Reprinted from

Eighth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Crop Inventory and Monitoring

July 7-9, 1982

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

ACQUISITION HISTORY SIMULATION FOR EVALUATION OF LANDSAT-BASED CROP INVENTORY SYSTEMS

J.H. SMITH, J.T. MALIN, C.C. LIN

Lockheed Engineering and Management
Services Company, Inc.
Houston, Texas

M. DVORIN

Omnitape
Houston, Texas

ABSTRACT

This paper describes the development and evaluation of a simulation procedure which produces patterns of Landsat data loss attributable to cloud patterns that are characteristic of a crop region. This simulation procedure is part of a simulation system under development which evaluates the performance of crop inventory system components over a number of years and under a variety of conditions.

I. INTRODUCTION

An important factor in the success of Landsat-based agricultural prediction is nonresponse. Unlike many remote-sensing applications, agricultural prediction requires repeated observation of the same scene throughout the crop year. Furthermore, these observations must fall in specific crop-stage-related time intervals, or biowindows, for accurate estimation of crop proportions in the scene. Since Landsat coverage occurs only once every 18 days, the loss of one or two critically timed data acquisitions can be very costly.

The simulation procedure (simulator) is part of a simulation system being developed in support of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program. The procedure has been designed to support studies of variance and bias in crop estimates that may be caused by nonrandom data loss because of cloud cover. In order to do so, acquisitions and losses from a sequence of periodic observations of each Landsat scene are simulated. Simulation of these sequences, scene acquisition histories, is based on previously observed cloud patterns in the crop region of interest. Studies of acquisition histories to be expected in foreign areas, particularly in tropical or subtropical areas with frequent cloud cover, should provide guidance concerning the feasibility of alternative crop inventory systems. Assessment of the potential benefits of two or three satellite systems is an additional capability.

*Under Contract NAS 9-15800 at the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, Texas 77058.

II. AgRISTARS DATA FLOW

Most current technology development in the AgRISTARS program is based on the sample segment, a 5- by 6-nautical-mile or 3.6- by 6-nautical-mile scene for which crop proportions are estimated. Regional estimates may then be produced by aggregating the segment-level estimates. The simulator is designed to mimic the process through which segment data are collected, preprocessed, and screened for quality. Figure 1 depicts this process, described in brief below. More detailed discussions of the process are provided in the literature.[1,2]

Once the segment has been located, a 10- by 11-nautical-mile scene containing the segment is extracted from the 100-nautical-mile-wide swath of data collected by Landsat. This scene, called the search area, is then screened by computer for excessive cloud or snow cover. The scene is discarded if more than 10 percent of its area exhibits a "whiteness" value above a set threshold. Approximately one-half of all potential data acquisitions are lost at this stage.

If the search area data are not rejected, the next step is registration of the segment within the search area.[3] That is, the segment is matched spatially with data collected earlier in the season (if this is not the first data collected in the current crop year). About 10 percent of all potential acquisitions are lost through failure of the registration algorithm. In the context of this paper, if a segment is successfully extracted, screened, and registered, an acquisition of the segment is said to have occurred on the date the segment data were collected. These acquisitions are screened by trained analysts for excessive cloud, cloud shadow, haze, and snow cover, as well as unacceptable misregistration. The percentage of

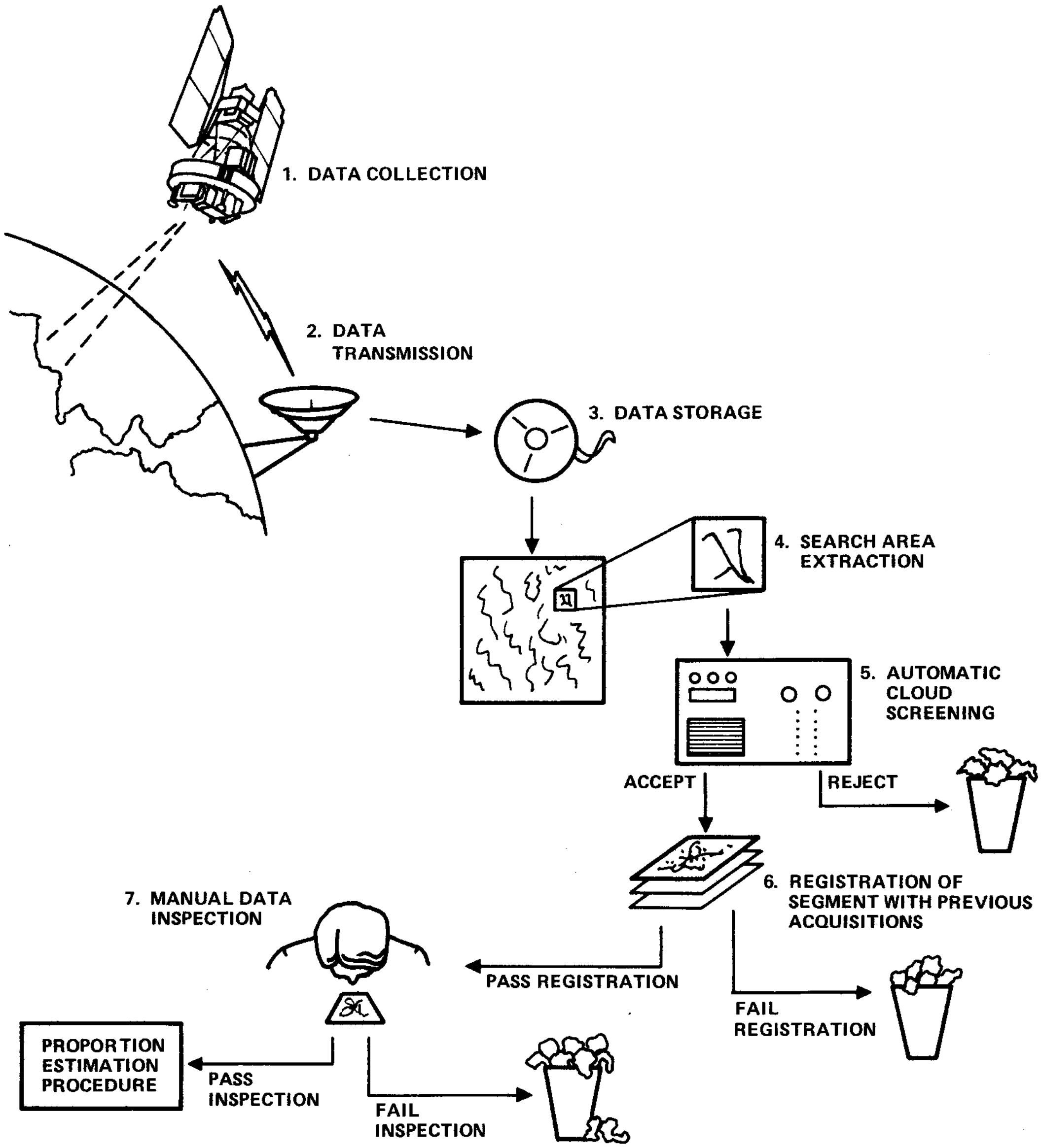

Figure 1. AgRISTARS data flow.

acquisitions lost at this stage depends on the requirements of the crop proportion estimation procedure employed.

III. DESIGN OF THE SIMULATOR

The acquisition history of a segment is simulated by first locating the segment geographically, then determining the dates on which the segment receives Landsat coverage. Next, a historic year is selected, and archived meteorological data are used to reproduce the observed weather patterns from that year. Thus, cloud cover itself is never actually simulated.

Cloud percentage at each segment location on a given date in an archived year is determined, and then acquisition or loss is simulated with probability that is a function of cloud percentage. The resulting acquisition history is compiled, and it may then be compared with crop-stage-related data to determine the type of crop estimate that the segment acquisition history could support.

A. CLOUD COVER

While a number of existing cloud cover models[4,5,6] appear to be quite adequate in most contexts, it is felt by these authors that no model is capable of capturing both the spatial and temporal patterns observed in nature. Thus, data bases containing archived meteorological data were established so that their contents could be retrieved and observed weather patterns could be re-created. One data base, limited in area to the U.S. Midwest, contains ground-based observations of sky cover.[7] A more general data base consists of albedo and heat flux readings worldwide from the National Oceanic and Atmospheric Administration (NOAA) series of polar-orbiting meteorological satellites (metsats). These data are extracted from the NOAA Operational Heat Budget Archive.[8]

To determine cloud cover at a segment location using the ground-based data, the observations are first corrected for observer bias and then interpolated to the segment location. The bias correction was performed because it has been demonstrated[9] that ground observers usually overestimate the amount of sky cover in an area.

To determine cloud cover using the metsat data, the first principal component of albedo and heat flux (P) is calculated, and this value is interpolated to the segment location. This value is then transformed to cloud cover by a relation developed as follows: The cumulative distributions of cloud cover and P were modeled with polynomials. Then, since cloud cover is essentially monotonic in P, the percentiles of the two variates were matched. That is, if an observed value of P falls on the n^{th} percentile of the cumulative distribution function of P, it is transformed to the n^{th} percentile of cloud cover. This transformation was employed because it forces the density of predicted cloud cover to match the density of observed cloud cover.

B. SIMULATING ALGORITHMS

Once cloud cover has been determined at a segment location, it is necessary to find the probability of obtaining an acquisition on the given date. The relation between cloud cover and acquisition probability was found by taking a set of 25 segments in the U.S. Corn Belt and interpolating ground-based cloud cover to each segment location for approximately 400 coverage dates. If the segment was acquired, a one was coded; otherwise, a zero was coded. These data were then used to fit the parameters a and b of the model

$$P(C) = a + bC$$

where C represents interpolated cloud proportion and P(C) is acquisition probability as a function of C. The actual parameter estimates were determined by the maximum likelihood method.

C. PROCESSABILITY

For each simulated segment acquisition, a within-segment cover percentage is generated, based on the percentage of cloud cover when the segment was acquired. Five procedure acceptability criteria were considered (0, 10, 20, 30, and 40 percent segment cover); then each acquired segment was coded as one (acceptable) or zero (otherwise), and a linear function was fit for each criterion in the manner described above. The simulated segment cover represents the acceptability of the acquisition, given a proportion estimation procedure. This final step is analogous to manual data screening. A segment with acceptable acquisitions in the required biowindows is processable.

IV. EVALUATIONS

Developmental testing of the simulator has recently been completed. A summary of the results is presented below.

A. DESIGN OF THE TEST

The test was based on segment processability as determined by the automated spring small grains proportion estimation procedure SSG4.[10] Actual segments from 1976 and 1977 in the U.S. northern Great Plains were used, with 10 simulated acquisition histories compiled for each segment, using NOAA-4 and NOAA-5 metsat data. Each of these 10 acquisition histories, along with the observed acquisition history, was input into procedure SSG4, which decided on segment processability. Comparisons were then made based on segment processability, as well as on an acquisition-by-acquisition basis. The simulation was run from Julian date 70 to 310 (March 11 to

November 6) to ensure that all biowindows defined for procedure SSG4 were covered. Because some early and late season acquisitions appear to be missing for reasons other than cloud cover or failed registration (e.g., deletion from the order list), it was decided that observed and simulated acquisitions would be compared only from Julian dates 130 to 230 (May 10 to July 29).

B. RESULTS OF THE TEST

Tables 1(a) and 1(b) compare observed and simulated segment acquisitions for the years 1976 and 1977. Tables 2(a) and 2(b) compare segment processability for the same set of segments. In each case, the results are a composite of the results obtained for 10 replications of the simulation. Both sets of tables give observed cell frequencies, followed by the frequencies which would be expected if the simulated and observed data were independent. In each of the four cases, a chi-square test of independence was performed, and the results were very highly significant, indicating a definite correlation between the observed acquisition history of a segment and its simulated acquisition histories.

C. DISCUSSION

A comparison of the marginal totals shows reasonable to very close agreement between observed and simulated acquisition and processability rates. Note that the simulated acquisition rate for 1977 is somewhat low. Subsequent investigation has revealed that a possible cause of the problem is a calibration

Table 1. A Comparison of Simulated and Observed Segment Acquisitions. Note: A = acquired; N = not acquired; 0 = observed; E = expected.

(a) 1976 (57 segments)

Observed		Simulated A	Simulated N	Total
A	0	1442	1038	2480
	E	1090	1390	(45.3%)
N	0	962	2028	2990
	E	1314	1676	(54.7%)
Total		2404 (44.0%)	3066 (56.0%)	5470

$\chi^2 = 368.7$

(b) 1977 (86 segments)

Observed		Simulated A	Simulated N	Total
A	0	1459	1821	3280
	E	1100	2180	(41.7%)
N	0	1176	3404	4580
	E	1535	3045	(58.3%)
Total		2635 (33.5%)	5225 (66.5%)	7860

$\chi^2 = 303.3$

Table 2. A Comparison of Simulated and Observed Segment Processability. Note: P = processable; N = not processable; 0 = observed; E = expected.

(a) 1976 (57 segments)

Observed		Simulated P	Simulated N	Total
P	0	343	57	400
	E	316	84	(70.2%)
N	0	107	63	170
	E	134	36	(29.8%)
Total		450 (79%)	120 (21%)	570

$\chi^2 = 37.3$

(b) 1977 (86 segments)

Observed		Simulated P	Simulated N	Total
P	0	246	174	420
	E	201	219	(48.8%)
N	0	165	275	440
	E	210	230	(51.2%)
Total		411 (47.8%)	449 (52.2%)	860

$\chi^2 = 38.3$

error between the metsat used for most of 1976 (NOAA-4) and the satellite used in 1977 (NOAA-5). Further investigation is currently underway.

The marginal totals of table 2(b) show that the simulated processability rate for 1976 is somewhat higher than the observed rate. This higher rate is due in part to the absence of a number of late season acquisitions in the operational data base, for reasons not accounted for by the simulator, and is not a major cause of concern. The simulated 1977 processability rate is, in fact, probably too low, even though it is very close to the observed rate.

V. SUMMARY

Generally, the evaluations of the simulator indicate that the primary goal of producing reasonable acquisition histories for segments has been attained. While some minor adjustments need to be made, it appears that the simulator will, in fact, be operational by the end of the year and that the simulated output will be of a realistic and useful character.

REFERENCES

1. Horn, T. N.; Brown, L. E.; and Anonsen, W. H.: Acquisition and Preprocessing of Landsat Data. Proc. Technical Sessions: The LACIE Symposium, vol. I. JSC-16015, NASA/JSC (Houston), July 1979, pp. 147-155.

2. Anonsen, W. H.: End of Project Report for Data Acquisition, Preprocessing, and Transmission Subsystem/Large-Area Crop Inventory Experiment (DAPTS/LACIE). NASA Goddard Space Flight Center (Greenbelt, Md.), 1981, 33 pp.

3. Grebowski, G. J.: 1978: LACIE Registration Processing. Proc. Technical Sessions: The LACIE Symposium, vol. I. JSC-16015, NASA/JSC (Houston), July 1979, p. 87-97.

4. Bean, S. J.; and Somerville, P. N.: Some New Worldwide Cloud Cover Models. J. Applied Meteorol., vol. 20, no. 3, 1981, pp. 223-228.

5. Smith, O. E.; and Somerville, P. N.: World-wide Cloud Cover Model. Fourth NASA Weather and Climate Program Scientific Review, NASA Goddard Space Flight Center (Greenbelt, Md.), 1979.

6. Greaves, J. R.; Speigler, D. B.; and Willard, J. H.: Developing a Global Cloud Cover Model for Simulating Earth-Viewing Space Missions. NASA Contractor Rep. CR-61345, 1971, 141 pp.

7. Local Climatological Data 1974-80. U.S. Dept of Commerce, NOAA Environmental Data Service (Asheville, N.C.).

8. NOAA Operational Heat Budget Archive: Courtesy of NOAA Environmental Data Information Services (Washington, D.C.).

9. Barrett, E. C.; and Grant, C. K.: Comparison of Cloud Cover Evaluated From Landsat Imagery and Meteorological Stations Across the British Isles. ERTS Follow-On Program, Study No. 2962A, Third Quarterly Report, NASA-CR-148216, NASA Goddard Space Flight Center (Greenbelt, Md.), 1976, 45 pp.

10. Dennis, T. B.; Cate, R. B.; Nazare, C. V.; Smyrski, M. M.; and Baker, T. C.: SSG4: An Automated Spring Small Grains Proportion Estimator. Eighth Int. Symposium on Machine Processing of Remotely Sensed Data, LARS-Purdue, Univ. (West Lafayette, Ind.), 1982.

ACKNOWLEDGEMENT

The authors would like to express special thanks to Diana Quinn for her painstaking work in testing and evaluating the simulator.

AUTHOR BIOGRAPHICAL DATA

James H. Smith. Currently, he is engaged in the development and testing of simulation models to evaluate the effects of loss of remotely sensed agricultural data due to atmospheric conditions and in the development of an inexpensive method of using remotely sensed data to estimate the year-to-year change in crop area within a given region. Areas of specialization include simulation and modeling, nonparametric statistics, multivariate statistics, and experiment design. He received a B.S. in mathematics in 1974 from Texas Christian University and an M.S. and Ph. D. in statistics in 1976 and 1980, respectively, from Texas Tech University.

Jane T. Malin. Currently, she is responsible for support to the planning and implementation of remote-sensing-based crop production estimation research activities in the Development and Evaluation Department. Dr. Malin previously served as task leader for technical planning and integration of the corn and soybean consortium project and for sampling and aggregation technology development and evaluation. She organized and led the agricultural information system simulation project. Areas of specialization include statistics and experiment design, modeling and simulation of human information processing, and artificial intelligence. She received a B.A. in psychology from the College of Wooster in 1967 and a Ph. D. in experimental psychology from the University of Michigan in 1973.

Marion Dvorin. At Lockheed, he was involved in the prediction and filtering approach for stationary and nonstationary time series analysis and autocorrelation analysis. Currently, he is employed by Omnitape, Inc., in Houston, Texas. He received an M.S. and Ph. D. in theory of probability and mathematical statistics from Moscow State University (U.S.S.R.) in 1964 and 1972, respectively.

Chih C. Lin. Currently, he is employed with Lockheed Engineering and Management Services Company, Inc., at the Goddard Space Flight Center, after transferring from the Houston office. Areas of interest include nonparametric statistics, binary data analysis, empirical Bayesian statistics, and computer applications. He received a B.S. in mathematics in 1974 from the University of Washington and an M.S. and Ph. D. in mathematical statistics from Texas Tech University in 1977 and 1980, respectively.