GEORGE HANUSCHAK'S COPY

JSC-14550

INDEPENDENT PEER EVALUATION OF THE LARGE AREA CROP INVENTORY EXPERIMENT

THE LACIE SYMPOSIUM

OCTOBER 1978

NASA
National Aeronautics and
Space Administration

Lyndon B. Johnson Space Center
Houston, Texas 77058

PREFACE

The Large Area Crop Inventory Experiment made extensive use of independent peer reviews during its three phases. These reviews, conducted at regular intervals, consisted of in-depth appraisals of the LACIE technology by recognized experts from government, industry, and universities. Many of the significant technology improvements implemented during LACIE had their origins in those Peer Reviews.

In preparing for the LACIE symposium, it was decided that an independent peer evaluation of LACIE, including its accomplishments, shortcomings, and implications for future remote sensing of agriculture, should be included.

This document contains two peer evaluation papers: the first is an overview as compiled by the Peer Plenary Team; the second is a comprehensive review that is divided into six chapters, each representing the work of the technical review teams.

CONTENTS

An Independent Evaluation by the Plenary Peer Review Team

General Chairman—D. Paarlberg, Purdue University

L. Eisgruber, Oregon State University—Cochairman, Experiment Results Team
B. Scherr, Data Resources, Inc.—Cochairman, Experiment Results Team
H. O. Hartley, Texas A & M University—Chairman, Experiment Design Team
D. Ingram, IBM—Cochairman, System Implementation and Operations Team
J. Quirein, Schlumberger—Cochairman, System Implementation and Operations Team
D. Goodenough, Canadian Center for Remote Sensing—Chairman, Data Processing Systems Design Team
G. Nagy, University of Nebraska, Lincoln—Chairman, USDA Applications Test System Team
R. Holmes, General Motors Institute—Cochairman, Supporting Research Team
R. Shay, Oregon State University—Cochairman, Supporting Research Team

INTRODUCTION

It is the intent of this paper to synthesize, in summary form, the peer group's technical evaluation of the Large Area Crop Inventory Experiment (LACIE) and to assess the potential utilization of the LACIE technology base to augment current global commodity production forecast systems.

In this summary paper, the following major issues will be addressed.

1. How capable was the technology in meeting its primary goal of providing improved wheat production information in the specific regions where it was evaluated?
2. How likely is it that the technology will succeed in other important wheat regions?
3. How applicable is the technology to other crops?
4. Would a future crop inventory system based on the technology be affordable (in view of the value of the information produced by it)?
5. Does the operational implementation of the technology appear to be feasible?
6. What important technical problems need to be surmounted in the development of an operational crop inventory system based on LACIE technology?
7. What are the future possibilities for pursuing the implementation and further development of this survey technology?

We will address each of these to the best of our knowledge and ability on the basis of our understanding of the results of LACIE and our view of how such a technology could be utilized.

OBJECTIVES OF THE EXPERIMENT

It is important in the technical evaluation of an experiment design and its achievements to pay due consideration to the experiment objectives. In our review of the LACIE documentation, we found that such objectives had been stated at the outset of LACIE and had not changed considerably. The general objectives of LACIE were to

1. Demonstrate an economically important application of repetitive multispectral remote sensing from space.
2. Test the capability of Landsat, together with climatological, meteorological, and conventional data sources, to provide improved wheat production information in important foreign producing regions.

These general experimental objectives engendered more specific objectives; namely, to

1. Provide, from an analysis of Landsat data acquired over a sample of the potential crop-producing area in major wheat-growing regions, estimates of the area planted to wheat; the use of ground-acquired identification of crop type was excluded from the analysis process to ensure a viable analysis approach in inaccessible foreign regions.
2. Provide, from an analysis of historical and real-time meteorological data, estimates of wheat yield

and combine the area and yield estimates to infer production.

3. Provide data processing and delivery techniques so that selected samples can be made available to the LACIE analyst teams for initiation of analysis no later than 14 days after acquisition of the data.

4. Provide a LACIE system design that requires a minimum of redesign and conversion to implement an operational system within the U.S. Department of Agriculture (USDA).

5. Monitor and assess crop progress (calendar) from a surface data base and evaluate the model potential for yield using surface data.

Ancillary goal-oriented activities included to

1. Make periodic crop assessments during the growing season from planting through harvest.

2. Achieve accuracy commensurate with U.S. Department of Agriculture requirements for foreign regions.

3. Improve crop inventory methodology and performance through supporting research and development.

4. Provide an objective test and evaluation program to quantify results from supporting research and development.

Periodic crop assessment reports were prepared by LACIE in Houston on a monthly basis during the crop season and mailed to the USDA LACIE office in Washington, D.C., the day before each corresponding official USDA report was released. In recognition of the objective to provide improved foreign wheat production information, an accuracy goal was set for production estimates at harvest to be within ±10 percent of true country production 90 percent of the time (referred to as the 90/90 criterion). An additional goal was to establish the accuracy of these estimates from early in the season (the first quarter of the crop cycle) through the harvest period.

LACIE RESULTS

The LACIE results to date clearly demonstrate that present remote-sensing capabilities can be combined with or substituted for conventional methods of collecting information in order to improve foreign crop production estimates. However, experiment results are not uniform throughout the growing season or from region to region. In certain regions, the LACIE results indicate that technology improvements are needed. It is important to recognize these differences and to assess experiment results in these various phases rather than simultaneously in their entirety.

Primarily because of such factors as field size and climatological conditions which influence management practices and alternatives, LACIE achieved its best results in estimating wheat production in the U.S.S.R. Information on wheat production in the U.S.S.R. from traditional sources has been and remains poor, particularly from the point of view of timeliness. At the same time, the U.S.S.R. is periodically a large purchaser of U.S. grain with significant impact on domestic as well as international prices. This, then, can only lead to the conclusion that LACIE results provide much needed information not now available from other sources. In the U.S.S.R., LACIE is of practical and immediate value.

Because of small fields, strip farming, and difficulties of separating "confusion crops," experiment results are not as impressive for the spring wheat regions of the United States and Canada. New technology such as improved satellite resolution may be required to solve these problems.

Although the original list of LACIE countries included China, India, Australia, Brazil, and Argentina, no production estimates were generated for these countries. Given the magnitude of the task and the available resources, we view the original goal of providing production estimates for all these countries as having been too ambitious. However, LACIE did conduct exploratory investigations within these countries. Although these investigations indicated that the LACIE approach would be generally applicable in these countries, additional investigations of the kind conducted in the United States, Canada, and the U.S.S.R. will be required to establish the degree of applicability.

In addition to the accuracy of commodity forecasts, another important consideration is timeliness. The LACIE data system and operations were experimental and thus not optimized for timeliness. The LACIE system required on the average 45 to 60 days between Landsat acquisition and completion of analysis of a segment. However, much of this time resulted from weekend or overnight hold time or time during which data lay on a desk awaiting analysis. Without such delays, it appears quite feasible to design an operational system that can complete analysis of the Landsat data within approximately 2 weeks of acquisition.

Regarding the question of the transferability of the LACIE technology from wheat to other crops (corn, soybeans, rice, forage, etc.), the technical approaches developed in LACIE should be generally applicable; however, considerable additional research and development may be required to adapt this technology to the characteristics of each crop. To assess more fully the utility of remote-sensing technology for the multiple crop application, additional experimental efforts will need to be made.

Yield models, which are used in conjunction with acreage estimates to arrive at production estimates, are an important component in the LACIE experiments. Given the overall performance in the U.S.S.R., a reasonable evaluation would be that yield models for that region of the world produce better information than is available currently. Such models are therefore useful, although replacement with more advanced models that account for wider ranges of weather variation must definitely be considered. Because the current stage of model development would have to be assessed as being marginally adequate for the United States and Canada, it is believed that a major effort should be put forth to develop models which do a more adequate job of yield estimation. It is recommended that such models have a good theoretical, physical, and physiological basis rather than relying exclusively on linear statistical regression, which produces models that are limited in their response to extreme conditions. It is realized that the less sophisticated models will probably be required for some time in many countries because of the limited availability of meteorological and historical data. However, techniques that permit the use of satellites to provide additional meteorological data—such as solar radiation, surface temperature, precipitation, snow cover, etc.—should be further developed. With such techniques, environmental satellites should play an increasingly important role in providing such meteorological data at finer spatial intervals and in areas where access to such data is otherwise restricted.

The primary benefit of improved yield models will most likely be improved near- and at-harvest estimates of yield. Up to a month or so before harvest, the major source of yield forecast uncertainty will be uncertainty in the weather through harvest. In this period of uncertain yield, the component of production which can be known with the greatest certainty is the growing area of the crop. While the LACIE estimates of acreage are accurate from mid-season on, they are biased low early in the season, before the crop is completely detectable in the Landsat data. It is believed that the LACIE early-season acreage estimates could be greatly improved by using statistical bias reduction techniques such as multiyear ratio estimation. An additional issue that must be resolved is the difficulty in using Landsat data to distinguish wheat from other small grains, primarily spring barley and winter rye.

Results of attempts to evaluate the economic importance of LACIE technology seems to us to be of limited usefulness to date. This limitation stems primarily from the inability to develop good estimates of the value of improved commodity forecast information. In addition, it appears to us that the economic evaluations that were conducted did not reflect an in-depth understanding of the current LACIE methodology and results or the potential results of an improved future technology.

Concerning the transfer of LACIE technology to the USDA, the LACIE technology components are being transferred partially or in full to the department's Applications Test System; however, in response to recently shifting USDA priorities, LACIE technology is being used by the Applications Test System primarily to detect and assess unusual crop conditions abroad rather than to make the quantitative commodity production forecasts evaluated by LACIE. While the Applications Test System has demonstrated it can produce a complete and timely assessment for a restricted foreign region, it will be difficult to determine the accuracy of these assessments in regions with a lack of confirming data. There has as yet been no feedback from the Foreign Agricultural Service, the intended major client of the Applications Test System. Thus, while the time is proper to review LACIE accomplishments, a review of the Applications Test System at this time is perhaps premature.

Despite the obvious room for additional research and improvement, the Large Area Crop Inventory Experiment must be considered a success. Our assessment of the experiment results leads us to the conclusion that for global wheat regions such as the U.S.S.R., the LACIE technology can be made operational and that for regions where the technology requires improvement (such as regions having small fields), funding for further research and development should be continued.

TECHNICAL PROBLEMS RESOLVED/ACHIEVEMENTS

The acquisition and handling, as well as the accurate and timely analysis, of the sheer volume of data required for a global commodity inventory including some of the most inaccessible regions of the world was recognized at the outset of LACIE as a formidable problem. Before LACIE, Landsat data had been evaluated for crop identification only in U.S. experiments and on a relatively modest geographic scale—never more than a few test sites within a state. Landsat data analysis techniques were manually oriented, requiring highly skilled professionals to spend from several hours to as much as several days to utilize these techniques successfully. Furthermore, these analysis techniques required ground observations of a relatively large sample of crop types to familiarize the analyst or to "train" the computer to recognize these crops. Such observations would not be available in inaccessible foreign regions.

The expertise to solve these problems was contained within many different disciplines: mathematics, physics, agronomy, meteorology, and engineering. Thus, it was necessary to assemble a multiagency research and management team since no one government entity contained adequate professional experience and expertise to attack the variety of problems involved. In addition, previous experience with remote-sensing technology development and transfer dictated the involvement of a user first to define realistic requirements and later to intelligently transfer a new and complex technology into an operational environment.

It was with this set of problems that LACIE began and it is within this context that the following comments and appraisements are offered.

It is our assessment that LACIE successfully assembled the remote-sensing and yield estimation technology into a data processing system capable of timely and reliable analyses at a sufficiently high data volume to monitor the vast regions required for global commodity surveys. LACIE successfully used sample survey technology to reduce the data load to a manageable size. For example, in the nine U.S. Central Plains states, about 2 percent of the total area was sampled by Landsat and monthly meteorological data were taken from approximately 200 weather stations. Furthermore, the LACIE experience indicated that the data required for area, yield, and production forecasts can be reliably acquired in most foreign areas where access to ground-acquired data would be infeasible. Possible exceptions are the historical meteorological and yield records required to develop the LACIE-type weather/yield models in China, Brazil, and Argentina. Development of models less dependent on historical data may be required for such countries. However, LACIE was quite successful in assembling the base of spectral, agronomic, and meteorological data necessary to understand the relationships among the observables and the desired forecast quantities, thus permitting the development of models required for global applications research.

The development of Landsat data analysis procedures requiring no ground observations was a critical LACIE accomplishment that made commodity surveys possible in inaccessible foreign regions. These procedures rely primarily on Landsat data acquired on multiple dates so that crops with differences in seasonal growth patterns can be reliably differentiated. A key to the rapid analysis of these data was the development of interactive computer procedures to process the multidate Landsat acquisitions for crop identification and mensuration. The 3 years of LACIE substantially transformed the state of the art of machine processing such data. The initial technology was heavily oriented toward analyst involvement. There was no capability to process multiple Landsat acquisitions from a single target. In the course of LACIE, machine-processing advances led to this multidate data analysis capability. At the same time, these advances reduced analyst involvement from 12 hours per segment to just over 3 hours per segment and reduced machine-processing time from hours to minutes. These achievements resulted in the capability to totally enumerate the acreage within a 30-square-mile area with just 3 hours of manual involvement and only minutes of machine-processing time.

An accomplishment that was key to the development of a global monitoring capability was the provision of reasonably simple mathematical yield models. These models used weather data that were routinely available on a global scale through established weather networks and enabled LACIE to provide more accurate and timely forecasts of foreign crop yield than were available from other existing systems. While improvements can and should be made to the models tested, they were shown in the Phase III U.S.S.R. evaluation to be capable of responding to and quantifying the effect on crop yield of both significantly above average and significantly below average growing conditions.

At this point, a word of caution should be injected. The simplicity of these yield models renders them extremely easy to implement and operate. It will be tempting, therefore, particularly given the desire to achieve immediate improvements in available agricultural information, to implement this capability and overemphasize the predictive capability of the models. Such action would not utilize the full range of LACIE technology, since yield is only half of the production equation. While these models produce reasonably reliable yield forecasts near harvest, weather uncertainties before harvest are perhaps the largest source of uncertainty about the ultimate size of the expected crop. Thus, early in the crop season, it is extremely important to have a reliable estimate of the land area planted to the crop. For the first two or three quarters of the crop cycle, acreage is perhaps the most reliable predictor of wheat production. In addition, the forecasts from the simple yield models can be viewed much more reliably if corroborated by indications from other data sources, such as Landsat, crop development stage, and other agronomic indicators. Furthermore, if the technology is to be improved, all available data must be considered in making the crop production forecasts. An overemphasis on the yield at this stage in the development of remote-sensing technology could detract from the advancement of the technology.

Another very important achievement of LACIE was the development of an accuracy assessment technology. A user of agricultural information is typically concerned with the reliability of the available data. The LACIE accuracy assessment effort developed the approach and technology required to estimate the accuracy of all forecasts. For each area, yield, or production forecast, a probability-of-error statement was quantitatively developed, standard deviations were computed, and bias was estimated. A remaining need is to exercise the technology under a wide range of agricultural and meteorological conditions in order to gain experience and to determine how well the technology performs in response to adverse weather events and abrupt changes in cropping practice introduced by government policy. The accuracy assessment program also produced a very comprehensive base of spectral ground observations and other data which will permit further development of the wheat forecast technology. At this point, we should reiterate a very important lesson learned in LACIE: that is, assessing the accuracies in foreign countries is difficult at any geographic level other than the national level simply because reliable, timely government assessments are not available below that level.

The experience of LACIE to date indicates clearly that the future of the technology is strongly dependent on a multiagency effort. The involvement is necessary for two reasons: (1) it brings a major potential user (the USDA) to the table, and (2) it permits the assembly of the multidisciplinary talents necessary to identify and solve the problems. The rate and degree to which the LACIE technology will be transferred to an operational status is not clear at present. In the near future, the utilization by the Foreign Agricultural Service of the LACIE Applications Test System should provide indications of the USDA approach to transferring the diverse technology developed in LACIE. If there is a criticism of the program at this point, it is that LACIE needed more involvement of the private sector, specifically agribusiness, a user that differs from the government sector in both information requirements and expertise.

Since the expansion of the LACIE technology to other countries and other crops is heavily dependent on advancements within the scientific community, we would like to comment on the aspect of basic and applied research. LACIE had a strong research effort, as evidenced by the general advancement in remote-sensing technology in those 3 years. Further, it appears that the actual research issues and problems related to commodity production forecasting have been reasonably well defined and structured in priority. LACIE is a prime example of the concept that applied research conducted as a joint effort by operationally oriented agencies such as USDA and the National Oceanic and Atmospheric Administration (NOAA) and research-oriented agencies such as NASA can produce focused research geared toward timely and realistic answers to important national needs. A major applications research effort such as LACIE benefits greatly from an integrated multidisciplinary research effort. No one institution has the capacity, either in trained researchers or in facilities, to gather and analyze data on the scale required to develop and evaluate a technology to be operated on a global scale. This is not to say that good, basic, independent research cannot contribute or is not necessary; however, the research must be conducted in accord with focused, definitive objectives if the fruits of that research are to be relevant and realistic.

In regard to the future use of this technology and the manner in which information derived from it is dispersed to the interested public, there are many important questions that need to be addressed. The trend toward on-farm storage of the nation's wheat supply makes every serious wheat farmer an alert student of prospective global supplies and their effect on the current and future market price of wheat. As a result, there is concern as to who has access to information generated by a LACIE-type system, when such access is permitted, and how it is done. Should the raw data be made available to all who desire them so that they can make their own analysis and interpretation? Should USDA and other governmental agencies provide analysis and interpretation for in-house use as well as for dissemination to the public? We recommend that these and other policy issues be addressed explicitly as part of further LACIE-type efforts. The major development of agricultural remote sensing needed at this time is the establishment of an organized user community to begin to acquire operational experience that will in turn provide needed information inputs to policy formulation and further system evolution. A next logical step is research and evaluation of the way in which the technological developments of LACIE could be organized and managed by a genuine user community.

PROBLEMS IN NEED OF SPECIAL ATTENTION

While many advancements were made by LACIE, the technology can still be considered as first generation. In the immediate future, research dollars wisely used should provide additional significant advancements. In our review of LACIE, it is apparent that the following areas deserve special attention and are likely to have the biggest payoffs in terms of technology improvement.

The further development of yield models should be based on daily or weekly rather than monthly averages of temperature and precipitation. To enhance model response to extreme conditions, the models should more realistically reflect the biological function of the plant. The increased use of environmental satellite data to augment temperature and precipitation information in regions with sparse ground networks should also be pursued. These efforts will render yield models more reliable where deviations from normal are extreme and should also improve their applicability in foreign countries with inadequate historical data for model development. However, perhaps the largest source of forecast uncertainty early in the season is the weather. Yield model developments will not bring as big a payoff in early-season performance as they will at harvest.

The LACIE results indicate that crop acreage in countries with typical field size on the order of Landsat resolution will be difficult to estimate. While improved Landsat spatial resolution can achieve better results, there are cautions to be observed in increasing resolution. As resolution increases, so do data loads and the associated processing and handling costs. Selection of resolution parameters for future satellite systems should be based on a better understanding of how resolution affects accuracy as a function of field size and spatial misregistration. Alternatives to coping with resolution-induced estimation error should also be explored, such as analysis techniques that utilize the spatial component of Landsat data. Finally, continued application of sound sampling survey practices should be employed to minimize the increased data loads inherent to increased satellite spatial resolution.

While LACIE demonstrated that the technology was successful for U.S. winter wheat and for the U.S.S.R., further developments are needed to extend the technology to wheat in other important regions. Effort is needed in more southerly regions where crop varieties, practices, and climatological factors are significantly different from those in the United States, Canada, and the U.S.S.R.; for example, India. In such regions, for instance, the effects of cloud cover on the acquisition of usable Landsat data at critical periods of the growing season need to be better quantified.

In LACIE, the lag between Landsat acquisition and reporting was typically 45 to 60 days. Much of this delay resulted from the nature of the experimental system; that is, from data backlogs, weekend and overnight time, etc. A review of the LACIE system operations indicates that there is no technical limitation to reducing this time to a few days. However, considerable expense will be involved in the time reduction. Thus, attention should be given to trading off system costs against the benefit of reducing the time between data acquisition and reporting.

FUTURE POSSIBILITIES

The experiment has provided the technology and experience base to pursue selected objectives from the following major activities.

1. Exploitation of the current technology for wheat regions such as the U.S.S.R. where it has proven able to provide better information than is currently available.
2. Additional experimentation to address the critical issues identified in the initial experiment; that is, improving the technology to monitor wheat production more accurately, more efficiently, and under a greater variety of conditions; examples of such additional experimentation include separation of wheat from barley, monitoring of small fields, and improved yield models.
3. Adaptation, application, and evaluation of technology to the monitoring of other major crops, such as corn, soybeans, rice, pasture, etc., in regions of importance.

It would appear that pursuing some combination of tasks from these categories would enhance the possibility of realizing both a short-term and a long-term gain from this technology.

APPENDIX—DESCRIPTION OF REVIEW PROCESS

This review, the last of a series of periodical peer reviews since the 1974 inception of LACIE, involved approximately 45 reviewers and was conducted over a 4-month period from March to June 1978. The review group was organized into seven teams as shown below. Each team conducted detailed reviews in specialized technical areas such as Experiment Design and Experiment Results. Teams were composed of discipline and technical specialists, some familiar with LACIE from previous reviews and some who had had no previous contact. An overview briefing was given to the peer group on March 3 to acquaint newcomers with LACIE and to update others on recent results. At this initial meeting, a peer review approach was proposed, consisting of a 2-day in-depth review of the LACIE operations in April and a 3-day review in June of all papers to be presented at the October symposium. Some members were not available for all three briefings, which created some difficulties in continuity. In addition, a better balance between time for review and time for preparation of peer review comments and documents would perhaps have increased the detail with which the groups clarified and supported their more general findings.

Following the June review, two papers were written to be presented at the symposium—this paper to be presented in the Plenary Session and a more detailed paper to be presented in the Experiment Results Session. The detailed paper consists of technical critiques of many issues not directly treated in this overview of the peer group's findings.

In general, the reviews were well organized and the informational needs of the review group were handled effectively. The reviews were intense and detailed and gave the peer group access to every aspect of LACIE. The reviewers were introduced to and had discussions with a wide range of LACIE personnel, including the project management staff, the research staff, and the operational analysts. The atmosphere was open and informal. Any aspect of the experiment could be explored in as much detail as time and energy would permit. As a result, the peer group believes that its findings represent a well-informed, independent assessment of the LACIE experience.

PEER GROUP REVIEW TEAMS

Plenary Team

D. Paarlberg, Purdue University—General Chairman

R. Balwin, Cargill, Inc.—Member at Large

L. Eisgruber, Oregon State University—Cochairman, Experiment Results Team

B. Scherr, Data Resources, Inc.—Cochairman, Experiment Results Team

H. O. Hartley, Texas A & M University—Chairman, Experiment Design Team

D. Ingram, IBM—Cochairman, System Implementation and Operations Team

J. Quirein, Schlumberger—Cochairman, System Implementation and Operations Team

D. Goodenough, Canadian Center for Remote Sensing—Chairman, Data Processing Systems Design Team

G. Nagy, University of Nebraska, Lincoln—Chairman, USDA Applications Test System Team

R. Holmes, General Motors Institute—Cochairman, Supporting Research Team

R. Shay, Oregon State University—Cochairman, Supporting Research Team

Experiment Results Team

L. Eisgruber, Oregon State University—Cochairman

B. Scherr, Data Resources, Inc.—Cochairman

F. Hall, NASA Johnson Space Center—NASA Cochairman

B. Blad, University of Nebraska

W. Coberly, University of Tulsa

A. M. Feyerherm, Kansas State University

G. Hanuschak, USDA Economics, Statistics, and Cooperatives Service, Washington, D.C.

K. Heiss, ECON, Inc.

F. Lamb, Eastern Oregon Farming Company

R. E. Selzer, Development Planning & Research Associates

Experiment Design Team

H. O. Hartley, Texas A & M University—Chairman

R. Heydorn, NASA Johnson Space Center—NASA Cochairman

J. Chromy, Research Triangle Institute

L. Guseman, Texas A & M University

D. Heerman, USDA Science and Education Administration, Fort Collins, Colorado

R. Thomas, University of California at Berkeley

L. Thompson, Iowa State College of Agriculture

System Implementation and Operations Team

D. Ingram, IBM—Cochairman

J. Quirein, Schlumberger—Cochairman

C. Johannsen, University of Missouri—Cochairman

J. Dragg, NASA Johnson Space Center— NASA Cochairman

S. DeGloria, University of California at Berkeley

D. Saxton, NOAA Environmental Data and Information Service, Washington, D.C.

J. Sos, NASA Goddard Space Flight Center

S. Wall, University of California at Berkeley

Data Processing Systems Design Team

D. Goodenough, Canadian Center for Remote Sensing—Chairman

J. Sulester, NASA Johnson Space Center—NASA Cochairman

J. Kast, Laboratory for Applications of Remote Sensing, Purdue University

T. Phillips, Laboratory for Applications of Remote Sensing, Purdue University

USDA Applications Test System Team

G. Nagy, University of Nebraska, Lincoln—Chairman

J. Murphy, USDA Foreign Agricultural Service, Houston, Texas—USDA Cochairman

D. W. Cary, Central Intelligence Agency, Washington, D.C.

H. Harkness, Sparks Commodities, Inc.

R. Head, USDA Office of Automated Data Systems, Washington, D.C.

R. Henderson, MITRE Corporation

R. LeGault, Environmental Research Institute of Michigan

R. McArdle, USDA World Food and Agricultural Outlook and Situation Board, Washington, D.C.

Supporting Research Team

R. Holmes, General Motors Institute—Cochairman

R. Shay, Oregon State University—Cochairman

J. Erickson, NASA Johnson Space Center—NASA Cochairman

W. Anderson, U.S. Geological Survey EROS Data Center

J. Estes, University of California at Santa Barbara

C. Hay, University of California at Berkeley

R. Jensen, NOAA National Weather Service, Honolulu, Hawaii

R. W. Leamer, USDA Science and Education Administration, Weslaco, Texas

B. Liska, Laboratory for Applications of Remote Sensing, Purdue University

R. Welch, NASA Ames Research Center

Detailed Results of Peer Group Review Teams at the Johnson Space Center

This section is a collection of six papers representing the summary findings of six peer review groups who reviewed LACIE from March through June 1978.

1. LACIE Results and Accuracy Group
2. LACIE Design Group
3. LACIE System Implementation and Operations Group
4. LACIE Data Systems Design Group
5. USDA Applications Test System Group
6. LACIE Supporting Research Group

As the titles of these groups suggest, each group reviewed a specialized technical area within LACIE. Each group was composed of 6 to 10 disciplinary and technical specialists, some familiar with LACIE from five previous reviews (dating from December 1974) and some who had had no previous contact with the project. Briefings by LACIE personnel to the group consisted of a 1-day overview of LACIE results and approach (March 3, 1978); 2 days of detailed briefings, discussions, and tours of LACIE facilities (April 4 and 5, 1978); and a 3-day review of all LACIE symposium presentations (June 21 to 23, 1978).

Generally, the peer groups were asked to review each technical area for technical adequacy to support the stated LACIE objectives. To achieve this aim, the groups generally addressed the following questions.

1. Was the LACIE design adequate to assess the usefulness of remote-sensing technology for global crop surveys?
2. What do the LACIE results indicate? Under what conditions will the technology perform adequately? Where does the technology need improvement?
3. What improvements to the technology are required? Is the LACIE research program adequate?
4. Can the technology be made operational?
5. Is the technology affordable in view of the value of the information produced by it?
6. What was learned about the amount of Landsat and meteorological data required to make global commodity surveys? Can the data be processed in a timely, reliable manner? Can the large volumes of data be adequately managed?
7. How can the technology best be transferred to a user community? What is the status of these efforts?
8. What was learned about systems design for future large-scale applications research?

The papers that follow represent a summary of each group's findings, as collated by the peer group chairman of each team.

Section 1

Findings of the Experiment Results Peer Group

Peer Group Members

L. Eisgruber, Oregon State University, Cochairman
B. Scherr, Data Resources, Inc., Cochairman
F. Hall, NASA Johnson Space Center, NASA Cochairman
B. Blad, University of Nebraska
W. Coberly, University of Tulsa
A. M. Feyerherm, Kansas State University
G. Hanuschak, USDA Economics, Statistics, and Cooperatives Service, Washington, D.C.
K. Heiss, ECON, Inc.
F. Lamb, Eastern Oregon Farming Company
R. E. Selzer, Development Planning & Research Associates

ASSESSMENT OF RESULTS

The LACIE results to date clearly indicate that present and future remote-sensing capabilities can be combined with, or substituted for, conventional methods of collecting information for improved crop production estimates. However, LACIE results also suggest that the current and likely future successes are not uniform throughout the growing season or from country to country or even within a country. It is important to recognize these differences and to assess LACIE results in these various phases rather than simultaneously in their entirety.

The LACIE results have been more successful for estimating wheat production in the U.S.S.R. primarily because of such factors as field size and climatological conditions, which influence management practices and alternatives. Information on wheat production in the U.S.S.R. from traditional sources has been and remains poor, particularly from the standpoint of timeliness. However, the U.S.S.R. is periodically a large purchaser of U.S. grain and has a significant impact on domestic and international prices. This, then, can only lead to the conclusion that LACIE results provide useful and much needed information not now available from other sources. In this area, LACIE is practical and useful right now.

The United States and Canada were included as LACIE test countries because the good baseline information available in these countries, as well as the accessibility to ground-observed data, permitted a detailed understanding of the effect of various agricultural and climatological conditions on LACIE performance. The primary implications of the U.S. and Canadian results is in regard to LACIE performance in foreign regions with similar characteristics. LACIE results in these two countries indicate that the Landsat and yield technology performed well in the U.S. winter wheat regions with fairly large field sizes and for a reasonable range of weather. However, the Landsat technology did not perform as well in the northern U.S. and Canadian spring wheat regions because of small fields, strip-fallow practice, and difficulties in separating confusion crops. In addition, the yield models performed poorly for extreme durations in the weather. The LACIE tests indicate that the technology would be marginal to unsatisfactory for 90/90 estimates in the United States and Canada.

This assessment of the usefulness of the LACIE results for regions similar to the northern United States and Canada is probably unduly critical for the following reasons. First, the LACIE 90/90 criterion was applied on the basis of the assumption that the standard benchmark estimates for U.S. wheat production are without error; however, this is not so. There is limited evidence that, if the error of estimate of the benchmark is considered, the benchmark estimate and the LACIE estimate cannot be shown to be statistically significantly different. Second, improved resolution of satellite imagery is expected to improve area estimates, particularly when small fields are involved. Third, remote-sensing technology can provide information that is not considered at all in the 90/90 criterion and that is not available from other sources. An example of such information is the various season estimates of the greening index (permitting conclusions about moisture stress and biostage information). Finally, a system designed for operational purposes can provide wheat production estimates more quickly than traditional methods. Under such circumstances, accuracy can be traded off for timeliness. In general, at-harvest estimates are much more accurate (relative to the benchmark data) than early-season results. Thus, the LACIE technology is promising even for regions similar to the northern United States, although the usefulness is not as immediate as it is to the U.S.S.R.

Although the original list of LACIE countries included countries such as the People's Republic of China, India, Australia, and Argentina, LACIE did not generate production estimates for these countries. Given the magnitude of the task and the available resources, the peer groups view the original goal of providing production estimates for all these countries as having been too ambitious.

During LACIE, the time lag between Landsat acquisition and reporting results was typically 45 to 60 days. Much of this delay resulted from the nature of the experimental system; that is, from data backlogs, weekend and overnight time, etc. Exploration of this topic with the LACIE staff has convinced the peer group that, without question, the technological and systems expertise exists to design an operational system that would meet a 14-day turnaround or even exceed it. However, considerable expense could be involved in reducing this time lag. Thus, careful attention should be given to trading off systems cost with the benefit of reducing the time between data acquisition and reporting.

The question naturally arises of the transferability of the LACIE results from wheat in certain countries to wheat in other countries or to other crops (corn, rice, forage, soybeans). Despite specific examples of how the technology could be transferred with good success, LACIE does not provide sufficient objective evidence (nor did it plan to do so) to permit a full assessment of this issue.

ASSESSMENT OF SOME PROCESSES AFFECTING THE RESULTS

Yield models, which are used with acreage estimates to arrive at production estimates, were an important component in LACIE. However, the performance of the yield models raises significant questions. Given the overall performance of the U.S.S.R. model, a reasonable evaluation would be that yield models for that region of the world are adequate and useful, although a replacement of these models with advanced models that are more sensitive to weather variations is expected to result in improved estimates. For the United States and Canada, the current stage of development would have to be assessed as being marginally adequate to inadequate. There is definitely a need to improve these models.

The yield models, used operationally in LACIE, work reasonably well when weather conditions do not deviate greatly from normal or average conditions. Since a major share of the success in making production estimates rests with making adequate predictions of yields, it is believed that a major effort should be made to develop models that do a more adequate job of yield estimation. Increased model sensitivity must be developed through (1) a more precise quantification of the relationship of grain yields to soil moisture, temperature, solar radiation, cultural practices, nutrient application, and soil characteristics; and (2) a denser network of meteorological observations than is presently available on a global basis. Some research has been conducted on improving relationship sensitivities; more is highly recommended. More research to achieve a denser network of meteorological observations is recommended through the use of satellites to provide data on soil moisture, biomass, solar radiation, precipitation, surface temperature, and snow cover.

An evaluation of the use of crop calendar models provides mixed results. The good LACIE results with respect to winter wheat and U.S.S.R. spring wheat would lead one to the conclusion that crop calendar models were adequate. Similarly, analysts clearly were of the opinion that crop calendar models were of substantial value to them. However, available statistical analyses about the impact of the use or nonuse of crop calendar models on the accuracy of classification were inconclusive. The overall evaluation is that improvement in crop calendar models for wheat is still needed.

Crop calendar models are needed for all major crops, and better starter models are needed. If crop calendar models are to provide the type of information that an analyst needs to label spectral classes for automated processing or to define decision algorithms for partially automated labeling algorithms, then the necessary correlation of crop biostage to spectral response needs to be addressed. The possibility of developing spectral crop calendar models should be explored, since the temporal-spectral response characteristics are the physical characteristics being used to identify the crops. With the current approach, all analysts must now develop their own concept of a spectral crop calendar based on past experience and "traditional" crop calendar information. Greater accuracy and less variability could be achieved by providing direct spectral crop calendar information to the analysts rather than forcing the analysts to develop their own subjective correlations based on limited data.

For estimation of wheat production, the issue of identification of confusion crops is one that needs to be resolved in certain areas of the world. Although the peer group has no specific suggestions on how this problem might be resolved, it recommends continued research to bring a satisfactory solution to the problem.

Current LACIE results indicate that production estimates are more accurate at harvest. However, one of the major advantages of a program like LACIE is its ability to make more accurate preharvest estimates of crop production than are currently available. Despite difficulties in making long-range weather forecasts, it appears desirable to expend some resources and efforts toward providing timely preharvest estimates. While it may be unrealistic to suggest that preharvest estimates be made more than 2 months before harvest time, reasonably accurate estimates made 1 to 2 months before harvest would appear to be feasible.

ASSESSMENT OF RELATED ACTIVITIES

Results of attempts to evaluate the economic importance of LACIE technology seem to be of limited usefulness to date, primarily because of the inability to develop good estimates of the value of more accurate commodity forecast information. In addition, it seems that the economic evaluations that were conducted did not reflect an in-depth understanding of the current LACIE methodology and results nor of the potential results of an improved future technology.

The LACIE technology components are being transferred partially or in full to the USDA Applications Test System (ATS); however, in response to recently shifting USDA priorities, LACIE technology is being utilized by the ATS effort primarily to detect and assess unusual crop conditions abroad rather than the quantitative commodity production forecast application evaluated by LACIE. While the ATS has demonstrated its ability to produce a complete and timely assessment for a restricted foreign region, it is not known how useful this assessment has been. There has as yet been no feedback from the Foreign Agricultural Service, the ATS's putative major client. Thus, while the time is proper to review LACIE accomplishments, a review of the ATS effort at this time is perhaps premature.

GENERAL ASSESSMENT AND CONCLUDING COMMENTS

Despite the obvious potential for additional research and improvement, LACIE must be considered a success. An assessment of the LACIE results leads the peer groups to conclude that for global wheat regions such as the U.S.S.R., the LACIE technology can be made operational and that for regions where the technology requires improvement (such as small-fields regions), continuing funding for further research and development should be pursued.

Interagency cooperation is one of the important factors that made LACIE a success; therefore, the peer groups recommend the continuation of fostering such cooperation on a long-term basis. Lack of such cooperation is likely to result in reduced funding, increased instability, and a lower quality program. The peer groups question whether any one agency has the capability to operate and develop a first-rate program and, more specifically, whether the U.S. Department of Agriculture (USDA) would have the capability to operate and develop the necessary technology or whether it should develop such a capability. Conversely, the peer groups question whether the National Aeronautics and Space Administration (NASA) or the National Oceanic and Atmospheric Administration (NOAA) would have the necessary agricultural (physical, biological, economic, and institutional) background to operate the program. Thus, proliferation would seem to be detrimental to the program.

From the standpoint of usefulness of results, two important questions will have to be addressed in the very near future: who will have access to information generated by a LACIE-type system, and when and how will such access be permitted? One model would be that raw data would be made available to everyone for analysis and interpretation. In addition, USDA and other governmental agencies should provide analysis and interpretation for in-house use and for dissemination to the public. However, this is only one approach; and even with this one approach, many specific issues need to be resolved. This issue must be addressed explicitly as part of future LACIE-type efforts.

Section 2

Findings of the Experiment Design Peer Group

Peer Group Members

H. O. Hartley, Texas A & M University, Chairman
R. Heydorn, NASA Johnson Space Center, NASA Cochairman
J. Chromy, Research Triangle Institute
L. Guseman, Texas A & M University
D. Heerman, USDA Science and Education Administration, Fort Collins, Colorado
R. Thomas, University of California at Berkeley
L. Thompson, Iowa State College of Agriculture

INTRODUCTION

This group reviewed the design of procedures in LACIE, which covers the sampling design for the acreage subsystem, the Classification and Mensuration Subsystem (CAMS) design, the yield model design, and the methodology for accuracy assessment. The first-generation LACIE yield subsystem relies almost exclusively on yield models developed from historical data banks provided by existing information systems such as the NOAA weather stations and from historical survey designs such as the USDA Statistical Reporting Service (SRS) county statistics for wheat yields. This discussion, therefore, is confined to an assessment of the statistical methodology used in yield model construction and testing and to the estimation of wheat production based on the sampling and CAMS designs and on the assessment of the statistical methodology. This assessment does not cover the methodology adopted in growth stage and crop calendar research and development work.

SAMPLING DESIGN FOR THE ACREAGE SUBSYSTEM

The first question that may be raised is why sampling? If LACIE must develop a computerized technology to convert tape-recorded channel measurements of reflected light over a unit area into estimation of wheat acreages, why not use this technology 100 percent? LACIE decided quite early and correctly that sampling of areas was desirable, if not essential. It became apparent that this conversion of the channel measurements to wheat acreage estimates could not be accomplished by an automatic computerized procedure but had to be done with the participation of human intelligence (photograph interpretation by analyst-interpreters). The time-cost element of this participation had to be assessed against the efficiency of LACIE sampling techniques. It was found that the sampling error (approximately 2 percent) resulting from quite moderate sampling fractions (again approximately 2 percent) was comparable if not smaller than the percentage error resulting from measurements, the so-called classification error. Cost-effectiveness considerations therefore dictated a sampling strategy.

Sampling Frame

The basic sampling frame consists of elementary units called pixels (areas of about 1 acre) with channel measurements (i.e., the reflected light intensities associated with them). These measurements are recorded on the basic Landsat tape that is to be sampled.

Sampling Units

For various reasons, it is impractical to use the astronomical number of pixels as sampling units. Instead, LACIE decided to use an area unit, the segment, and to record the channel measurements for all pixels within the area unit as the sample information. This should be regarded as a cluster sample of pixels. The size of the area unit is 5 by 6 nautical miles. It may be argued that this unit is too large from the standpoint of sampling efficiency (it contains approximately 23 000 pixels). The size of this unit may not be optimum; however, the following practical considerations dictated the use of a unit of at least a comparable size.

1. It was necessary to register the acquisition of data from segments acquired during the various passages of the satellite over the same segment. The technology of identifying the same segment in these various passages requires key points within the segment that are easily recognizable and, in turn, this requires a segment of an adequate size.

2. Again, the Landsat imagery and its interpretation by the analysts, as well as the computation of signatures custom-made for the segment, requires an adequate size, as does Procedure 1 (see subsequent discussion).

3. LACIE addressed the problem of how much the variance of the statistical sample could be reduced by using areas of smaller size; the gains did not justify changing from the above segment size to a much smaller area in view of the aforementioned and other practical limitations. The peer group, therefore, has no criticism of the choice of the segment size.

Sampling Design

The sampling design is a multistage stratified design. In the U.S. LACIE design, strata are either counties or groups of counties. Historical agricultural-census wheat acreages are used to classify the counties into three groups: Group I, high wheat density; Group II, moderate wheat density; and Group III, trace wheat density. For Group I, a precomputed number of sample segments is drawn. For Group II, the design is in two stages. The counties are primary sampling units which are drawn with probability proportional to their historical wheat acreages, and the segments are secondary units. One segment per sampled county is drawn. For Group III, no sample segments are drawn, but the total wheat acreage is estimated through a ratio estimator using historical agricultural-census wheat acreages. The possible bias of the estimate in Group III has been addressed and has been found to be outweighed by the reduction in variance through the design efficiency, avoiding a wastage of sample segments to measure trace wheat acreages. The allocation of the segments follows the principle of optimum allocation and is again based on historical wheat acreages.

Subsample of Ground-Truth Segments

The subsample of ground-truth segments thus far has been used only for special studies to control the quality of the LACIE estimates of segment wheat acreages. One reason for this is that ground-truth information is obviously not available in remote areas of the LACIE countries. Ground-truth segments, in providing independent information on the true wheat acreages of segments, may be regarded in sampling terminology as a "special record check." The use of these data for quality control agrees with the practices adopted by Federal agencies in socioeconomic surveys. Rarely, if ever, are such special records used to adjust the survey estimates. However, the accumulation of the ground-truth data should make it feasible to use them for bias estimation and bias elimination. This problem is addressed below.

Strong Points

The strong points of the survey design are its effective utilization of historical data and its flexibility. As already indicated, the survey design is efficient in the reduction of the sampling variance of the acreage estimates. LACIE has addressed the problem of improving the stratification based on historical areas, particularly since these areas are very large in foreign LACIE countries, such as the U.S.S.R. and the People's Republic of China. The most effective way of improving the stratification is by intersecting the historical areas with agriclimatological areas, based on both historical information (such as soil-type maps and climatological isobars) and Landsat imagery. Although such substrata are more likely to be homogeneous with regard to wheat density, historical information on the latter will not be available and must be specially estimated.

Because of the Group III (areas of low wheat density) features and the associated ratio estimator, the design is flexible to cope with segment loss through cloud cover and other contingencies. However, special care has to be exercised by LACIE to control possible biases arising from the transfer of strata which have lost all their segment information to Group III.

Weak Points and Suggested Remedies

The weak points of the design concern bias, nonsampling variances, and measurement bias. More work is needed to assess the possible bias through cloud cover and other courses of nonresponse. In the LACIE estimation procedure, it is assumed that a loss of a segment through cloud cover is not correlated with the segment's wheat density within a stratum. This means that if a segment within a county is lost through cloud cover, this segment may have a wheat density that may be above or below the county average with roughly equal probability. While such an assumption has been found to be fairly well satisfied for small areas like counties, there is considerable doubt whether it would also be satisfied for areas the size of an oblast. Indeed, LACIE has found in special studies that there are correlations between cloud cover and wheat density in larger strata. This problem again emphasizes the desirability of reducing the strata size in foreign LACIE countries, even if the historical data for optimum allocation are not very reliable.

Other contingencies that may lead to bias are the treatment of boundary segments in contingent strata and the pseudorandom selection of segments to avoid contingent sample segments in general. More work is needed to examine the possibility of biases from these practices.

The variance formulas for wheat acreage do not include the correlated nonsampling errors. For example, if certain analysts have a tendency to underestimate the wheat acreage of segments and others have a lesser tendency to do so, their idiosyncrasies would increase the true variance of the wheat acreage estimate. Attempts should be made to use the recently developed methodology of estimating nonsampling variances from survey data into the LACIE variance formulas.

The LACIE has found that there is definitely a measurement bias which leads to an underestimate of the wheat acreage of certain areas such as a state or a country. Attempts have been made to estimate such biases by comparing LACIE segment estimates with those obtained by ground-truth measurements. Unfortunately, the ground-truth data bank is limited at the present time; consequently, a direct estimate of the bias has a large variance and cannot be used for an adjustment of LACIE wheat acreage estimates. It is suggested that the ground-truth subsample be redesigned to reduce such a variance. Moreover, the ground-truth data bank could be utilized to reduce the bias at the expense of a possible increase in variance by a method known as Procedure 1.

Anticipated Future Developments

Transfer of bias estimation to foreign countries.—The bias estimation referred to in the preceding subsection is, of course, only available for the United States. A transfer of this bias estimation to foreign countries would have to be based on a model of Landsat predictors. LACIE has made various attempts, most of them unsuccessful, to establish regression models. A more hopeful procedure is Procedure 1.

Multiyear designs.—At the present time, LACIE estimates of wheat acreages are "one-shot" estimates in that they are based on the Landsat data for the current year only. An attempt should be made to utilize the Landsat data of preceding years. LACIE has developed a methodology for accomplishing this, and a paper on this topic will be included in the symposium proceedings. Such a procedure obviously would require the rotation of segments; i.e., the retention of a fraction of the segments each year and the replacement of a fraction by new segments. The implementation of such a technique should result in an improvement of wheat acreage estimates for a region for the current year. It would also provide the opportunity to monitor gross classification errors of individual segments by relating the segment to acquisitions of previous year(s) using a "year by segment within strata" analysis of variance model.

Quality control of analyst labeling procedure.—As the ground-truth data bank is increased, it may be possible to establish quality control charts to monitor individual analyst labeling errors. This would enable LACIE to introduce remedial action for certain analysts by conducting special briefing sessions.

CLASSIFICATION AND MENSURATION SUBSYSTEM

Initial Approach

The initial CAMS design was a single-system approach that incorporated available machine classification techniques in modular form and used manual interpretation techniques for classifier training. As new machine-processing techniques were developed, they could be tested for possible incorporation into the prototype system. In addition, new manual interpretation techniques and better defined ancillary information could be incorporated.

The initial machine-processing design required the classification of each pixel into wheat or nonwheat categories. The proportion of wheat in the segment was then defined to be the number of pixels classified as wheat divided by the number of identifiable pixels in the segment. For the machine-processing approach to work, manual analysis of Landsat and ancillary data was required to select and label sample fields of wheat and nonwheat. The sample fields constituted the set of training data mentioned previously. The initial machine classification model assumed multivariate normal classes and used a Bayesian classifier.

By the end of Phase II, adjustments to this approach were required. The biggest problem in estimating wheat production in the United States occurred in spring wheat areas. A relative difference of −22 percent was noted between LACIE estimates and SRS estimates, and most of the errors could be traced to the acreage estimate. The initial classification and mensuration approach required the analyst to determine the number of statistically separable spectral classes in a LACIE segment, select sample fields representing each such class, label each field as wheat or small grains versus "other," and evaluate the classification results.

A key problem with this approach was that, in general, attempts to do multitemporal classifications failed. This was a basic problem because the information for discriminating crop types with Landsat data was in the relationship of the canopy spectral-reflectance changes over time and its crop calendar (assuming that two different crop types will have different crop calendars). A major contributor to this problem was the failure of the clustering algorithm which was required to provide the analyst a good sample of data with which to train the classifier. In lieu of this, the analyst was used to determine the clusters, or spectral structure of the data, by simply examining color imagery. The subtleties of the sampling required to obtain good estimates for classification parameters were not always apparent in the color image. This problem became even more acute when training for a multitemporal classification was required. The number of unique spectral classes increased significantly. Moreover, greater spectral variability required even larger sample sizes, often beyond those reasonably obtainable within prescribed time limits.

Another problem with the approach was that there was no way to easily detect and correct machine-processing errors. First, the "tests" for adequacy of classification performance applied by analysts to classified segments tended to be somewhat subjective. These evaluation procedures included a visual assessment of the wheat/nonwheat classification map and the percentage of correct classification for training fields and for a very small sample of independent test fields. A more statistically reliable procedure for analyst evaluation of classification results was desirable. Second, once a machine-processing run was judged to be inadequate, no provision was made for allowing the analyst to correct (in a statistically valid manner) the machine wheat estimate without recycling through the entire machine-processing sequence. Field labeling by the analyst was hampered by inadequate information on the growth cycle and cropping practices applied to crops that could be spectrally confused with wheat. This was particularly true in Phase I. In addition, techniques and opportunities for tracking the digital reflectance values through time for training fields were very limited. This spectral aid information can be valuable in the labeling process.

Analyst contact with machine-processing functions in the initial system was relatively inefficient. Of particular note were the lengthy periods required for analyst-supervised digitization of training field boundaries, limited opportunities and capabilities for direct interrogation of the Landsat digital data base using near-real-time interactive systems, and lengthy periods between submission of training data and return of classification results for evaluation. The digitization procedure significantly added to analyst time. The long turnaround time to obtain classification results would have made interactive definition of training statistics problems more difficult for the analyst.

Procedure 1—The Second-Generation Approach

The Procedure 1 design was an attempt to correct some of the deficiencies of the initial design. Several basic modifications resulted. First, the analyst was no longer required to select a sample of fields for classifier training. Instead, he was provided with a random sample of dots and was required to label them as being small grains or non-small-grains. Next, research efforts had provided a working clustering algorithm that was added as a processing step before classification to automatically estimate the subclass structure in the data required for the classification process. Analyst-labeled dots were used to label the clusters automatically, and, in turn, all pixels in a cluster were used to estimate classification parameters. Finally, Procedure 1 provided a means for quantitatively evaluating and treating classification errors other than simply reworking the segment until a satisfactory answer emerged, as was done in the first design. In Procedure 1, the classification result was treated as a stratification of the segment into potential small grains and non-small-grains areas. A second set of dots was then labeled and used to obtain a stratified areal estimate of the small-grains proportion in the segment. In effect, a two-stage sampling process is done by the analyst. One stage is used to produce a stratification of the area, and the next stage is used to effect the final estimate.

The Procedure 1 design provided more accurate answers in a way that made more efficient use of the analyst, when compared to the first design. In addition, multitemporal machine processing was now being done routinely. Overall, the design appears to be an improvement over the Phase I and II design in the following areas.

1. Training statistics definition: Clustering was used to define (spectrally) the structure of subclasses in the data required for classification. This had been a time-consuming and error-prone task for the analyst. Moreover, all pixels in the segment could be used to estimate such class mean vectors and covariance matrices, thus reducing the sampling size problem also present in the initial design.

2. Analyst labeling: The accuracy of the Procedure 1 estimate depends to a large extent on accurate, or at least unbiased, analyst labeling of the dots. Procedure 1 does provide the analyst with certain products, such as trajectory and scatter plots, which attempt to summarize the numerical data. The trajectory plots provide a summary over time of the spectral movement of a pixel. The scatter plots provide a summary of the spectral proximity of each dot for a given acquisition. These products have been helpful as a labeling aid. In addition, more information concerning confusion crops was available for Procedure 1 than was available in the initial LACIE design.

3. Stratified areal estimate: Using the machine classification as a segment stratification into wheat/small grains versus "other" enabled the analyst to introduce information on machine bias using a ratio estimation technique. Unfortunately, machine classification error is high enough that the efficiency benefits gained through the use of such a stratification appear to be marginal. It is more efficient than using only the Type 2 labeled dots for the estimator; however, when compared to an estimator formed by using both Type 1 and Type 2 dots, there may not always be a gain in efficiency.

Key Technical Issues

The key technical issues regarding Procedure 1 are as follows.

1. To some extent, the Procedure 1 design has attempted to improve analyst performance. However, such errors are still a dominant factor in the bias and variance of the estimates; therefore, designs which are more tolerant of some analyst error should be sought. In particular, machine-oriented procedures that smooth analyst errors should be considered.

2. There is also a question of efficiency in Procedure 1. In particular, how does the variance of a simple random-sample estimate by the analyst compare to the variance of the estimates from Procedure 1?

3. A sample of 40 dots (Type 1) may be insufficient to label clusters, particularly when wheat proportion is low or when a significant proportion of confusion crops may be present. Errors in combining clusters to produce categories of small grains and non-small-grains could become significant in these instances. An alternative would be to group all clusters into probability of wheat strata based on

some spectral index (preferably multitemporal) and then perform only one dot sample (Type 2) within an early stratum to determine the proportion of wheat.

4. Procedure 1 provides for a more efficient interface between the analyst and machine processing and machine turnaround times are lower; however, improvements can still be made. In particular, real-time availability of scatter plots, trajectory plots, and other spectral aids would be valuable to the analysts. At present, this plot information is available only for the previous Landsat pass for a given segment, not for the most recent one. In general, the opportunity for real-time integration and analysis of Landsat digital data will tend to improve labeling performance.

5. Boundary dots are not labeled in the Type 1 labeling process; therefore, this extra labeling requirement would have to be imposed before such a comparison is made. If the concept in Procedure 1 is to be pursued further, designs in which Type 1 labeling is done in only a small fraction of segments and all segments are classified (i.e., a signature extension concept) need to be considered.

6. Procedure 1 makes no use of the spatial information available in Landsat data. Techniques which could prove useful as replacements for existing clustering and classification techniques are currently available for testing within the Procedure 1 framework. In addition, segment field structures based on multitemporal data can be provided to the analyst as labeling aids.

7. It may be appropriate to consider the use of ground data in the U.S. Great Plains region to bias-correct analyst labels. Designs incorporating this feature have already been discussed. As an alternative, within-segment measurement error strata might be constructed. Expected analyst bias correction factors could then be applied by stratum after analyst labeling.

8. In some instances, manual dot or area analysis rather than a man-machine combination may provide lower variance, lower bias, and/or lower cost wheat proportion estimates. A set of small experiments should be designed to evaluate the circumstances under which given processing combinations provide the real-time estimates least subject to variance and bias constraints.

9. Procedure 1 is still experiencing classification biases that usually lead to underestimates of the wheat acreage of an area, particularly for early-season estimates. It is believed that much of this bias arises from areas in the Landsat color photographs for which the analysts are uncertain as to the identity of the crop. The following is a very brief sketch of a modification of Procedure 1 which may be helpful in eliminating bias due to labeling errors.

The analysts should be permitted to delineate these areas of uncertainty on each color photograph as "doubtful areas" of various "types" and exclude these from Procedure 1. The processing by the analysts of Landsat segments for which ground truth is available (for both previous-year and current-year acquisitions) would then produce a ground-truth data bank of actual wheat proportions within "doubtful areas" of all the types, leading to "expected wheat proportions" for such doubtful areas. The aggregation of such a ground-truth bank would eventually lead to expected wheat proportions of sufficient reliability for operational use. This would mean that, in processing operational segments, the analyst would merely delineate "doubtful areas" of various "types" and the computer program would convert the delineated areas into total acreages and multiply them by the expected wheat proportions from the ground-truth data bank. No claim could be made for estimating the wheat acreages of individual segments. However, with a sufficiently detailed classification into types of doubtful areas, this method would lead to essentially unbiased estimates of wheat acreages for an area such as a state. The transfer of this technology (if successful) to foreign countries would undoubtedly be difficult and is not discussed here.

ASSESSMENT OF THE STATISTICAL METHODOLOGY USED IN YIELD CONSTRUCTION AND TESTING

NOAA Yield Model

The data banks used for the building of the NOAA yield model were the USDA SRS county statistics program for wheat yields covering the years from 1932 to 1972 and the records of precipitation and temperature from NOAA weather stations for the same time range. More specifically, the following predictors of wheat yield were selected using the advice of wheat experts associated with NOAA. The predictors are those for Kansas; slight modifications were used in other states.

1. Overall constant

2. Linear technology trends

3. Precipitation (mm) from August to February, departure from normal

4. March precipitation (mm) minus potential evapotranspiration (PET) computed from monthly temperature, departure from normal

5. March precipitation (mm) minus PET computed from monthly temperature, squared departure from normal

6. May precipitation (mm), squared departure from normal

7. May degree days above 90° F

8. June precipitation (mm), departure from normal

9. June precipitation (mm), squared departure from normal

It was realized from the outset that the selection of these weather summaries represents a possibly gross oversimplification of the plant growth physiology for wheat. Indeed, it may be argued that NOAA records are inadequate for construction of a scientific plant growth model. However, it must be remembered that, for a model to be useful, the predictors involved in it must all be readily available on a real-time basis and for a representative sample of wheat locations covering the total area of, for example, a state. Therefore, very detailed measurements at the weather stations (such as daily records of precipitation and evaporation) do not necessarily apply to the wheatfields for which the USDA SRS sample of wheat operators report. Moreover, the more detailed meteorological measurements will not be available for some of the weather stations. It was believed (probably rightly so) that it is more important to have a better coverage of the prediction area for the basic meteorological measurements than to have more detailed measurements from a very small sample of weather stations.

It would clearly be a folly to attempt to predict wheat yield over a span of 30 years or so from meteorological variables alone. Obviously, during this timespan, wheat technology (involving increase in fertilization, improvement of varieties, and managerial factors) will have increased the average wheat yield. This model represented wheat technology trend by a broken linear trend having a small rate of increase in the initial period, a more rapid rate of increase in the second period, and a somewhat reduced rate of increase in the final period. The breakpoints in this technology trend were related to stages in the historical wheat technology development by the wheat expert associated with NOAA.

Strong Points of the NOAA Yield Model

The NOAA model was compiled in a relatively short time to meet real-time operational needs. To appreciate this achievement, one has to remember that the compilation and collation of data supplied by two different Federal agencies is by no means a simple task.

The model is simple. This feature is particularly important when it is realized that the LACIE predictions require an estimation of the yield per acre on a real-time basis. A more complicated model can, of course, be readily covered by an appropriate computer program. However, the key point about simplicity is that the data acquisition required for the use of a model is likely to be very difficult if the input requires complicated predictors that are often unavailable.

The yield model has been incorporated in the 10-year test designed to validate the 90/90 criterion. This test, of course, only validates whether or not the model is adequate to satisfy this criterion.

Weak Points and Suggested Remedies

The previously mentioned technology trend was assessed by "eyeballing" the data to which it was supposed to be fitted. Such a procedure does not permit the computation of statistical errors in the estimated coefficients involved in the law. There should be no difficulty in replacing the fitting of such a broken linear trend by a genuine "spline regression" technique. The resulting law would still be a linear regression law, provided that "breakpoints" in the technology are predetermined by wheat experts.

All the meteorological predictors were entered in the model as linear functions, and, as such, the model is not sensitive to weather effects that can be considered "extreme."

The estimated variance for the wheat predictions is valid only for predicting a current-year USDA SRS estimate of wheat yield. This would tally with the LACIE objectives only if it assumed that the USDA SRS survey estimates of wheat yield have an error that is negligible. LACIE has already developed an alternative method of yield error prediction. To satisfy the objectives of LACIE, it is necessary to predict the average wheat yields for all fields in a state. Therefore, the estimated variances of wheat predictions should account for the sampling error introduced by the fact that the NOAA stations can only be regarded as a sample of locations over the state. Computations indicate that the yield prediction error from this method may indeed be appreciably larger than that estimated by NOAA unless the NOAA weather stations supplying the input information represent a stratified sample over the state. Care should thus be taken to compute the input predictors as stratified averages using at least one weather station per county.

ESTIMATION OF WHEAT PRODUCTION

Production as a Product Estimate of Acreage and Yield

The first-generation LACIE technology does not provide for direct estimation of wheat production but rather computes wheat production as the product of acreage and yield. The two factors in this equation are estimated by their separate LACIE subsystem technologies. These two subsystems are not entirely independent, although at the present time the acreage subsystem uses Landsat data almost exclusively and the yield-per-acre subsystem uses only historical data. However, there are occasional overlaps in the data used. For example, the crop calendar information used by the analysts to interpret Landsat imagery for the early-season acquisitions is highly correlated with the early-season rainfalls used in yield prediction. However, at the present time, the design of LACIE procedures is completely covered by the design of the two subsystems discussed.

Implementation of Statistical Estimation Procedures for Production

The statistical estimation procedures for estimating the product of acreage and yield have been competently implemented and have led to an approximate assessment of the 90/90 criterion in which biases are allowed for. Since acreage and yield estimates are based on essentially separate data banks, they have been rightly regarded as essentially independent in the derivation of variance formulas for the product of the two factors. However, in later generations of LACIE technology, the dependence of the estimates may have to be accounted for.

Section 3

Findings of the System Implementation and Operations Peer Group

Peer Group Members

D. Ingram, IBM, Cochairman
J. Quirein, Schlumberger, Cochairman
C. Johannsen, University of Missouri, Cochairman
J. Dragg, NASA Johnson Space Center, NASA Cochairman
S. DeGloria, University of California at Berkeley
D. Saxton, NOAA Environmental Data and Information Service, Washington, D.C.
J. Sos, NASA Goddard Space Flight Center
S. Wall, University of California at Berkeley

ASSESSMENT OF SYSTEM IMPLEMENTATION AND OPERATIONS

The summary findings of the LACIE System Implementation and Operations Group are as follows.

The Applications Evaluation System (AES) provided successful integration of remote-sensing and data-processing technology into a large-scale crop inventory system, especially under the following constraints.

1. Use of existing capability
2. Schedule and funding
3. Continuing changes in LACIE design
4. Without ground truth
5. Limited historical information

The AES demonstrated that the technology will support production estimation for selected wheat-growing regions before harvest and established a base of interagency, multidisciplinary, experienced personnel for additional remote-sensing agricultural applications research and development. System design and operational experience will provide for the development and implementation of future operational systems such as a first-generation USDA hardware/software system. The AES identified and brought into focus key technical issues such as better classification procedures, crop calendars, and data registration for remote-sensing crop inventory. It provided a stimulus for improved access, quantity, and quality of Landsat and meteorological data. It provided first-time testing of technology over a sufficiently wide range of agronomic and climatological conditions to evaluate the technology and first-time evaluation of technology under realistic foreign situations without ground truth or historical crop information.

Based on these findings, the peer group makes the following recommendations.

1. In a large-scale applications project like LACIE, the analyses should be supported by project-dedicated systems to achieve timeliness and efficiency.
2. System design should be flexible and interactive for incorporating changes and improvements.
3. Such systems will ensure adequate development and utilization of global data bases (Landsat, meteorology, crops, soils, etc.).
4. Crop production information should be released to the public sector consistent with commodity laws.
5. Techniques for research-operations technology transfer internal and external to the project should be improved.
6. Maintenance and improvement of the resource base developed within LACIE for future resource inventory systems should be continued.
7. It should be ensured that the remote-sensing research community has access to timely Landsat data.

The sections that follow summarize the findings for each of the AES subsystems.

DATA ACQUISITION, PREPROCESSING, AND HANDLING

A key element in any operational resource inventory process is data management. Before LACIE, inventories aided by remote sensing had yet to address data management in a large-scale operational context. Data base management in LACIE operations consisted of three major components: data acquisition; data preprocessing; and data handling, storage, and retrieval. The primary objective of this data management system was to ensure availability, consistency, and timeliness of the data required for the global-level productivity estimation process of LACIE.

Data acquisition and preprocessing operations were conducted by the NASA Goddard Space Flight Center (GSFC). During the 3-year operational effort, evolution of these components increased project throughput, provided additional analysis data, and improved the operational efficiency of LACIE. Major advancements that enhanced project operations resulted from (1) improved digital processing that allowed automatic cloud screening and shadow detection, (2) streamlining of the interface and integration of altitude and orbital data, (3) removal of the time-consuming video-terminal screening process, and (4) consistent extraction and registration of multitemporal sample segments from full-frame Landsat digital data. Experience gained in using foreign ground-receiving stations, direct data transmission to the NASA Johnson Space Center (JSC) via land lines, and quality assurance film library and processing procedures also contributed significantly to the improvement of LACIE operations.

Data handling, storage, and retrieval operations were conducted at JSC with responsibilities centered on ensuring operational data flow through the analysis process, determining status and tracking of data through the analysis system, and archiving and retrieving data as required by the analysts. These data-handling responsibilities were conducted via a multicomputer system configuration that dictated the efficiency of the operational data flow within LACIE. Though some compatibility problems existed, they were minimized by using a consistent data format for passing estimates and other data to other subsystems.

As a result of the operational data management experience of LACIE, the following major recommendations are made.

1. Expand, improve, and extend the quality and availability of registered, multitemporal, full-frame satellite data sets to the remote-sensing and user communities.
2. Improve the availability and timeliness of full-frame imagery and computer-compatible tape (CCT) delivery to the research and user communities.
3. Explore the use of direct (ground) data-transmission lines to state- or regional-level data centers to improve the timeliness of data acquisition.
4. Provide geometrically corrected data as a standard product.

CROP ASSESSMENT

Crop assessment in the LACIE context is a system used for the generation of wheat production estimates by combining data from three other subsystems that provide wheat area, wheat yield, and historical and ancillary data for the level being aggregated. The system was designed, developed, and implemented at JSC as an interactive system to accomplish the following in an operational and highly flexible manner.

1. Expand sample segment proportion estimates to any desired level of the aggregation hierarchy.
2. Derive a weighted yield estimate.
3. Combine area and yield estimates to generate the production estimate.
4. Compute statistics associated with estimates.
5. Routinely generate comprehensive crop reports as required.

Because of its interactive nature, the system is highly flexible, allowing input of sample segments of varying size, system improvements as technology develops, multiple crop analysis, and implementation by operational and research communities.

Three major components of the interactive system are data management, aggregation, and report processing. The core of the system is a unique data aggregation software system that combines area and yield estimates in light of historical and ancillary data to generate a production estimate and associated statistical descriptors for evaluating the accuracy and reliability of the estimates derived. For each estimate, descriptors include standard error, coefficient of variation, probability of 10-percent error, and 90-percent confidence limits. The other major component is a unique reporting system that publishes the estimates in tabular format as generated by the aggregation system. The reports also contain crop condition assessments, yield analyses, and discussions of segment data in narrative format.

The system evolved and expanded over the 3-year period of LACIE from a batch to an interactive system with significant improvements noted in enhanced data entry, increased statistical descriptors, simultaneous operation of analyst stations, increased aggregation hierarchy output, increased timeliness in report generation, and integration of multicrop aggregation design concepts. Based on 3 years of operations, the Crop Assessment System (CAS) (1) has identified those elements critical for achieving routine crop reports regardless of operational environment; (2) has recognized the improvements required to reduce report generation time following changing crop conditions; and (3) has noted that system software should be designed to accommodate advances in remote-sensing processing technology.

As a result of the operational CAS of LACIE, the following major recommendations are made.

1. Expand, improve, and generalize the aggregation system software to allow for universal application of the developed technology.
2. Improve the data entry mechanism to minimize time required to input data of varying format via cards or terminals.
3. Remove the restricted security imposed on reports for 120 days, adhering to commodity laws only.
4. Ensure that constructive, timely criticism and feedback are provided from the groups responsible for evaluating the crop assessment reports.

CLASSIFICATION AND MENSURATION SUBSYSTEM

The function of the Classification and Mensuration Subsystem (CAMS) was to provide an estimate of wheat acreage extracted from the LACIE sample segments. These acreage estimates, combined with yield estimates and historical agricultural data, were input to CAS, which in turn generated the final area, yield, and production estimates for evaluation. CAMS was designed to provide these estimates of spring and winter wheat for domestic and foreign areas using classical classification techniques. As in all LACIE systems, the basic premises of using no ground data, deriving estimation from data obtained by the Landsat multispectral scanner, and conducting LACIE in an operational environment were maintained. In reviewing CAMS, the peer group defined a number of key issues and findings.

1. CAMS developed and implemented techniques for the classification and mensuration of Landsat data over large, spatially separated, geographic areas.

   a. Phase I: 700 segments (13 440 000 acres)—U.S. Great Plains (all) and exploratory areas

   b. Phase II: 1700 segments (32 640 000 acres)—U.S. Great Plains (all), U.S.S.R. (partial), Canada (partial), and exploratory areas

   c. Phase III: 3000 segments (57 600 000 acres)—U.S. Great Plains (all), U.S.S.R. (all), and Canada (partial)

2. The estimates provided by CAMS were produced without the aid of ground data.

3. CAMS designed and implemented a highly structured procedure for the analysis of Landsat data. This system, designated Procedure 1, was basic to the successful operation of LACIE. The key points of Procedure 1 are that it strengthened the analyst-machine interface by identifying those functions best suited for machine processing and analyst decisionmaking, that it was designed to produce estimates in domestic and foreign small grains areas, that it was as accurate as previously used methods, and that it was less time consuming than previously used methods.

4. CAMS identified a set of data products believed to be necessary for optimal analyst-machine processing interaction. These products include agricultural data such as detailed, adjustable crop calendars; historical agricultural data; agricultural statistics; and soils information. Films products such as registered Landsat color-infrared and specialized enhancements and full-frame Landsat imagery were also needed. Spectral aids such as scatter plots and trajectory plots as well as the numerical data obtained from the satellite system were used extensively in Phase III of LACIE and are considered a necessary and integral part of the analysis-interpretation procedure. Clustering and classification maps were additional data products identified by CAMS as being extremely valuable. The key to the correct detection and identification of wheat (and other crops) rests on the availability of multitemporal imagery and ancillary data; their availability is considered vital to the success of a LACIE-type system.

5. Within a short time, CAMS made significant improvements in the efficiency of the analysis procedure: in Phase I, 12 man-hours per acquisition; in Phase II, 6 man-hours per acquisition; and in Phase III, 3 man-hours per acquisition. Future improvements would also be expected.

6. CAMS maintained data production throughput with significant changes in quantity of data analyzed and changes in personnel. In Phase I, there were 2000 acquisitions, 1100 of which were machine-processed;

in Phase II, 9000 acquisitions, 2400 of which were machine-processed; and in Phase III, 18 000 acquisitions, 5000 of which were machine-processed.

As a result of the experience gained during LACIE by CAMS, three major recommendations were defined by the peer group.

1. The techniques learned in CAMS should be applied to nonagricultural problems.
2. Continual development and refinement of spectral aids is vital.
3. Ground data should be used in the development of analyst procedures (particularly in cooperative foreign programs).

OPERATIONS MANAGEMENT, CONTROL, AND REPORTING

The primary functions of the operations management, control, and reporting (OMCR) activity were tracking the status of all elements within the AES, day-to-day coordination of operations, and identification and tracking of system-level discrepancies. In addition, the quality assurance group performed audits of operational procedures of functional groups whenever these procedures were changed. The OMCR activity was managed by the LACIE operations manager. The primary tool for tracking and reporting the activities was the Automated Status and Tracking System (ASATS).

The OMCR activity was in place at the onset of the project but continued to change to meet the requirements of management and the system throughout the three phases of LACIE. Initially, the tracking was performed manually; ASATS was developed later to accommodate system complexities that were greater than anticipated and to respond to management in a more timely manner. The OMCR required approximately 10 percent of the project's manpower; i.e., 10 to 12 people. In general, the performance of this LACIE element was excellent. It provided good visibility and responded effectively to management's changing requirements for status of a wide variety of items within the system. It was able to function in an environment of an increasing workload and a continuously changing, very complex, geographically distributed system configuration. The Operations Coordination Center was an effective vehicle for the coordination of daily activities. The use of audits of operational procedures contributed significantly to the validity of the total LACIE. All LACIE subsystems except the Yield Estimation Subsystem (YES) had good configuration control procedures. The lack of configuration control in YES hampered somewhat the work in the OMCR element.

The peer group concluded that any system of this complexity requires consideration of the OMCR function in the earliest planning phases of the system and incorporation of this function in its design and that it should be easier to carry out the OMCR function in a project-dedicated system.

Section 4

Findings of the Data Processing Systems Design Peer Group

Peer Group Members

D. Goodenough, Canadian Center for Remote Sensing, Chairman
J. Sulester, NASA Johnson Space Center, NASA Cochairman
J. Kast, Laboratory for Applications of Remote Sensing, Purdue University
T. Phillips, Laboratory for Applications of Remote Sensing, Purdue University

FINDINGS

An extensive, complex, high-volume data processing system which served a large-scale agricultural remote-sensing inventory experiment was designed and assembled. The system was implemented on several computers and effectively served a changing production process for LACIE results.

The production process was supported in three major areas: data acquisition and preprocessing; information storage, retrieval, and management; and applications evaluation support (as discussed in the symposium paper by D. H. Hay). During LACIE, the changing production process was controlled through an extensive test and evaluation program that was also supported by the data processing system.

Although some system concepts were tested before LACIE, a data processing system to support the LACIE requirements did not exist at the beginning. The initial system was successfully implemented and then substantially developed and enhanced throughout LACIE. The final LACIE data processing system provides a significant understanding of the functional specifications and requirements for future systems.

Many factors are discussed in the papers presented and in other LACIE documentation; they point to the significance of the LACIE data processing system. The review team believes that five items warrant further discussion in this paper.

1. The design and implementation of a large data base that is flexible and readily accessible for production and research and development processing
2. Successful demonstration of effective man-machine interactive processes for many remote-sensing data processing functions
3. The integration of a large set of diverse system components into a single system
4. The development of system configuration alternatives and an objective process for their evaluation
5. An assessment of the ability of the data processing system to meet its designed goals

LACIE Data Base

The processes that required data bases in LACIE are discussed by Westberry in the symposium paper entitled "The LACIE Data Base: Design Considerations." In this paper, Westberry quickly pointed out that the history, fields or dot, and process control data bases did not present a major problem for design and implementation. However, the storage and retrieval of the multispectral data did require careful design. It is the implementation of these data into an "image" data base meeting the LACIE requirements which provides an understanding of the key future requirements for an agricultural remote-sensing inventory of data processing systems.

The LACIE requirement for an image data base was approximately 3 billion bytes, which Westberry calculated by considering 4 acquisitions for 3840 sites and 16 acquisitions for 960 sites, giving a maximum acquisition total of 30 720. Each acquisition yielded 94 790 bytes. Since there was little control of the time at which a given segment would be processed, most of these data were required to be on-line to support the throughput requirements of the system. Immediate availability of all data was required to support the research, test, and evaluation needs of LACIE.

To meet these needs, the analysts stored data for a single phase or crop year in a disk-implemented data base in a segment or site orientation. The data base was designed and implemented so that at the beginning of a new crop year, data from the previous crop year would be available for research, test, and evaluation. As segments were acquired for the new crop year, data were inserted into an expanding data base and acquisitions from the previous crop year were deleted, making optimal use of the disk system provided for the data base.

The reviewers believe that the design and the implementation of the data base were flexible and provided readily accessible data for production requirements and for the support of the research, test, and evaluation requirements of LACIE. They also believe that even though LACIE made significant advancement in the state of the art of data bases supporting remote-sensing processing requirements, future systems will present even more complex requirements. This topic is further discussed in Westberry's paper and in the section of this paper entitled "Key Issues for Future Systems Design."

Man-Machine Interactive Processors

During LACIE, the roles of both man and computer were rapidly changing. The procedures for data analysis were improved as the project matured. The data processing system designed for LACIE supported the initial data analysis procedures and the changing requirements. The symposium paper entitled "Man-Machine Interfaces in LACIE/ERIPS" by Duprey addresses this subject.

Initially, the data analysis procedures required extensive interaction of the analyst with the segment data. The use of menus allowed productive interaction with the computer by users with diverse levels of experience. The menus were tutorial (thus decreasing analyst training requirements), could be modified as the system changed, and provided some structure to the software. As experience was gained from the interactive analysis of data, many of the analysts' tasks were found to be implementable on the computer. In the later phases of LACIE, high-volume production was required. Thus, noninteractive use of menus was also designed into the data processing system.

Duprey also discusses error recovery and checkpoint/restart capabilities in the LACIE data processing system. The automatic saving of the environment of a job at the conclusion of each major process was an important development for a system with many different activities at various processing stages, Both these capabilities, error recovery and checkpoint/restart, served to increase the effectiveness of the man-machine analysis team.

In summary, LACIE successfully demonstrated the effectiveness of the man-machine interactive processors for many remote-sensing data processing functions serving large-scale agricultural inventories. The use of menus, selected batch processes, interactive color terminals, error recovery, and checkpoint/restart procedures is expected to continue contributing to effective processing capabilities. However, as data loads increase, further improvements in efficiency will be required.

Data Processing System Integration

The LACIE integrated a large set of diverse system components (primarily from locally available equipment) into a single system, thus demonstrating several scales of processing capability. The demands placed on the data processing system for LACIE were complex and diverse. Each of these major demonstration subsystems required differing forms and amounts of data processing support. In addition, LACIE had additional processing requirements for support of the research and development and the test and evaluation activities.

At LACIE's inception, the existing data processing capability and the available facilities were unable to support the required functions or throughput; projected data processing system implementation schedules did not support first- or second-year throughput goals; limited funding was available for data processing system expansion; and few skilled remote-sensing data system design personnel were available. Consequently, a 3-year program with gradually increasing capability and throughput was planned; the program was contingent on available funding, improvement in analysis techniques and data system design skills, and equipment and software procurement schedules. This program depended on the use of a dispersed data processing system to benefit from in-place equipment. As a result, data processing tasks were parceled out to machines that were neither designed for, nor dedicated to, the processing of remotely sensed data or the formulation of crop production forecasts.

The environment of LACIE systems included the following major factors.

1. Data management. The LACIE project re-

quired the collection, preparation, processing, and tracking of data collected by satellite, ground observers, weather stations, field sensors, and aircraft sensors. There was a need for manipulation and traffic control of many large data sets, many of which were acquired by several functional subsystems of the Applications Evaluation System (AES) and by the research and development community and during test and evaluation activities. LACIE data management tasks are discussed in the symposium papers by Westberry and Rader.

2. High-speed processing. Lyon discusses the need to perform numerous repetitive computations on large groups of image elements during the data processing session. The processing goal of 4800 segments during the third year of the LACIE project implied the requirement for up to 60 hours per day of the serial processing capabilities of the most powerful machine available to the LACIE project. The solution to this problem reached by the LACIE systems design team was to augment the processing system with a programmable, peripheral high-speed processor. As reported by Lyon, this processor greatly reduced the clock time required to perform the computerbound portions of the processing for the AES.

a. Man-machine interaction. Differing amounts of man-machine interaction were required to satisfy the needs of various elements of LACIE. Troubleshooters, reworking segments whose classification results were inconsistent, and researchers and developers required very flexible and interactive systems. Analysts working in the mainstream of the production system required a highly automated (inflexible) high-production system. Duprey discusses how the man-machine interface was accomplished for the high-production system in the Ground Data Systems Division.

b. Seasonal processing demand. Because crops are grown on a seasonal basis, the data processing requirements for monitoring crops in a timely manner are also seasonal. The data processing peak is even more accentuated. As the crop year progresses, data are collected repeatedly over each sample segment as the satellite passes over the segment. Since the spectral response of a crop changes with time, data from as many as four dates are concurrently analyzed. Therefore, when segment A is analyzed in May, data from two satellite passes may be analyzed, whereas an analysis of the same segment in July may consider data from four acquisition dates. Processing data collected on four dates for a segment requires about six times as many computer operations as processing the data collected on a single date.

The peer review team believes that the task of designing and assembling the components of a data processing system for a project of the magnitude and complexity of LACIE was difficult, especially because of the decision to use in-place equipment (some performing non-LACIE processing as well) and the universal lack of experience in the design of high-volume operational remote-sensing data processing systems. LACIE successfully met this data processing challenge. In the process, valuable experience and insight were gained for the design of more efficient and effective systems for remote-sensing research and application.

Alternative System Configurations

The LACIE systems experience has led to the development of reasonable configuration alternatives for production and research and development system designs as well as for a more objective process for system selection. As mentioned in the preceding subsection, LACIE involved two levels of processing.

1. Quasi-production—processing of a high volume of data in a fairly automated manner
2. Research and development—processing of a relatively small volume of data in a highly interactive manner

In addition, LACIE provided experience in the use of mainframe computers (IBM 360/75, IBM 360/148, IBM 360/195, Univac 1108/1110), minicomputers (three PDP-11/45's and an Image-100), and a special-purpose processor (Goodyear STARAN). Thus, LACIE familiarized its project staff with the use of various machines and combination of machines to work a wide variety of tasks in both production and research and development.

The LACIE experience has pointed to the need for the following research and development system data processing requirements.

1. Rapid incorporation of apparently fruitful research software into the test and evaluation system
2. Access to large data bases by all elements of the research, development, test, and demonstration effort
3. The capability to concurrently support several modes of computer processing
4. The capability to track and manage all aspects of the data flow

5. The ability to transfer the demonstrated technology to user agencies

6. The flexibility to alter and upgrade any portion of the software system

7. The capability to support and effectively utilize certain commercially available software (such as statistical software packages)

With this background, future data system usage has been projected at five distinct levels, and several centralized and distributed processing systems have been analyzed on a cost-effectiveness basis (discussed in the symposium paper by Gregor and Spitzer). The findings of this study indicate that (1) all system configurations considered would perform more effectively than the system used by LACIE, (2) the development cost for any of the new systems considered would be recovered in lower operating costs, and (3) the 10-year life-cycle costs of the system with single-machine architectures were slightly superior to the 10-year life-cycle costs of the distributed systems considered.

The peer group review team believes that the question of distributed versus centralized processing systems will remain a key issue for some time. For more discussion of this subject, refer to the key issues section of this report.

The LACIE also sponsored an examination of the criteria used for computer system selection (see the symposium paper by Poole). This work was undertaken recognizing a need for a formalized system selection process based on quantifiable criteria since selection of a computer system to solve a particular problem is a complex task, comparison of various systems is difficult, and these factors often result in selection of the wrong systems.

Poole's formalized evaluation process proceeds in two phases. During the first phase, user requirements are defined and candidate configurations are formulated. The capability of the candidate configurations is then compared to combinations of computer processing requirements that are typical representations of the total system workloads during specific periods. These typical combinations of user computation requirements are called benchmarks. If a particular system configuration fails to handle the benchmarks adequately, it is either altered and retested or it is discarded.

Entry into the second phase of the evaluation process is predicated on successful satisfaction of the adequacy requirements. In the second phase, user-specific selection criteria are identified and segregated into criteria categories. The criteria categories are then weighed, as are the individual criteria components of each category. Each candidate system is then given a numerical rating for each of the criteria components, and an overall analysis of all candidate systems is thereby quantified.

The review team believes that the quantification of candidate system configurations was a useful process and would contribute to the objectivity of the selection process. It should be realized, however, that some very important variables (user expectations and transportability of processing techniques) are not easy to quantify with any degree of accuracy. Nor is it always reasonable to linearly sum the weights of the various criteria categories. For example, if the schedule for hardware delivery is too late for the system to be used, it makes no difference how transportable the system code will be or how inexpensive the system is. Conversely, if the cost is beyond a certain limit, delivery schedule and system availability and maintainability make little difference. Such relationships between factors would indicate the need for bounds on the criteria categories and/or a multiplicative rather than an additive system selection model. The reviewers do believe that although the evaluation model will help to add objectivity to the process, the need for some subjective judgments will remain in the evaluation and selection of data processing systems.

Evaluation of the LACIE Data Processing System

The LACIE planned to reach a processing capacity of 4800 segments over eight countries and to demonstrate a 14-day segment turnaround time from collection of Landsat data to availability of classification results for aggregation. These goals were not achieved. The LACIE data processing systems plan (Hay) called for a phased 3-year program with gradually increasing capability and throughput based on available funding, improvement in analysis and data system design skills, and equipment and software procurement schedules. Planned segment processing capacities and actual processing achieved (from Hay) are as follows.

Year	*Segments processed*	
	Planned	*Achieved*
Phase I (1974-75)	693	700
Phase II (1975-76)	1800	1800
Phase III (1976-77)	4800	3000

Although LACIE successfully achieved its processing goals for the first 2 years, it fell significantly short of its segment processing goal for the third year. This shortfall may be attributed to three factors.

1. During the second year of the LACIE project, one of the two tape recorders onboard the Landsat-2 satellite failed. To preserve the capability to monitor areas of the Earth during periods when the satellite was not in contact with a Landsat ground station, the decision was made to delete countries in the Southern Hemisphere from the scope in the third year of the LACIE project.
2. To improve the classification accuracy, the decision was made to acquire and process all Landsat data collected during the growing season over the segments monitored. Thus, the volume of data acquired and processed increased to a greater extent than the number of segments processed. As a result, the preprocessing capacity at the NASA Goddard Space Flight Center was saturated.
3. Given the decision to acquire and process all data collected over a segment during a growing season, more resources would have been necessary for every processing subsystem if the number of segments processed was to exceed 3000.

The timeliness goal of the LACIE project was for the data to be analyzed and available for production estimates 14 days after acquisition by Landsat (see the plenary paper by Hall). Because of the decision to work a single 40-hour-per-week shift and the fragmented nature of the processing system (which resulted from the decision to utilize in-place skills and equipment), the 14-day turnaround was not demonstrated. The feasibility of a 14-day turnaround was demonstrated.

Members of the peer group review team believe that the failure to meet the timeliness and segment volume processing objectives initially outlined for the LACIE project does not reflect negatively on the LACIE data processing staff. In fact, given the constraints of time, money, available equipment, and magnitude of the task, the LACIE data processing achievements were excellent.

KEY ISSUES FOR FUTURE SYSTEMS DESIGN

The review team in its deliberations identified five key issues that need to be considered in the design of future systems. All these issues were considered in the design of the LACIE system, but they continue to be important for future crop information systems.

Technology Transfer

The first key issue in system design is technology transfer. The development of a routine, operational crop information system must incorporate those system and operational constraints to be expected in the end organization ("user") that will operate the system. A highly skilled research and development team may transmit the new technology to the user in a number of ways. First, the technology may be transferred in the form of scientific papers and system documentation. In this case, the user must develop his own system but does not need to pursue research in interpretation algorithms. Second, the technology may be transferred as a high-level language software system capable of being run on a number of different computers. Unfortunately, such systems are inefficient and cannot usually achieve the throughput required of an operational crop information system. Third, one may transfer a complete system, software and hardware, from the research and development organization to the user. Such a transfer minimizes the necessity of creating inefficient software modules and preserves the integrity of the developed system. Fourth, one may incorporate into a system transfer the movement of operations personnel from the research and development organization to the user organization. By consideration of the end objectives of the user organization, an optimal future crop information system can be designed for the user. In embarking on such a development, the user organization must not radically change its end objectives during the system development. Both the user and the research and development organization are constrained in technology transfer to define objectives, performance criteria, the final system environment, and the end operational structure. The primary purpose of transferring technology is to eliminate costly duplication of research and development efforts. The pursuit of technology transfer by documents, hardware and software, system, and personnel complicates the initial design of any future operational computer system. Ultimately, however, this approach is the best one for future designs of crop information systems.

Data Base Access

The LACIE required on-line access to a 3-billion-byte data base. This data base included the data for one complete crop year plus a few acquisitions for certain segments for many years for research, test, and evaluation. A future crop information system will involve data from the thematic mapper of Landsat-D and from other satellites. The data volume of the thematic mapper will be seven times that of the existing Landsat multispectral scanner. The incorporation of synthetic aperture data could lead to a data volume comparable to the thematic mapper of Landsat-D. Therefore, the design of a new crop information system must consider alternative approaches to the construction of the data base with respect to size, cost, flexibility, speed of access, operational procedures, and error recovery.

The higher spatial resolution of the thematic mapper means that there will be greater correlation between adjacent picture elements (pixels). Important reductions in the physical size of the data base might be achieved through image coding. A future system could include a code/decode hardware processor for image movement from data base to display devices. The inclusion of cartographic ancillary data suggests that the designers of a future system may wish to have access to the data base in the coordinates of one or more map projections. Some geocoded data are more efficiently stored as polygons. A future data base might incorporate both grid data and polygons. In determining the size of the data base available online to the analysts, consideration should be given to the operational structure of the system. It may be possible to reduce the size of the on-line data base by scheduling the flow of acquisitions from tape to data base and back to tape in such a way as to maintain a minimum number of acquisitions with on-line access in the data base. LACIE achieved the design and implementation of a flexible data base. Faced with at least a sevenfold increase in data rate and with a multiplicity of digital data sources, the designers of the future data base will find the requirements for access to the large data volumes a key design issue.

Man-Machine Interface

The third key issue identified by the review team is the identification of the efficient division of tasks for man and machine to minimize costs. The assignment of tasks to the analyst or to the user depends on their level of experience. Sophisticated interactive algorithms require highly trained users. The paucity of such users often leads to a series of selected parameter values which less experienced users may apply. Two methods of communicating with the user and the software are menus of programs and a command language. Command language offers the operator the ability to accomplish a wide variety of operations, many of which might not have been considered when the system was originally designed. However, as with all languages, one must continually use it to maintain fluency. Operational procedures which involve many users but in which the user has only infrequent access to the system are better suited to menu structures. The menu of programs provides a general structure to the information extraction problem. For example, one might have such menu titles as "input, preprocessing, feature extraction, classification, and output." Interaction with programs selected from a menu could involve English text with brief explanations of the actions required. LACIE selected menus for its current system. The choice of menus will need to be reconsidered for a future crop information system since the number of programs may exceed a manageable menu structure.

Large numbers of programs make a system seem formidable to users less familiar with complex computing systems than system designers. In the new system, the designers will need to combine batch processing with interactive processing. One approach would be to develop a "learning" capability in the system. A sophisticated analyst would place the system in a "learn" mode and proceed to execute the complex sequencing of tasks for the recognition of a particular crop. The system would "learn" the task sequencing and record the selected parameters. The sequence of tasks might be stored as a command file. Then, a less experienced user could run the command file and associated parameter files and would be required to enter only a few parameters.

The system should be capable of accepting new algorithms easily for experimentation. Proven algorithms could then be entered into the menu structure easily; for example, if the menu structure were stored in a disk file that could be quickly edited. At the end of each process, the LACIE system saved the state of the system, a desirable feature in any future system. With large numbers of programs, the state of the system needs to be recorded so that an analysis session can be interrupted or restarted after system or analyst errors are encountered. Error recovery, well handled in LACIE, also needs to be considered for the future system.

Financially, more effort is placed on software than on hardware in large system developments for crop information extraction. Hardware is important, especially display devices, for effective man-machine interaction. Future display devices should incorporate features that allow the full expression of the analytical activity of a human perceiving an image. For example, it should be possible to display the optimal spatial frequencies, colors, and contrasts for an observer. Real-time magnification with resampling, enhancements, and a variety of cursor modes is also required. The display devices should be in a comfortable physical environment with low noise. Response times should be fast to minimize human fatigue during a session.

The determination of the efficient division of tasks for man and machine to minimize costs is difficult and will require considerable research for the design of the next crop information system.

Distributed Versus Centralized Systems

The review team concluded that the trade-off between cost and performance of distributed (physically dispersed) versus centralized (local) systems was a key issue requiring further investigation for future systems design. As with technology transfer, the selection of a distributed or centralized system should depend, in part, on the organizational structure in which the system will be finally placed.

In a distributed system, one has a complex communications network. Operational control is dispersed. Often, the modes would be minicomputers with array processors and special-purpose hardware. In a distributed system, one gains parallel processing and possibly more system tolerance of hardware failures. The initial costs for such a system are lower. However, the greater difficulty of programing minicomputers boosts software costs higher than those costs encountered with a large computer (refs. 1 and 2).

A centralized system has a large main computer. The initial costs are high, but software development costs for a crop information system are lower. It is easier to maintain a secure system with a centralized facility, and operational control is relatively simple. The design of a centralized system is less costly than that of a distributed system. Consideration of centralized versus distributed systems will be needed in the design of a future system and will be strongly influenced by the user's operational structure.

Alternative Processing Strategies

Four alternative processing strategies are proposed: the use of spatial and contextural properties, parallel processing, flexible sampling, and feature selection.

1. Since LACIE began, a number of new techniques that should be considered in the design of a future system have been developed. The multispectral scanner of the Landsat contains most of its information in its spectral values. There is, however, significant information in the spatial content of the image. Spatial information has not been used to improve LACIE acreage estimates. For example, it has been shown that a simple postclassification context filter can reduce classification errors of agricultural crops by a factor of 2 (ref. 3). The use of ECHO (Extraction and Classification of Homogeneous Objects) (ref. 4) has improved the accuracies of classifications significantly. The thematic mapper of Landsat-D, with its higher resolution data, will necessitate the use of spatial and spectral techniques in order to achieve maximum accuracy in crop acreage estimation. A third approach to the incorporation of spatial information into the analysis of Landsat imagery has been the development of texture features (ref. 5), which, when combined with spectral features, have yielded higher classification accuracies. Fourth, the computationally inefficient methods of edge detection, image segmentation, and field classification (ref. 6) have produced the highest accuracies for single-date crop classifications. Future crop information systems should include the use of spatial features and contextural properties for acreage estimation.

2. Inclusion of spatial features in crop image analysis necessitates operations that act on windows or groups of pixels rather than on the earlier per-pixel processes. The extraction of spatial features for various window sizes and the introduction of large data volumes necessitate the development of processes that act in parallel on the image. Parallel processing can be achieved by multiple simple processors performing identical operations on different data windows or by multiple sophisticated processors performing different operations on the same data sets. The architecture of a parallel processing system and its programing are radically different from the existing LACIE system. Research is required to identify the optimal mix of parallel and sequential processing, the optimum architecture for

high throughput and ease of programing, and the processes which best fit a parallel processing system. Although it is premature in remote sensing to introduce hybrid optical/digital systems, by the early 1980's a successful parallel processing system might include a combination of optical processors and computers. Operations such as spatial filtering could be performed in the optical processor rapidly while pattern extraction was executed in the computer. The review team recommends, therefore, that the design of a future crop information system include a thorough review and consideration of parallel processing approaches.

3. The design of a crop information system depends on the amount of data to be processed, which in turn depends on the sampling strategy. Sampling strategy (reviewed in much greater detail by other LACIE teams) should be more flexible for future systems in order to profit by good weather scenes, major changes in crop acreages and distributions, and the availability (for some countries) of accurate ground information. Particularly desirable are changes in sampling strategy that reduce the number of segments and hence total system costs.

4. The LACIE computed the best set of three Landsat dates to be combined with the most recent Landsat acquisition. Feature selection involves the computation of a distance measure between two classes and the selection of that feature set which permits the best recognition of the two classes. To compute the distance measure, one needs the statistical parameters, such as spectral means and covariances, for each class. LACIE personnel carried out exhaustive feature selection; that is, they calculated the best feature set for all possible date combinations and classes. New algorithms have been developed for which the optimal feature set can be determined without calculating all possible feature and class combinations (ref. 7). An alternative processing strategy for a new crop information system would be the ability to compute from all the acquisitions the best m-feature subset to identify a particular crop, with features drawn individually from any date. A reduction in the number of features in the recognition process would significantly reduce system costs. The review team recommends that a new system incorporate efficient algorithms for feature selection.

REFERENCES

1. Burr, William E.; and Gordon, Robert: Selecting a Military Computer Architecture. IEEE Computer, vol. 10, Oct. 1977, pp. 16-23.
2. Fuller, Samuel H.; and Burr, William E.: Measurement and Evaluation of Alternative Computer Architectures. IEEE Computer, vol. 10, Oct. 1977, pp. 24-35.
3. Goldberg, M.; and Goodenough, D. G.: Analysis of a Spatial Filter for Landsat Imagery. J. App. Photog. Eng., vol. 4, 1978, pp. 25-27.
4. Kast, J. L.; and Davis, B. J.: Test of Spectral/Spatial Classifier. LARS Contract Report 112877, Purdue University, 1977.
5. Haralick, R. M.: Texture-Tone Study With Application to Digitized Imagery. University of Kansas Remote Sensing Laboratory Report 182-6, 1974.
6. Goodenough, D. G.; Goldberg, M.; O'Neill, K. J.; and Teillet, P. M.: Spectral and Spatial Features for Remote Sensing Classification. Proceedings of the Fifth Canadian Symposium on Remote Sensing, (to be published).
7. Goodenough, D. G.; Narenda, P. M.; and O'Neill, K. J.: Feature Subset Selection in Remote Sensing. Canadian J. Remote Sensing, vol. 4, 1978, pp. 143-148.

Section 5

Findings of the USDA Applications Test System Peer Group

Peer Group Members

G. Nagy, University of Nebraska, Lincoln, Chairman
J. Murphy, USDA Foreign Agriculture Service, Houston, Texas, USDA Cochairman
D. W. Cary, Central Intelligence Agency, Washington, D.C.
H. Harkness, Sparks Commodities, Inc.
R. Head, USDA Office of Automated Data Systems, Washington, D.C.
R. Henderson, MITRE Corporation
R. LeGault, Environmental Research Institute of Michigan
R. McArdle, USDA World Food and Agricultural Outlook and Situation Board, Washington, D.C.

ASSESSMENT OF THE USDA APPLICATIONS TEST SYSTEM

The USDA Applications Test System (ATS) has been the subject of careful systems studies. These studies have resulted in a system design that utilizes the state of the art in digital image analysis hardware to allow efficient processing of the Landsat imagery. At the same time, the system is flexible enough to allow for the accommodation of new processing requirements.

The system as currently configured is based on a PDP-11/70 minicomputer that coordinates the activities of the image analyst stations, the imagery data base, and the array processing unit. There are three image analysis stations, each containing two interactive color displays driven by I^2S model-70 refresh memories, an interactive DECwriter terminal, and an interactive color graphics terminal.

The segment imagery is displayed on the interactive color displays which allow for the simultaneous display of five acquisitions of Landsat data. All analyst work, including pixel selection and identification, field delineation and identification, and pixel masking, is done on the color displays using a trackball cursor. Classification and clustering maps of the segments are displayed as requested by the analyst.

The imagery and the results of classification and clustering are stored on the imagery data base, which currently consists of two Ampex 300-megabyte drives. The data are directly transferable via the PDP-11/70 massbus to the array processor and via the unibus to the image displays.

All classification and clustering is done on a Floating Point Systems AP-120B programmable array processor. As currently configured, full maximum likelihood classification of one segment acquisition (four channels; 117 lines by 196 pixels) into 30 classes is possible in only 10 seconds. Because the AP-120B is fully programmable, it will allow needed flexibility as processing requirements change or evolve.

The ICS interactive color graphics terminal has a microprocessor controller and allows the graphical display of historical and meteorological data for the segments at the request of the analyst.

FINDINGS

The peer review of the USDA ATS resulted in the following findings.

1. Because of the relatively late formulation of USDA goals within the LACIE time frame, there is a certain discrepancy between the LACIE objectives and current USDA requirements. For instance, it appears that commodity analysts will continue to make national and global production forecasts and estimates, with agricultural/meteorological and Landsat data used only as inputs in the process, rather than as part of an automated end-to-end system as envisioned in LACIE. Early assessment of crop condi-

tions will have a high priority, as opposed to the harvest-time accuracy stressed in LACIE. Hence, direct and complete transfer of the LACIE technology, procedures, and objectives is neither desirable nor currently taking place.

2. Many elements from LACIE are being transferred partially or fully into the current USDA system. Among the most significant of these are the stratification concept through agrophysical units; the segment image data base; the Phase I, II, and III classification and clustering procedures and algorithms; the yield models and data bases; spectral and other analyst aids; image and ancillary data acquisition techniques; and trained personnel.

3. New elements included in the USDA system are the modular hardware and software design based on dedicated minicomputers, the integrated cellular data base design, the provision for output products which played only an intermediate role in LACIE, and the operation of a rigorous cost estimation and budgeting program.

4. The current organization and system design of the USDA ATS operation provides a significant amount of flexibility for meeting new and changing objectives and is an important tool for testing procedures and products in an operational environment.

ISSUES AND RECOMMENDATIONS

1. Some of the LACIE procedures incorporated into the ATS, particularly as components of the early assessment, have not yet been validated with respect to the new objectives.

2. Some of the procedures and techniques developed outside LACIE and not included in LACIE because of LACIE's restricted objectives and strict ground rules need to be investigated further.

3. Some of the LACIE procedures now being transferred to the ATS, such as the sampling and aggregation schema, may not be directly applicable to the changing objectives of the USDA and therefore need to be modified.

4. Detailed objectives and priorities with respect to items 1, 2, and 3 urgently need to be defined for incorporation into the research and development design.

MINOR POINTS

1. The excessive downtime of the current system provides a serious handicap to meeting intermediate ATS objectives and needs to be reduced, perhaps through the engagement of an electronics technician or increased redundancy.

2. The proprietary software arrangement is difficult to administer, and the software itself is difficult to modify.

3. There is a lack of adequate hard-copy output (plotter) in the current system.

4. There is a lack of adequate graphic input (coordinate digitizer) that would greatly accelerate the compilation of portions of the data base such as the soils inventory.

5. In view of current developments in geographic data base technology, it might be desirable to reexamine alternatives and enhancements to the current grid-based data organization.

SUMMARY

The objectives of the ATS, originally announced in December 1974, have gradually shifted and are currently influenced most heavily by the recent USDA Secretary's initiatives, which place a high priority on early crop assessment and detection of unusual crop conditions abroad.

Early crop assessment abroad is currently performed by USDA analysts on a largely qualitative basis. It is not known how accurate these assessments are for different regions and under different conditions; therefore, no reasonable quantitative objective can be set for partial automation of this task. In other words, assessments produced either by ATS or by conventional methods cannot be verified. The ATS has demonstrated that it can produce a complete and timely early assessment for restricted regions, but there has been no feedback so far from the Foreign Agricultural Service, the ATS's putative major client, on the usefulness of these estimates. It is believed that very few persons in USDA understand remote sensing and the LACIE technology. This has proved a handicap to involving other USDA personnel in planning the utilization of ATS products in operational USDA activities.

The hardware and software configuration of the ATS represents a scaled-down minicomputer-based version of the LACIE capabilities. The current configuration is quite flexible and adaptable to changing requirements and to the incorporation of new technology. In the opinion of the peer group, the importance of high-speed classification hardware has been overemphasized at the expense of expanding the data-base capability. However, the processing speed of the system is likely to have higher priority than the ancillary data supply (soils, crop practices, historical information, meteorological data) for several years.

The turnkey approach to technology transfer adopted by the ATS has created several problems. Among these are the inordinate difficulty (including legal constraints) of introducing even minor software changes in the existing system, the heavy dependence on contract personnel, and the lack of hardware maintenance facilities and personnel at the ATS site. As a result, development and maintenance work frequently prevents ATS analysts from working on the system, contrary to the objective of providing a quasi-operational test environment. The unpredictability of system availability also makes it difficult to schedule other USDA personnel for familiarization.

Although this might be the proper time to review LACIE accomplishments, the ATS itself is not at a comparable stage. Inasmuch as it is not intended as a strictly experimental system, research goals of the LACIE 90/90 type cannot be substituted for indications of ATS integration into mainline USDA activities. The technical problems that would prevent such integration appear to have been solved. Since analysts using the computerized system will have all the information available under the current manual method, augmented by the classification, clustering, image processing, and rapid information retrieval capabilities of the computerized system, it is difficult to see how the computer-aided analyst could not outperform his unaided counterpart. Therefore, this peer group fails to find any unsurmountable obstacle to the accomplishment of the objective according to the current ATS plans and schedule.

Section 6

Findings of the LACIE Supporting Research Peer Group

Peer Group Members

R. Holmes, General Motors Institute, Cochairman
R. Shay, Oregon State University, Cochairman
J. Erickson, NASA Johnson Space Center, NASA Cochairman
W. Anderson, U.S. Geological Survey EROS Data Center
J. Estes, University of California at Santa Barbara
C. Hay, University of California at Berkeley
R. Jensen, NOAA National Weather Service, Honolulu, Hawaii
R. W. Leamer, USDA Science and Education Administration, Weslaco, Texas
B. Liska, Laboratory for Applications of Remote Sensing, Purdue University
R. Welch, NASA Ames Research Center

INTRODUCTION

The LACIE baseline review in December 1974 and the LACIE documentation of initial project design detailed many concerns about wheat area, yield, and production estimation techniques and procedures. The review occurred at the beginning of the LACIE project; foresight indicated that the main tasks for the supporting research, test, and evaluation program would be the following.

1. In wheat area estimation (ref. 1):
    a. Signature extension strata determination, sampling, and analogous site identification
    b. Classification accuracy and bias, including human analyst effects; boundary pixel effects; training field homogeneity testing; detection of harvested wheat; category thresholding; and multitemporal layered classification
    c. Data quality, including data dropout classification strategies and a cloud and shadow detection algorithm
    d. Feature selection and photograph interpretation techniques, including a search for an optimum biological phase for the discrimination of various types of wheat from nonwheat; improvement of human analysis-interpretation techniques; and concern for the man-machine interaction effects
    e. Testing and evaluation of 16 specific approaches to wheat area estimation related to the research tasks cited above
2. In wheat yield estimation (ref. 2):
    a. Selection of yield models using universally available data
    b. Identification of yield strata
    c. Testing and comparison of yield models
    d. Selection of historical data
    e. Selection of operational models
    f. Identification of crop calendars
    g. Identification of surface weather models based on environmental satellite data
    h. Comparison of phenological development and yield
    i. Comparison of other factors based on the spectral appearance of wheat and its relation to yield
    j. Research data and measurement requirements to support the research tasks
    k. Testing and evaluation of requirements for the research tasks cited above
3. In wheat production estimation (ref. 3):
    a. Crop assessment improvements through research in sampling strategies and error analysis
    b. Development of aggregation strategies that include attention to cloud cover problems and harvested wheat estimates

c. Development of a complete error analysis model for LACIE, including the effects of analysts' selection procedures and possible misidentification of training fields, signature extension, feature (channel) selection, sample sufficiency, misregistration, cloud cover, haze, data dropouts, different area estimation methods, and the number of classes of scene objects

Some general opinions about the supporting research, test, and evaluation program (the research program) are as follows.

First, the initial design of the research program was on target, with one forgivable omission in technology transfer. The initial plan was understandably broad and less focused than the research work which developed as the actual problem areas unfolded. All technical issues which arose were recognized as potential problem areas in the original plan. However, it now appears that there is a need for research into the managerial, political, and economic aspects of the development of a genuine user community; technology transfer has not been as easy as first thought and the principal deterrents do not appear to be technical.

Second, the LACIE research program is judged to have been rapidly responsive to needs expressed and problems uncovered in the more operational aspects of LACIE. One example is the complete redesign of the analysts' interaction with data and machines to decrease wheat area estimation bias. Another example is the complete change of direction on signature extension when the original strategy fell short of expectations. A third example is the increasing emphasis on yield modeling and crop calendar modeling when these proved to be major gaps in scientific understanding.

Third, the research program has served to define and "prioritize" the actual problems to be solved in the furtherance of a global crop-monitoring system. It is common for a carefully time-controlled high-technology program to leave a rich legacy of maps and routes for future directions; LACIE is no exception to this rule.

COMMENTS ON LACIE RESEARCH

In wheat area estimation, the most notable research development was made in careful prompting of the analyst in the selection and labeling of training data for machine classification. The simple single name for a series of improvements over initial area estimation techniques is Procedure 1. This procedure ensures a random sample of "pure" (nonboundary) pixels from a grid overlay of a 5- by 6-nautical-mile segment to be labeled by the analyst and to provide starting vectors for machine spectral clustering of total scene pixels. The procedure also attempts to remove classification bias through a forced labeling of a second random sample of the scene using both pure and boundary pixels. Primary aids to the analyst are accurate crop calendars, spectral data feature extraction plots of greenness versus brightness (Kauth variables), and crop temporal trajectories in greenness-brightness space. Stripped of details and LACIE-specific names, the development amounts to this: careful attention to the design of a random sampling strategy with careful attention to the factors of human analyst interaction with data and machines decreased classification errors and decreased 5- by 6-nautical-mile segment wheat area estimation time from 12 to 3 hours. The development of Procedure 1 also made multitemporal analysis much easier than with the original analysis technique employed; this has served to increase wheat area estimation accuracy significantly.

Sun-angle and haze corrections and clustering algorithm improvements also aided Procedure 1 because of increased consistency in data and data-processing techniques. It now appears that a formal machine-to-analyst questioning process may aid in enhanced consistency in analysts' labeling decisions.

Detailed classifier strategies studies or, equivalently, proportion estimation methods did not appear to result in a significant gain in acreage area estimation accuracy. In general, different methods evoke slight changes in decision boundaries. More advancement appears to result from attempts to estimate efficiently and accurately the pixel total-population density function in spectral-temporal space as provided by Procedure 1.

The initial approach to signature extension (i.e., training a classifier on one segment and processing nearby segments from that classifier) simply failed. Geographically close segments are not necessarily spectrally close. A more promising approach to achieve the classification efficiency of signature extension is based on minimization of a multiple segment random sample after segments are grouped into strata of similar environmental factors and after data are corrected for haze and Sun-angle effects.

There are two problem areas in LACIE in which research has provided little progress. First, classification accuracy in regions with small fields is not as good as in regions with large fields because of boundary effects and analyst labeling difficulties. Higher resolution satellite data appear to be the best hope here; various edge-correction algorithms and spatial clustering of per-field classifiers have not yielded dramatic performance increases. Second, it has proven extremely difficult to distinguish spring wheat from spring barley. A possible but less than satisfying approach is historical, cultural, and economic modeling of wheat to small grains ratios. There is some hope that better crop calendar models and more frequent satellite overpass data may enable more subtle spectral and temporal discrimination of wheat from confusion small grains.

Wheat yield estimation research made reasonable advances during LACIE if one considers the level of effort that was funded. Simple statistical models developed from adequate historical data, meteorological data, and an estimate of cultural practices effects work reasonably well under average conditions and have been used in LACIE to date. If historical data are sparse, there is an inadequate basis for model development. The present weather station network is more sparsely distributed than the LACIE segment sites, making it possible to miss localized severities. Even when an adequate historical data base is available, current models fail to respond to extreme episodic weather events.

The present approach to this problem of yield modeling is to generate stratified statistical empirical models using more frequent and more spatially fine-grained input data of weather, soils, and cultural practices. This modeling approach is hopefully a productive intermediate step between simple statistical models and mechanistic (physiological) system models. Mechanistic or physiological yield modeling is in the embryonic stage of development with no short-term operational usage expected soon.

It is generally believed in remote-sensing circles that yield information may be contained in the spectral response of satellite data and its development over time. However, it is only now that sufficient field spectral ground measurements are beginning to be available, together with ancillary data that might permit correlation of yield-dependent causes and spectral effects.

It is the peer group's opinion that wheat yield estimation by any means—satellite or otherwise—is less advanced than wheat area estimation at this time and is thus a major problem to be addressed through research in wheat production accuracy achievement.

A problem related to yield estimation and of great importance in aiding the analyst in wheat area estimation is the generation of accurate crop calendar (phenology) models. There is a need to develop accurate, spatially dependent starter models to correct nominal crop calendars for the difference between normal and actual planting dates. It is also necessary to correct for variable dormancy periods for winter wheat. It was believed that the Robertson biometeorological time scale model offered the best approach for LACIE with correction for dormancy and possible modification for inclusion of a soil moisture variable. Here again, field spectral measurements have only recently been available in sufficient quantity and variety to conceive of modeling the inference of crop calendars from spectral-temporal satellite data.

The major advance in wheat production estimation was achieved by an improved sampling strategy that assigned segments according to naturally and agriculturally similar areas rather than political subdivisions. This permitted a significant reduction in the number of 5- by 6-nautical-mile segments in the U.S. Great Plains and in the U.S.S.R. necessary to achieve LACIE goals. Further sampling efficiencies have been proposed but not well tested in LACIE in its short timespan. Here again, as in the case of Procedure 1 for the analysts' aid, sampling strategy has been a fruitful area of research for the advancement of the technology. Sampling has also proved to be a far more complex issue than was apparent at the outset of LACIE.

SUGGESTED FUTURE RESEARCH AREAS

Future research for the advancement of global crop monitoring technology should emphasize the following areas.

1. Organizational design and management for the establishment of a genuine user community
2. Yield and crop calendar modeling with increased emphasis on physiological modeling and inference from satellite data with close attention to sampling design for statistical models
3. Further improvements in efficiencies of area estimation by even more attention to man-data-machine interaction, which the peer group believes will reduce analysis time significantly even from current levels
4. Analysis of the type and quality of environmental data employed by analyst-interpreters

5. Sampling and aggregation design, including signature extension efficiencies

REFERENCES

1. Classification and Mensuration Subsystem (CAMS) Requirements. LACIE-C00200, Vol. II, NASA Johnson Space Center, Houston, Texas, Dec. 1974.
2. Yield Estimation Subsystem (YES) Requirements. LACIE-C00200, Vol. III, NASA Johnson Space Center, Houston, Texas, Dec. 1974.
3. Crop Assessment Subsystem (CAS). LACIE-C00200, Vol. IV, NASA Johnson Space Center, Houston, Texas, Dec. 1974.

NOTES