-1A. First, the modifications will be described, then the evaluation procedures will be presented and finally, the results will be given.

The first modification involves the selection of spectral inputs to the algorithm. The change was to select at least one acquisition prior to spectral emergence of corn and soybeans, and not to use the Brightness channel of early-season acquisitions. The necessity for this change arises since sufficient information must be present in the spectral inputs so that the important crops can be distinguished. Without the change BLOB was often unable to distinguish classes such as pasture from corn or soybeans, and so these classes were sometimes lumped into the same field. The early-season Brightness channel was eliminated since at that time of year, Brightness information was sometimes found to falsely signal a boundary.

The second modification was intended to improve the purity of quasi-fields by making BLOB more sensitive to crop spectral differences that are present only within short intervals in a growing season. In order to do this, separate spectral decision thresholds were established for pre-season acquisitions and corn/soybeans separation acquisitions. A difference flagged by any one of these thresholds could then force separation into two fields.

Some terminology used in describing BLOB and its performance is needed at this point. Each pixel in a scene is assigned to exactly one blob, such that each blob consists of spatially connected and spectrally similar pixels. A pixel is in the interior of a blob if the pixel and all of the four strong neighbor pixels fall in the same blob (algorithm STRIP); otherwise the pixel is in the exterior. A big blob is a blob that has at least one interior pixel. Thus a segment is composed of three strata -- big blob interiors, big blob exteriors, and little blob exteriors. In the context of the C/S-1 and C/S-1A procedure (Section 3.2), a subset of big blobs is selected as labeling targets by a randomizing procedure, and the selected blobs are labeled according to the spectral character of the interior pixels. The blob labels are aggregated to form a segment estimate.

The evaluation consisted of computing and analyzing several performance measures listed below:

(1) Fraction of Scene
   (a) in big blob interiors
   (b) in big blob exteriors
   (c) in little blobs

(2) Purity
   (a) of big blob interiors
   (b) of big blob exteriors

(3) Impure Big Blobs Interiors (purity 80% rule)
   (a) number of them
   (b) percent by area of all big blob interiors

(4) Bias Indication
   (a) purity of corn big blob interiors
   (b) purity of corn big blob exteriors

The ground truth used for evaluation was established in the form of fraction of an area that is corn, soybean, other and unknown ground truth. Blobs containing more than 50% unknown were not used in the evaluation and other blobs containing some unknown ground truth were treated by reassigning the unknown area in proportion to the remaining three classes.

Purity (of a blob, or of a stratum of a scene) was computed as the largest of percent corn, percent soy, percent other, after the correction for unknown ground truth. Then mixed quasi-fields were identified as one whose interior pixels have purity less than a purity threshold. The threshold whose setting is an arbitrary matter of definition was held at 80% in the data that follows.

Purity values were given for corn blobs as well as for all big blobs since there was significant bias in favor of overestimating corn in the C/S-1 procedure. These values can help to understand the cause for some of this bias.

Three configurations of BLOB were tested. They are:

(A) the version used in C/S-1

(b) the same BLOB algorithm as in (A), but with revised acquisition selection procedure (first modification).

(C) the version used in C/S-1A. This involves both the revised acquisition selection and the spectral decision threshold modification (first and second modifications).

The two modifications, especially the change in spectral inputs, clearly improved the performance. Blob purity was improved, dramatically from about 85% to about 90% and the fraction of the scene in mixed blob interiors was reduced from 26% to 16% (Tables 3.13, 3.14).

However, there was a negative side to the changes. The percent of the scene in blob interiors was decreased by 8% and the percent of the scene in small blobs (with no interior pixels) was increased by 11%. This factor by itself could increase bias in a segment estimate unless methods for extending estimates to this stratum are sufficiently robust.

Of the two modifications, the most significant one is the change of spectral inputs. Most of the increased purity and decreased occurrence of mixed blob interiors was due to its effect. The unwanted changes in balance between big and little blobs was due about equally to each of the two changes.

The net impact on the procedure of making the two changes was positive. Additional evidence of this net positive impact has been given in Section 3.2 in which procedure test results were discussed.

An important effect that was observed in the C/S-1 procedure results was a positive bias in favor of corn. In order to examine this effect, purity was computed separately for the set of corn blobs. The results show that corn blobs are very consistently less pure than all blobs taken together. The magnitude of the difference in purity is about four percentage points for blob interiors and about eight percentage points for blob exteriors, and these differences hold true independent of which configuration of blob was used.

TABLE 3.13. RESULTS OF BLOB SUBCOMPONENT TESTS

Parameter	Configuration of Blob	Segments 202	205	800	832	842	Average
Fraction Interior	A	0.38	0.37	0.38	0.30	0.35	0.36
	B	0.34	0.33	0.37	0.24	0.30	0.32
	C	0.29	0.31	0.34	0.21	0.26	0.28
Fraction Exterior	A	0.53	0.50	0.50	0.55	0.51	0.52
	B	0.51	0.49	0.48	0.50	0.48	0.49
	C	0.50	0.50	0.47	0.47	0.45	0.48
Fraction Little	A	0.09	0.14	0.13	0.15	0.14	0.13
	B	0.14	0.17	0.15	0.26	0.22	0.19
	C	0.20	0.20	0.19	0.31	0.30	0.24
Number of Mixed Big Blobs (interior purity 80%) Compared to Total Number of Big Blobs	A	108/470	108/401	90/439	130/551	100/460	107/464
	B	58/475	83/402	66/403	81/533	68/442	71/436
	C	55/486	82/423	65/418	110/570	56/436	74/467
Percent of Big Blobs That Have Mixed Interiors	A	23.0	26.9	20.5	23.6	21.7	23.1
	B	12.2	20.6	16.4	15.2	15.4	16.3
	C	11.3	19.4	15.6	19.3	12.8	15.3

TABLE 3.13. RESULTS OF BLOB SUBCOMPONENT TESTS (Continued)

Parameter	Configuration of Blob	Segments 202	205	800	832	842	Average
Percent of Big Blob Interior Pixels in Mixed Blobs	A	26.2	22.9	17.8	37.5	23.7	25.6
	B	10.3	19.9	14.6	21.2	10.9	15.4
	C	9.2	22.8	14.3	25.7	10.1	16.4
Interior Purity (all big blobs)	A	87.0	88.7	89.7	83.4	87.5	87.3
	B	94.0	90.1	92.3	89.6	92.6	91.7
	C	95.0	90.5	93.5	88.8	93.6	92.3
Exterior Purity (all big blobs)	A	70.7	72.9	72.1	66.3	68.2	70.0
	B	79.2	76.9	76.1	73.0	73.2	75.7
	C	80.6	77.5	77.9	75.1	76.8	77.6
Interior Purity (corn big blobs)	A	83.8	81.0	89.3	74.0	86.3	82.9
	B	91.8	81.3	92.7	81.2	92.3	87.9
	C	91.3	83.9	93.3	82.6	92.5	88.7
Exterior Purity (corn big blobs)	A	64.9	57.1	73.2	50.6	65.6	62.3
	B	73.9	62.1	75.5	56.0	71.4	67.8
	C	73.3	64.7	77.6	59.7	74.0	69.9

TABLE 3.14. SUMMARY RESULTS OF BLOB SUBCOMPONENT TESTS

	Blob Configuration		
	A	B	C
Fraction big blob interior	.36	.32	.28
Fraction big blob exterior	.52	.49	.48
Fraction little blob	.13	.19	.24
Number of mixed big blobs	107	71	74
Number of big blobs	464	436	467
Fraction of blobs mixed	23.1	16.3	15.3
Interior area fraction mixed	25.6	15.4	16.4
Interior purity (big blobs)	87.3	91.7	92.3
Interior purity (corn big blobs)	82.9	87.9	88.7
Exterior purity (big blobs)	70.0	75.7	77.6
Exterior purity (corn big blobs)	62.3	67.8	69.9

This observation fulfills the expectation developed in the preceding section that inaccurate blob boundary placement would sometimes occur and cause bias. In the next section, some reasons for this behavior are postulated.

3.3.4.3 Bias and Its Causes and Treatment

In the last section, it was shown that the BLOB algorithm acts in a biased way toward at least one specific crop. During a significant part of the growing season, e.g., when corn is rapidly accumulating biomass and then becoming ripe, corn's spectral distribution, first, is more narrow than most other crops, especially soybeans, and second, is more centrally located in spectral space. The first characteristic (narrow spectral distribution) is thought to interact with BLOB's algorithm in a way that tends to incorporate more variance into each corn blob before BLOB forces a new blob to be defined. The second characteristic (central spectral location) can allow certain mixtures of non-corn crops to look like corn and can cause a spectral mixing between corn and most other crops. Any or all of these explanations (or others) could be the cause of the observed low corn purity and bias.

Other quasi-field algorithms also are subject to similar effects, perhaps for different reasons. For example, if a fixed spectral decision line is used, scene spectral effects can cause bias in favor of one crop or the other.

In an example run of a different quasi-field algorithm [55] based on superposition of edges formed by spectral decision boundaries, the presence of non-uniform purity values among crops was also observed as shown in Table 3.15. We would expect this segment to exhibit an overestimate of soybeans, and an underestimate of the less pure corn and other categories. There is little guarantee that the

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 3.15. PURITIES BY CROPS IN CATE-DENNIS QUASI-FIELD ALGORITHM (One Segment)

	Purity of Corn	Purity of Soy	Purity of Other
Quasi-Field Interior	0.793	0.843	0.844
Quasi-Field Edge	0.647	0.810	0.664

direction of the bias for this algorithm is consistent, or that it would cancel out over an ensemble of segments. The effects described above should be taken into account in defining improved quasi-field techniques. An improvement in purity or a reduction in purity differences can have a favorable influence on the bias of a procedure. If any of the above-mentioned potential bias-causing mechanisms can be circumvented, possibly by using an edge detection and placement approach that does not rely on specific spectral conditions, bias may also be reduced.

3.3.5 ARGENTINA GROUND DATA PREPARATION

3.3.5.1 Introduction

In February 1981 a ground data mission in Argentina was successfully carried out by Supporting Research personnel from ERIM and UCB. This activity, described in Section 2.4.2, and also in the 1981 Ground Data Collection Report [20], generated numerous kinds of information, most notably crop identifications for visited fields in 15 segments. In this section, we describe an activity that used this information and one site visited by a USDA team* to produce a digital ground truth image, registered to Landsat data, for each segment visited.

This product, as discussed below, was configured to be as similar as possible to ground truth products, called UGTT's, commonly used in the AgRISTARS program. Three key differences of this product from the UGTT product are worthy of special note.

First, the nature of the Argentina survey did not permit wall-to-wall ground data collection in a segment. Since activities were limited to main roads, fields visited were in linear strings within a segment.

*Segment 685, San Pedro (33°57'S/59°46'W) by C. Caudell et al, 15 Dec 1980.

On the average 41 fields were visited in each segment (range of 18 to 117 fields) for a total of 651 fields in the 16 segments surveyed.

Second, the base map which the data collection team used for annotation of crop codes was not high-resolution aircraft photography, but rather Landsat imagery, enlarged to 1:85,000, for one date only. For eight segments, acquisitions were used that had been acquired within two months of the mission. For the rest, acquisitions used were acquired from five to six months prior to the mission.

And finally, the Landsat data that was used was provided in a form different from the form traditionally used. The pixels were sampled to form a 57 x 57 meter grid, rather than the usual 57 x 79 meter grid. The segment size remained 5 x 6 miles, but the number of scan lines was increased from 117 to 162. The ground truth information was sampled at the rate of 3 per scan line and 2 per pixel along the scan line, as in the UGTT products, but this scan line sampling rate is subject to the same resolution change as the associated Landsat data.

3.3.5.2 Approach

The following steps were carried out in making the digital ground truth products discussed above:

(1) Staff members familiar with the Landsat data, and with the data collection activity, delineated the position of field boundaries on the 1:85,000 base image that was annotated with the ground truth data, and assigned field numbers. The image used for this delineation was the base acquisition used at JSC for Landsat data registration.

(2) The delineated field boundaries were digitized on an x-y coordinate digitizer, and recorded in a polygon format. Descriptive information including field number and crop type was recorded with each field polygon.

(3) The digitizer coordinates were converted to Landsat line and point numbers. This step required no special registration step since the base image was already in Landsat coordinates.

(4) A computer algorithm effectively placed a 28.5 by 19 meter grid (1/2 x 1/3 pixel grid) over the field polygons, and assigned the proper field number to each grid position. For each pixel, the ground truth code for the associated field was placed into the output image.

(5) A quality assurance check of the encoded image data was carried out for each site. This check consisted primarily of the following two steps. First, a computer generated list of each field with its associated ground truth code was checked against the original list provided by the data collection team. Second, a map displaying field numbers was generated. The map was visually compared to the Landsat image to insure proper location, shape and relationship to other fields on the image. Once these steps were completed, any errors detected were corrected.

(6) Both the polygon data and the encoded image are retained in a data base. The encoded image data, which has been carefully checked, has been made available in the form described in the next section.

3.3.5.3 Data Base Description

The data prepared as described above exists in the form of UGTT products (images giving crop codes). This section describes the format of these products.

This data product makes use of the crop ground truth codes given in Table 3.16. These codes were, as much as possible, taken from those given in the 1981 Enumerator's Manual (JSC-16860). In a few cases additional codes (marked with * in Table 3.16) were defined in order to cover conditions found in Argentina that were not handled by the pre-existing codes.

TABLE 3.16. CROP CODES USED IN ARGENTINA GROUND TRUTH DATA PRODUCTS

Crop	Crop Code
Alfalfa	101
Corn	105
Oats	111
Peanuts	112
Soybeans	119
Sorghum	120
Sunflower	121
Winter Wheat	125
Grasses	131
Other Hay	132
Pasture	134
Trees ≥ 8 pixels	135
Water > 5 acres	136
Non-Agricultural	140
Idle Land/Fallow	231
Previous Year Residue/Stubble	232
Mixed Crop	233
Problem Field	99
Non-Inventoried	255
Bare Soil	128*
Internal Drainage, Drainage Way	129*
Chicory	130*
Natural Vegetation (Non-Ag)	141*
Corn or Sorghum	143*

*New codes unique to Argentina data

For convenience, Table 3.17 is provided to identify the status of related Landsat data. This table presents the Landsat acquisition used during field work, the Landsat acquisition that was used by JSC for registration (and that was used for delineation and digitization), and the number of acquisitions that exist.

The format of the UGTT product is Universal format [56], a format widely used at JSC. In this product, each pixel in the ground truth image consists of one channel ground truth code. Each 2-pixel by 3-scan line array of codes in the ground truth image represents one Landsat pixel. As previously noted, the Landsat pixel size used is 57 x 57 meters rather than the usual 57 x 79. The ground trutn code actually stored on tape is a modification of crop code presented in Table 3.16. If each code is interpreted as a positive 8 binary bit number, the modification is:

Table 1 Code	Action
less than 128	add 128
greater than or equal to 128	subtract 128

(This artifact is retained in order to conform to other UGTT products produced at JSC.)

One UG product for each of the 16 segments is stored in a data base that has been made available to JSC. This data base also includes special notations for each segment identifying any special comments or considerations.

TABLE 3.17. LANDSAT DATA ASSOCIATED WITH COLLECTED GROUND TRUTH

Segment	Acquisition Used as Base Map During During Ground Truth Inventory	Acquisition Used for Delineating Field Boundaries*	Number of 1980/1981 Crop Year Acquisitions	Number of Fair or Better Quality Acquisitions
527	24 Aug 80	24 Apr 81	9	5
692	12 Dec 80	24 Nov 80	14	10
616	12 Dec 80	24 Nov 80	12	8
681	29 Sep 80	5 Jan 81	17	9
520	10 Dec 80	19 Mar 81	8	6
561	24 Aug 80	5 Jan 81	17	9
685	---	5 Jan 81	8	5
649	10 Dec 80	19 Mar 81	15	9
611	12 Dec 80	24 Nov 80	14	3
511	24 Aug 80	5 Apr 81	17	9
556	12 Sep 80	11 Feb 81	18	8
570	10 Dec 80	10 Dec 80	13	6
682	23 Aug 80	5 Apr 81	7	5
677	12 Sep 80	12 Sep 80	17	5
604	12 Dec 80	24 Nov 80	14	10
578	10 Dec 80	19 Mar 81	9	6

*Also is base acquisition for Landsat Registration

3.4 INVENTORY TECHNOLOGY DEVELOPMENT CONCLUSIONS AND RECOMMENDATIONS

An end-to-end analyst-based, computer-aided crop inventory method for crop inventory without in situ training data has been developed and tested. This procedure, termed the Baseline Corn and Soybean Procedure sought to formalize an analyst interpreter based technology into one that would be essentially automatable. Detailed analysis of results enabled the development of procedural modifications that would improve the procedure's precision while automating certain processes, particularly the analyst logic for crop identification.

In addition to the research conducted in end-to-end estimation procedures, advanced component procedures have been examined. Initial understanding of the spectral/temporal nature of corn and soybean confusion crops, particularly sunflowers and sorghum has been formulated. The evaluation of analytical profile techniques as a method to extract features from multitemporal spectral trajectories revealed very promising results. Features related to a crop's rate of emergence and senescence, growing season length and peak spectral response were derived and found to contain sufficient discriminating potential to produce accurate crop area estimates. Examination of the appropriate target selection procedures for automatic labelers was initiated. It was found that current techniques for automatic definition of 'fields' as targets could introduce bias into estimates due to inconsistent treatment of pixels as a function of crop class. For example, the BLOB procedures tend to produce consistently larger targets that are predominantly corn (due to the central position corn occupies in spectral space).

As a result of the research conducted in support of the Inventory Technology Development Project, the following key recommendations are made:

PRECEDING PAGE BLANK NOT FILMED

• The development of completely automatic techniques for crop area estimation should be pursued; automatic technology, beyond its operational efficiency, enables the diagnosis of problem areas in a shorter turnaround time resulting in a more rapid development cycle.

• Much of the current research has stressed at harvest estimation, the development of early seasons methods remains critical.

• The adaptation of Landsat-based inventory technology from the U.S. to the Southern Hemisphere will encounter a crop mix and agricultural environment significantly different; emphasis should be placed on developing a thorough understanding of the spectral/temporal characteristics of key crops (corn, soybeans, rice, cotton, sorghum, sunflowers) as well as cropping practices (e.g., crop calendars); it should be well understood to what degree Landsat can support crop identification and discrimination in that environment so as to set realistic expectations on the technology.

• As seen in both SR analysis in the small grains application (Section 2.7) and in the ITD analysis (Section 3.3), profile-based technology is an extremely promising approach; efforts should be extended in this direction in addition to the expert-based methods; the two approaches coupled in a comprehensive research program would provide a penetrating understanding of the potential of Landsat-based crop inventory technology.

• The identification of an appropriate target provided to analyst interpreters or to machine classifiers remains an unresolved technical issue; the resolution of MSS results in mixture pixels that must be interpreted or classified in an unbiased manner; in addition, multitemporal analysis of such targets requires highly accurate acquisition-to-acquisition registration; both quasi-field-based and pixel-based labeling strategies need to be evaluated to establish their attributes with respect to the bias or variance that they introduce that are unrelated to sampling but to target feature selection; in addition methods should be explored that relax the registration requirement.

REFERENCES

1. Odenweller, J., C. Hay and B. Wood, Interpretation and Analysis of Scanner Data for Crop Inventory, NAS9-14565, University of California at Berkeley, Space Sciences Laboratory, Berkeley, CA (to be published October 1982).

2. Thelen, B. and D. Mitchell, Remote Sensing Augmentation of Conventional Crop Acreage Response Models for Estimation of Corn, Soybeans, and Sorghum Acreages, Report No. 152400-16-T, Environmental Research Institute of Michigan, Ann Arbor, MI (to be submitted for publication).

3. Pestre, C. and W. Malila, Using Knowledge of Agricultural Practices to Enhance Through-the-Season Interpretation of Landsat Data, Environmental Research Institute of Michigan, Ann Arbor, MI (to be submitted for publication in Proceedings of the Fifteenth International Sympoisum on Remote Sensing of Environment).

4. Umberger, D.E., M.H. Proctor, J.E. Clark, L.M. Eisgruber and C.B. Braschler, Econometric Models for Predicting Confusion Crop Ratios, The LACIE Symposium, Volume 1, NASA/Johnson Space Center, Houston, TX, pp.19-46, October 1978.

5. Houck, et al., Analyzing the Impact of Government Programs on Crop Acreage, NTIS, 1976.

6. Johnston, J., Econometric Methods, McGraw Hill, 1972, pp.312-313.

7. Mitchell, D., Department of Agricultural Economics, Michigan State University, East Lansing, MI, private communication.

8. Crist, E.P. and W.A. Malila, A Temporal-Spectral Analysis Technique for Vegetation Applications of Landsat, Proceedings of the 14th International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, MI, April 1980.

9. Badhwar, G.D., A Semi-Automatic Technique for Multitemporal Classification of a Given Crop, AgRISTARS Report SR-J0-00481, NASA/ Johnson Space Center, Houston, TX, July 1979.

10. Reboh, R., Knowledge Engineering Techniques and Tools in the Prospector Environment, Technical Note 243, SRI International, Menlo Park, CA, June 1981.

REFERENCES (Continued)

11. Shortliffe, E.H., Computer Based Medical Consultations: MYCIN, Elsevier Computer Sciences Library, Artificial Intelligence Series, Nils J. Nilsson Editor, American Elsevier Publishing Company, New York, NY, 1976.

12. Heydorn, R.P., R.M. Bizzell, J.A. Querin, K.M. Abotteen and C.A. Sumner, Classification and Mensuration of LACIE Segments, Proceedings of LACIE Symposium, NASA/Johnson Space Center, Houston, TX, October 1978, pp.73-86.

13. Cicone, R., E. Crist, R. Kauth, P. Lambeck, W. Malila and W. Richardson, Development of Procedure M for Multicrop Inventory, With Tests of a Spring Wheat Configuration, Final Report 132400-11-F, Environmental Research Institute of Michigan, Ann Arbor, MI, March 1979.

14. Perry, C.R. and C.R. Hallum, Sampling Unit Size Considerations in Large Area Crop Inventory, Using Satellite-Based Data, NASA/Earth Observations Division Technical Report, JSC-13767, 1979.

15. Wheeler, S.G. and P.N. Misra, Large Area Inventory of Crops With Landsat Data, Procedure 2: Functional Elements and Development Plan, IBM, Federal Systems Division, Houston, TX, May 1978.

16. Kauth, R.J. and W. Richardson, Procedure B: A Multispectral Training Selection and Proportion Estimation Procedure for Processing Landsat Agricultural Data, Technical Report 122700-31-F, Environmental Research Institute of Michigan, Ann Arbor, MI, November 1977.

17. Kauth, R. and W. Richardson, Signature Extension Methods in Crop Area Estimation, The LACIE Symposium, NASA/Johnson Space Center, Houston, TX, October 1978, pp.801-824.

18. Cochran, W.G., Sampling Techniques, Third Edition, John Wiley and Sons, New York, NY, 1977.

19. Hicks, D.R., Agronomic Characterization of the Argentina Indicator Region, SR-E2-04222, NAS9-15476 and NAS9-14565, ERIM #152400-19-T, Environmental Research Institute of Michigan, Ann Arbor, MI, December 1981.

20. Hicks, D.R., B. Sellman, E. Sheffner, G. Thomas and B. Wood, 1981 Argentina Ground Data Collection, SR-E1-04174, NAS9-15476, NAS9-14565, ERIM #152400-12-T, Environmental Research Institute of Michigan, Ann Arbor, MI, October 1981.

REFERENCES (Continued)

21. Gardner, W. and V. Johnson, Brazil AgRISTARS Area Frame Evaluation, Trip Report for February 2-20, 1981, USDA/ESS, April 1981.

22. Hicks, D.R., Notes for Brazil Sampling Frame Trip, SR-E1-04138, NAS9-15476, ERIM #152400-7-T, Environmental Research Institute of Michigan, Ann Arbor, MI, August 1981.

23. Sellman, B. and D.R. Hicks, Argentina Ground Data Collection Plan for 1981-1982 Crop Year, SR-E1-00644, NAS9-15476 and NAS9-14565, ERIM #152400-14-T, Environmental Research Institute of Michigan, Ann Arbor, MI, November 1981.

24. Sellman, B. and D.R. Hicks, Plan de Verificacion de Campo y Coleccion de Datos en Argentina Para El Ano Agricola 1981-1982, SR-EL-00644, NAS9-15476 and NAS9-14565, ERIM #152400-14-T, Environmental Research Institute of Michigan, Ann Arbor, MI, November 1981.

25. Crist, E.P., Cultural and Environmental Effects on the Spectral Development Patterns of Corn and Soybeans - Field Data Analysis, SR-E2-04224, NAS9-15476, ERIM #152400-20-T, Environmental Research Institute of Michigan, Ann Arbor, MI, January 1982.

26. Hay, C.M., et al., Development of AI Procedures for Dealing with the Effects of Episodal Events on Crop Temporal-Spectral Response and Development of AI Guidelines for Corn and Soybean Labeling, SR-B9-00434, University of California, Space Sciences Laboratory, Berkeley California, 1979.

27. Bauer, M.E., L.L. Biehl and B.F. Robinson, Field Research on the Spectral Properties of Crops and Soils, SR-PO-04022, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IND, 1980.

28. Dale,R.F., D.T. Coelho and K.P. Gallo, Prediction of Green Leaf Area Index for Corn, Agronomy Journal, 72:999-1005, 1980.

29. Sternberg, S.R., Cellular Computers and Biomedical Image Processing, Proceedings of the U.S.-France Seminar on Biomedical Image Processing, Grenoble, France, 1980.

30. Badhwar, G.D., Crop Emergence Date Determination from Spectral Data, Photogrammetric Engineering and Remote Sensing, 46:369-377, 1980.

REFERENCES (Continued)

31. Cicone, R.C., E. Crist, R. Kauth, P. Lambeck, W. Malila and W. Richardson, Development of Procedure M for Multicrop Inventory, with Tests of a Spring Wheat Configuration, Final Report 132400-16-F, Environmental Research Institute of Michigan, Ann Arbor, MI, 1979.

32. Kauth, R.J. and G.S. Thomas, The Tasseled-Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat, Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue/LARS, West Lafayette, IND, 1976.

33. Holmes, Q.A., R. Horvath, R.C. Cicone, R.J. Kauth and W.A. Malila, Development of Landsat-Based Technology for Crop Inventories, Volumes 1 and 2, Final Report 132400-29-F, Environmental Research Institute of Michigan, Ann Arbor, MI, December 1979.

34. Malila, W., J. Gleason and R. Cicone, Multispectral System Analysis Through Modeling and Simulation, Proceedings of the Eleventh International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, MI, April 1977.

35. Malila, W., J. Gleason, F. Sadowski, R. Cicone and E. Crist, Applications of Modeling to Analysis and Processing of Landsat Data, Proceedings of the Twelfth International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, MI, April 1978.

36. Ritchie, J.T., USDA Grassland, Soil and Water Research Laboratory, Temple, TX (Private conversation).

37. Jackson, R. and P. Pinter, USDA, U.S. Water Conservation Laboratory, Phoenix, AZ (private communication).

38. Suits, G.H., The Calculation of the Directional Reflectance of a Vegetation Canopy, Remote Sensing of Environment, 2, p.117, 1972.

39. Suits, G.H., The Calculation of the Directional Reflectance of a Vegetative Canopy and The Cause of Azimuthal Variations in Directional Reflectance of Vegetative Canopies, Remote Sensing of Environment, 2, (p.117 and p.175), 1972.

40. Allen, W., T. Gayle and A. Richardson, Plant Canopy Irradiance Specified by theDuntley Equations, JOSA, 60, p.372, 1970.

REFERENCES (Continued)

41. Pont, W.F., Spatial and Spectral Simulation of Landsat Images of Agricultural Areas, ERIM #152400-18-T, SR-E1-04206, Environmental Research Institute of Michigan, Ann Arbor, MI, March 1982.

42. Dye, R.H., Restoration of Landsat Images by Discrete Two-Dimensional Deconvultion, Proceedings of the Tenth International Symposium on Remote Sensing of Environment, Ann Arbor, MI, October 1975.

43. Crist, E.P. and W.A. Malila, Development and Evaluation of an Automatic Labeling Technique for Spring Small Grains, SR-E1-04065, ERIM #152400-3-T, Environmental Research Institute of Michigan, Ann Arbor, MI, August 1981.

44. Finkler, J.A. and E.P. Crist, Documentation of Computer Procedures for Labeling Spring Grains and Discriminating Between Spring Wheat and Barley Using Landsat Data, SR-E1-04131, ERIM #152400-5-T, Environmental Research Institute of Michigan, Ann Arbor, MI, August 1981.

45. Crist, E.P. and W.A. Malila, A Temporal-Spectral Analysis Technique for Vegetation Applications of Landsat, Proceedings of Fourteenth International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, MI, April 1980.

46. Clinton, N.J., LACIE Phase III Labeling Error Characterization, Final Report, Lockheed Electronics Company, Inc., Technical Memo LEC-13012, March 1979.

47. Clinton, N.J., Transition Year Labeling Error Characterization Study Final Report, Lockheed Electronics Company, Inc., LEMSCO-14056, October 1980.

48. Rice, D. and T. Wessling, Ground Truth Crop Proportion Summaries for U.S. Segments, 1976-1979, SR-E1-04189, ERIM #152400-13-T, Environmental Research Institute of Michigan, Ann Arbor, MI, October 1981.

49. Dailey, C.L., K.M. Abotteen and J.D. Nichols, Corn/Soybean Decision Logic: Improvements and New Crops, NASA Report FC-L0-00420, Lockheed Engineering and Management Services Company, Inc., Houston, TX, 1980.

C-4

REFERENCES (Continued)

50. Hay, C.M., C.A. Kuretz, J.B. Odenweller and E.J. Sheffner, Development of AI Procedures for Dealing with the Effect of Episodal Events on Crop Temporal-Spectral Response and Development of AI Guidelines for Corn and Soybean Labeling, Space Sciences Laboratory Series 20, Issue 24, Space Science Laboratory, University of California, Berkeley, CA, 1979.

51. Cicone, R., C. Hay, R. Horvath, M. Metzler, O. Mykolenko, J. Odenweller and D. Rice, Users Manual for the U.S. Baseline Corn and Soybean Segment Classification Procedure, AgRISTARS Report FC-E1-00712, Environmental Research Institute of Michigan, Ann Arbor, MI, March 1981.

52. Roller, N., K. Johnson, J. Odenweller and C. Hay, Analyst Handbook for the Augmented Corn and Soybean Segment Classification Procedure (C/S-1A), AgRISTARS Report FC-E1-00723, Environmental Research Institute of Michigan, Ann Arbor, MI, October 1981.

53. Pont, F., H. Horwitz and R. Kauth, Estimating Acreage by Double Sampling Using Landsat Data, AgRISTARS Report IT-E2-04246, Environmental Research Institute of Michigan, Ann Arbor, MI, January 1982.

54. Kauth, R.J., A.P. Pentland and G.S. Thomas, BLOB, An Unsupervised Clustering Approach to Saptial Preprocessing of MSS Imagery, Eleventh International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, MI, Vol. 2, 1977, pp.1309-1317.

55. Dennis, T.B., Spatial/Color Sequence Proportion Estiamtion Techniques, NASA Report SR-LO-04028, LEMSCO, Houston, TX, December 1980.

56. Earth Resources Data Format Control Book, Publication Number PHO-TR 543, NASA/Johnson Space Center, Houston, TX, October 1973.