GLOBAL CROP PRODUCTION FORECASTING: A SIMULATION ANALYSIS OF THE DATA SYSTEM PROBLEMS AND THEIR SOLUTIONS

Harvey Golden
James W. Neiers

ABSTRACT

Marshall Space Flight Center has been analyzing the NASA long-term objective of providing by 1985 a bi-weekly Global Crop Production Forecast of major crops having worldwide food and/or economic significance. The goal of this analysis was to determine probable data related issues in achieving this objective and the recommendation of alternatives for their resolution.

A quantification of the data volumes and the acquisition expectation for different system configurations are essential to establishing realistic design parameters. The required information must be obtained at precise times during the growing seasons. The opportunities are limited by the orbital mechanics of the satellites by climatological conditions. The use of simulation to combine the dynamics of crop phenology, cloud probabilities, and orbital dynamics to obtain a faithful representation of data acquisition and processing performance is discussed in this paper along with some results.

INTRODUCTION

One of the most significant potentials of space technology is its application in the areas of management of earth resources and of environmental understanding and prediction. An integral part of this potential is an increasing emphasis on commercially viable space data products. As a result, increasing demands for data quantity and quality and more cost effective and timely data products have caused new technology problems with the end-to-end data systems. Marshall Space Flight Center (MSFC) has performed a detailed analysis on several long-term earth applications objectives to determine probable data related issues in achieving these objectives and has recommended alternatives for their resolution. "Global Crop Production Forecasting" was one of those objectives analyzed because of its potential of being a major driver of the data system in the timeframe under consideration.

GLOBAL CROP PRODUCTION FORECASTING

The criticality of this objective is evident; for one of the major problems of the world today is its ability to feed its ever increasing population. Approximately 98 percent of the world's food comes from the land and approximately 2 percent comes from the sea. The best conceivable management can do no more than double this amount of food from the sea; therefore, any increase in available food must come from the land. Most of the World's food comes from grain such as wheat, rice, and corn. The reserve of the world food resource has shrunk from 26 percent of the annual consumption in 1959 to 7 percent in 1974. North America is the only major exporting region in the world and food exports are a major factor in U. S. world trade and balance of payments. Accurate global crop production forecasting would provide better information concerning impending crop failures; thus aiding improved decisions on the transporting and distribution of the available food. Aside from the humanitarian benefits, there are also benefits to national policy and economics. Earlier and more accurate information about world crops could help the U. S. and other countries better manage their agricultural production and minimize fluctuation in price and trade volumes. Grain exports could be planned with less disruption of domestic markets. To provide these benefits, the forecasting system must be timely and accurate.

STUDY APPROACH

Many organizations, including the National Aeronautics and Space Administration, USDA and the United Nations have funded recent studies concerning this significant issue. Specifically, the objective MSFC has been analyzing is to provide by 1985 a bi-weekly Global Crop Production Forecast of major crops having world-wide food and economic significance.

To accomplish the analysis of such a system, it was first necessary to be cognizant of previous existing, and planned related activities. Therefore, MSFC conducted a review of existing literature, a survey of existing and planned technologies and facilities, and interviewed potential users. From this effort information needs were defined and data system requirements were determined. A unique simulation facility, located at MSFC, called the Data System Dynamic Simulator (DSDS) was the primary tool used in the quantification analysis. The resultant requirements were then compared against the existing and planned data systems and deficiencies identified and analyzed. The objective was then further refined, new technology requirements identified and a conceptualized data system synthesized.

STUDY PARAMETERS

In order to obtain the information to make forecasts with the accuracies required, many physical parameters had to be identified and measured. Unfortunately, there is no unanimous agreement among experts as to which parameters should be measured and what influence these parameters have on crop forecasts, or the related data systems requirements. The parameters may be wholly or partially obtained by remote sensing from space, while in-situ measurements, statistical data, and historical data must be used to obtain others. The hypothetical global forecast assumed for this analysis is based largely on the use of multispectral data from a Thematic Mapper located onboard a satellite. The data are screened to select only those with an acceptable amount of cloud cover and are subsequently combined with ancillary data such as crop calendars and historic growing patterns to obtain crop identification and mensuration. Crop identification and mensuration are combined to give crop area. Using satellite, historical, statistical, and ground collected data, the meteorological, hydrological, and crop condition models feed into the agromet model which combines their outputs to determine yield. Yield and area are then combined to give crop production. This flow must be repeated for each different crop, region, and growing condition. All these repetitions are summed to obtain a global production forecast for distribution to the users.

Trade studies were performed as a first step toward defining concrete data systems requirements for collecting and processing the enormous quantities of remotely-sensed data needed to produce the information for a global forecast. The studies had two principals goals: (1) To determine the data load for an operational global crop production forecasting system as a function of data frequency, crop types, their biophases, cloud coverage, and number of satellites; and (2) to investigate and propose alternate strategies (e. g., editing or sampling) for reducing the data load in case it should exceed projected processing capabilities. Considering the complexities and unknowns involved in attaining these goals, certain basic assumptions had to be made. For example, only the front end, the data acquisition and ground preprocessing, of a total end-to-end data system was considered in detail. A system was assumed to be capable of processing 70 full scenes per day, similar to the currently planned GSFC Landsat-D ground preprocessing system (50 full and 50 partial Thematic Mapper scenes per day). Accuracy of classification and yield prediction capability which will be available were presumed to be adequate. Ancillary data, and data from non-space platforms or satellites beyond Landsat-D were not included. These simplifying assumptions, and others, had to be made to establish reasonable bounds within which the analysis could proceed (see Tables 1, 2).

TABLE 1

ASSUMPTIONS

1. SAMPLING STRATEGY EMPLOYED FOR MENSURATION
    A. 1 KM^2 SAMPLE SIZE
    B. 30 METER RESOLUTION/PIXEL SIZE
    C. NO. OF SAMPLES CALCULATED BY REGION
2. MULTI-TEMPORAL OBSERVATIONS REQUIRED
    A. ONE OBSERVATION PER CONFUSER CROP DURING GROWING SEASON
    B. SPECIFIC WINDOWS DEFINED WHERE THE DELTA REFLECTANCE BETWEEN CROPS EXCEEDS THE THRESHOLD OF SENSOR RESOLUTION
3. MAJOR PRODUCERS AND MAJOR CROPS ANALYZED
    A. 22 COUNTRIES - 36 REGIONS
    B. 4 CROPS

TABLE 2

COUNTRIES, REGIONS, CROPS, AND NUMBER OF SAMPLES USED FOR STUDY

Country & Region	Crops	Estimate of # of Operational Sample Segments
Argentina	W,C	1500
Australia	W	1000
Bangladesh	R	1000
Brazil North	C	1000
Brazil South	C,S,R	1767
Canada	W	1000
China North	W,C,S,R	2000
China Central	W,C,S,R	2000
China South	W,C,S,R	2000
Egypt	C	1000
France	W,C	1500
India Punjab	W,C	1500
India Ganges	W,C,R	1767
India Central	W,C,R	1767
India Bilaspur	W,C,R	1767
India Coastal	R	1000
Indonesia	R	1000
Italy	W,C,R	1767
Japan	R	1000
Mexico	C	1000
Pakistan	W	1000
Romania	W,C	1500
South Africa	C	1000
Philippines	R	1000
Thailand	R	1000
Turkey	W	1000
USA - Region A	W,C,S	1500
USA - Region B	W,C	2000
USA - Region C	W,C,S,R	2000
USA - Region D	W,C,S,R	2000
USSR Latvia	W,C	1500
USSR Ukraine	W,C	1500
USSR Transvolga	W,C	1500
USSR Volga-Ural	W,C	1500
USSR Siberia	W,C	1500
Yugoslavia	W,C	1500
TOTAL		51355

W = WHEAT, C = CORN, S = SOYBEANS, R = RICE

Using the DSDS to correlate the interrelated influences of orbital parameters, crop calendars, and cloud conditions, comprehensive sets of global data loading profiles were generated. The effects on the data load of different alternatives for assessing crops and their biophases that were investigated are listed in Table 3. There was emphasis on the number of satellites in orbit. Schemes for reducing the data load through cloud rejection editing and sampling were also studied. All these analyses produced information too extensive to be covered in its entirety in this summary report. However, an attempt has been made to present some of the more significant conclusions along with a more detailed description of the simulation modeling.

TABLE 3

TRADES PERFORMED

OPTIMIZATION OF REQUIRED NUMBER OF SATELLITES

DETAILED

- DIFFERENT REPEAT CYCLES
- INCREASED SWATHWIDTH
- VARIOUS CLOUD CRITERIA

CURSORY

- POINTABLE SENSOR
- VARIOUS NODAL CROSSINGS
- SUPPLEMENTAL COVERAGE
- OVERSAMPLING
- NON-SUN SYNCHRONOUS ORBITS

OPTIMIZATION OF DATA/INFORMATION RATIO

DETAILED

- CLOUD EDITING
- SAMPLE EXTRACTION
- REDUNDANT SAMPLE REMOVAL

CURSORY

- SAMPLE STRATEGIES
- COST TRADE-OFFS

Before discussing these, it should be remembered that the essential question is not what data volumes are required to perform global crop production forecasting, but the quality of the information which can be derived from these data and used in computing area and yield. So, although possible data loads and methods to reduce them were studied, this was done only to establish a base from which to answer the broader question of how to avoid processing data that is at best superfluous and at worst costly and time consuming.

DATA SYSTEM DYNAMIC SIMULATOR

DSDS was used to merge conditional events with probabilistic events to measure the performance of alternative data systems. The conditional events, while being deterministic, were not suitable for a strictly mathematical analysis because of their large quantity and the variations resulting from minor changes in the candidate systems. The magnitude of the problem was further increased when the probabilistic effects were introduced.

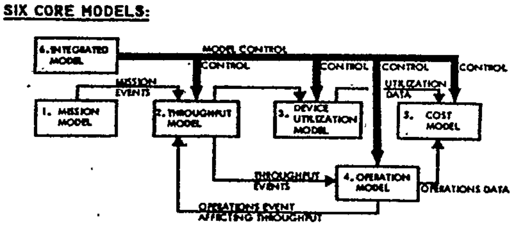
The objective of this simulation was to determine the success to be expected in obtaining requisite data under variations in the acquisition system. The conditions for obtaining data were that the area immediately above the sample was within the field of view of a satellite borne sensor. Additionally, the seasonal time was limited by a period of interest which was determined by a crop calendar for that region. These conditions were simulated by a series of models in DSDS. These models are in-place software for which interconnections and parameters are specified to represent the system under investigation.

The DSDS is an MSFC software simulator for measuring the performance of data systems. It is built on the concept of precoded modules for the elements in data systems. A data system model can be constructed by identifying the modules, their associated parameters and their interconnections. These features include system timing, control, sizing, personnel support activities, cost, and external influences. The use of precoded modules minimizes the effort required of the systems analyst but even more importantly, it greatly reduces the amount of validation required which is the traditional nemesis of simulation. In addition to the precoded modules for the data system elements, called DSEM's, there are standard core models: Mission Model, Integrated Model, Throughput/Resource Utilization Model, Device Utilization Model, Operation Model, and Cost Model.

Figure 1 shows the system of six core models and how they relate to each other.

FIGURE 1

DSDS CORE MODELS



1. The Mission Model (MM) is the simulation driver - it generages the sequence of events which cause data to be generated, routed and controlled.

2. The Integrated Model (IM) is the simulation executive - it provides the overall control of the other models.

3. The Throughput/Resource Utilization Model (TRUM) is the mathematics and logic which represents the functional operation of the Data System - it is the model through which the data "flows."

4. The Device Utilization Model (DUM) is the "instrumentation" of the simulation - it provides the history on what happens to the data as it travels through the Data System Model.

5. The Operations Model (OM) is the scheduling and allocation of people to support the flow of data in the Data System Model - it is a model of the supporting human activity.

6. The Cost Model (CM) is the cost estimator for the modeled data system - it provides an estimate of what it costs to buy and operate a given data system.

The DSDS Core Model System is an event oriented simulation system. The Mission Model creates events which "drive" the TRUM. It creates sensor, experiment, and communication events.

These events are sent to the TRUM which generates more events representing data "flow" through the system. Each event in the system carries with it a set of attributes. These attributes carry the information necessary for identification of the events in the TRUM.

All events are stacked in an event file with a designated removal time and destination. Each event removed from the event file is routed to a specific DSEM. This block of code recognizes the event by the attributes it carries with it. The DSEM acts on the event and generates new events as required. These new events are placed in the event file with a specified time for removal and a designated DSEM destination. This removal and replacement process continues until either all events have been removed from the event file and no new events have been generated to replace them, or until a time period specified on input has been reached.

During execution of the TRUM, when events are removed from the event file, processed by the DSEM's and replaced with new events as necessary, another process is accomplished. Each DSEM, as appropriate, gathers statistical data on events which pass through it. This data is stored and is utilized by the DUM for generating reports following the simulation.

From the DSDS generated reports, problems in mission scheduling, bottlenecks in data flow, and cost inefficiencies can be identified. The User can identify any conflicts in sensor scheduling. The computer "flag" apparent conflicts. There is a total of 20 Mission Model reports available to assist the User in analyzing the mission.

Once the mission is fully analyzed, and the User satisfied with the target coverage, the data system is analyzed. From the Throughput/Resource Utilization Model and Device Utilization Model (TRUM and DUM), the User is provided with a set of seven reports from which he can determine where bottlenecks are occurring, how much data is lost, and what elements are not fully utilized.

The Operations Model (OM) is run both concurrently with TRUM and following the running of the DUM. The portion which runs concurrently converts demands and supplies of skills into a work schedule for the various skills and logs skill's utilization resulting from the data flowing in the system. The OM portion running after the DUM, processes the skills utilization statistics for each facility and generates reports on these facilities.

The Cost Model calculates cost of the data system elements. Each Data System Element Model (DSEM) has associated cost parameters from which costs are calculated. Compiled costs are output in three reports detailing cost information about each data system element, operating cost both for manpower and material, and summary data in a time spread presentation including design development test and evaluation costs.

An additional feature of DSDS permits the construction and retention of higher level Data System Models (DSM's) using the DSEM's and the capabilities of the core models. These DSM's can be used as additional higher level building blocks for subsequent simulations. Again, this feature greatly reduces the modeling time and the validation required.

APPLICATION

The sensor platform position and the sensor field of view was determined as a function of time for the various satellite orbits using the Mission Model. The Mission Model includes both a

Mission Ephemeris Generator (MEG) and a Target data base. The MEG calculates the position of the satellite and the nadir coordinates in terms of latitude and longitude at each finite time increment. The agricultural mission uses an orbit that repeats every 16 days. For this study, the ephemerides were determined for one repeat cycle and repeated as necessary throughout the simulation.

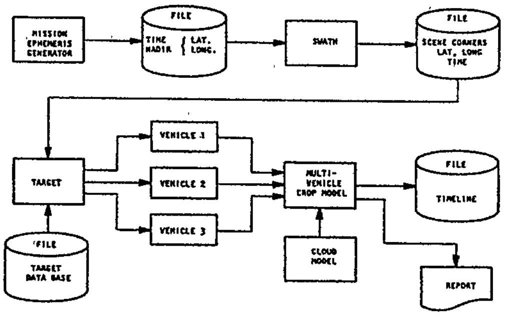
The first step in generating the ephemerides was to determine the nominal altitude and inclination to obtain a sun-synchronous integral repeat cycle. The current Landsat-D orbit is at 704 KM and repeats every 16 days. This is nearly the minimum achievable altitude that will not decay too rapidly because of drag. We looked at other orbits that repeated less frequently as well as some that provided only partial coverage of the earth's surface.

Starting with the nominal altitude and inclination calculated from the closed form formula we ran the MEG and adjusted the altitude to obtain a repetition in the ground trace. For the 16-day Landsat D orbit, we achieved a repetition of 735 meters after 233 orbits. The accuray of the repetition is a function of the number of steps used in the MEG and the number of orbits required to repeat. A trade-off decision was made to accept the 735 meter repeat error rather than increase the computer time for calculating the mission ephemerides.

The sequence of steps followed during the study are depicted in Figure 2. For economy of computer time, the MEG was only run for one repeat cycle for each insertion latitude and time. The mission ephemerides, in terms of nadir latitude and longitude, were generated in along track increments of 90 nautical miles, which corresponds to the along track spacing of standard Landsat scenes. This permitted a direct accounting of scenes as a measure of data volume that would be suitable for comparison with previous studies.

FIGURE 2

SIMULATION SYSTEM FOR GCPF



The data generated by MEG was used to generate pairs of latitude and longitude corresponding to the edge of the sensor field of view. One of the system alternatives investigated was the benefits of using a wider swath sensor. It was not necessary to repeat the MEG runs for each change in swath width.

The DSDS will recognize the existence of Targets within a defined sensor field of view. As part of this study, twenty-two countries were chosen based on the criteria that each contributed two percent or more to the world harvest of wheat, corn, rice, or soybeans. The larger countries were divided into geographic regions corresponding to statistical reporting districts. The point target capability of DSDS was used to measure variations due to geographic location and cloud conditions. A minimum number of samples in a simulation region was set at 30 for a geographic region containing only one crop. The number of samples was adjusted to a maximum of 60 when all four crops were grown in a region.

The samples were randomly distributed within each statistical reporting region. Each sample had a corresponding latitude and longitude designation. This approach resulted in 1553 sample targets for the world. These samples constitute the target data base that was used by the target model to identify potential sample acquisition for each new time step. The output of this model consisted of a time correlated file of potential data acquisition during the repeat cycle time. This data was saved for repeated use as input to the crop model. A typical report from the Target Model shows how many times each target was seen during a repeat cycle and summarized the results. Some targets are observed several times during a repeat cycle. This is a result of overlap between adjacent swaths, which, for a 185 km swath, reaches 100% at 57.46° latitude.

The time during which the data is obtained is important for forecasting crop production. The contrast between crops of interest and confusion crops reach maximum during certain times during the growing cycle. An early observation of the soil background is necessary before the effects of the plant canopy are predominant. Some specie classification is best performed by detecting flowering or by harvest time. All these temporal considerations impose additional constraints on the data acquisition time. These effects were handled in the crop model by a set of crop calendars. Each species and homogenous region had a crop calendar. It defined time periods or windows during which useful data could be acquired. Each target was assigned to one of the crop calendars.

The multivehicle crop model inputs the files of target acquisition for up to three vehicles. For each test case, the simulation was a full year. When the end of the repeat cycle was reached on any of the input files, the file was rewound and re-read until the full year was covered.

After a scene record was read from each satellite's input file, the following steps were performed.

a. A test was made to determine which scene occurred first.

b. A check was made to determine if any of the targets in the scene had crops in a growth stage of interest, i.e., an active window. This was determined by checking each target in the scene with its corresponding crop calendars. If none of the targets were active, the next scene was read from the corresponding input file and Step A was repeated.

c. If any of the targets in the scene were in an active window, the Cloud Model was called to determine the cloud category for the scene based on the current month, time of day and the cloud region.

The cloud model is based upon the Allied Research Association's statistics. The world is divided into 30 homogenious cloud regions for which the statistics were scaled to the 90 nm area of a standard Landsat scene. Each target is assigned to one of the cloud regions. The first target encountered in a scene was used to set the scene cloud region.

d. For each target in an active window, the number of opportunities for observation was incremented.

e. The scene cloud cover was compared with the scene cloud cover threshold for processing. This threshold was set as part of the editing variation at the start of the crop model simulation. If the scene cloud cover was below the threshold, a test was made on each target in the scene. The percent of targets that could be obtained from a scene is a function of the scene cloud cover as shown below. A separate test was performed for each target. If the target was clear, the number of clear observations was incremented.

Scene Cloud Category	Percent of Cloud Cover In the Scene	Percent Of Clear Samples Obtainable
1	<10	95
2	10, 20, 30	80
3	40, 50	55
4	60, 70, 80, 90	25
5	100	0

f. The Target Processing Timeline was created. For each clear observation, a test was made to see if the target had been seen for the desired number of times during the current window. If the desired number of clear observations had not been obtained, a record was written in the time line file with the following information:

- Target number
- Date
- Crop(s) of interest
- Window number

g. The Scene Processing Timeline was created. For each scene with at least one target in an active window, the following information was obtained:

- Date
- Region number
- Cloud region
- Scene cloud category

ALTERNATIVES ANALYZED

Number of Satellites

Four alternatives for optimizing the number of satellites were investigated. They all have the property of using the currently designed Landsat-D Satellite, in some cases with modification to the sensor system or operational procedures. The four systems with their characterizing features are illustrated in Figure 3.

FIGURE 3

FOUR CANDIDATE AGRICULTURAL DATA COLLECTION SYSTEM

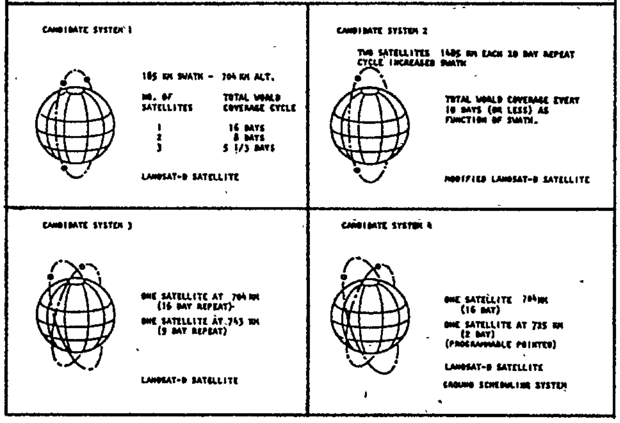


Table 4 gives a comparison of the results achieved with 1, 2, and 3 satellites. The mean and standard deviation for the percent of samples for which cloud free observations were missed are presented in Table 5. The two satellites data is based on 20 simulation runs and the three satellite data is based on 7 runs. The 50% and 95% confidence limits for obtaining samples were determined for this data. The confidence levels correspond to cloud free samples.

TABLE 4

PERCENT OF SAMPLES ACQUIRED FOR EACH TEST CASE

COLUMN NUMBER	1	2	3	4	5	6	7	8	9	10
NUMBER OF VEHICLES	1	1	1	1	2	2	2	2	3	1
VEHICLE ALTITUDE	704	704	704	704	704	704	704	704	704	1485
SWATH WIDTH (KM)	185	185	222	222	185	185	222	222	185	185
MAXIMUM CLOUD COVER ACCEPTED (%)	50	90	50	90	50	90	50	90	50	90
NO. OF RUNS AVERAGED	7	1	1	1	20	1	1	1	7	2
WORLD AVERAGE	87.11	92.14	92.31	94.91	96.99	98.84	97.88	99.30	99.18	83.12
ARGENTINA	93.3	97.8	98.9	98.5	99.6	100.0	100.0	100.0	100.0	92.8
AUSTRALIA	95.0	100.0	98.3	90.0	99.8	100.0	100.0	100.0	100.0	91.7
BANGLADESH	84.4	95.6	93.3	96.7	94.3	100.0	94.4	98.9	99.7	82.8
BRAZIL - NORTH	100.0	96.7	96.7	100.0	99.9	100.0	100.0	100.0	100.0	90.0
BRAZIL - SOUTH	99.1	98.1	100.0	98.1	99.5	100.0	100.0	100.0	100.0	92.5
CANADA	96.7	96.7	100.0	100.0	99.8	100.0	100.0	100.0	100.0	96.7
CHINA - NORTH	85.8	88.3	91.2	96.7	97.4	100.0	100.0	100.0	99.2	85.8
CHINA - CENTRAL	76.7	90.8	78.3	83.3	91.6	92.5	96.7	97.5	98.1	71.2
CHINA - SOUTH	87.5	90.0	90.0	96.7	98.8	98.3	100.0	100.0	100.0	91.7
EGYPT	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.3
FRANCE	86.7	97.8	92.8	95.6	98.5	100.0	100.0	100.0	99.8	86.7
INDIA - PUNJAB	87.8	98.9	96.7	98.9	99.2	100.0	100.0	98.9	100.0	91.7
INDIA - GANGES	75.0	85.4	95.8	98.1	91.9	98.1	86.8	98.1	98.1	63.0
INDIA - CENTRAL	72.3	90.6	85.8	94.3	91.7	95.6	96.2	97.5	97.6	71.1
INDIA - BILASPUR	82.4	89.9	84.3	88.7	94.8	97.5	95.6	100.0	96.3	66.4
INDIA - COASTAL	85.6	87.8	92.2	91.1	97.8	100.0	97.8	100.0	99.5	88.3
INDONESIA	85.0	91.1	83.3	88.9	96.6	100.0	96.7	98.9	99.4	75.6
ITALY	86.9	93.5	94.9	94.4	98.7	100.0	100.0	100.0	99.9	88.3
JAPAN	95.0	93.3	98.3	96.7	99.7	100.0	100.0	100.0	100.0	91.7
MEXICO	97.2	98.9	95.6	95.6	99.6	100.0	100.0	100.0	100.0	92.2
PAKISTAN	93.3	80.0	98.3	100.0	98.3	100.0	100.0	100.0	100.0	91.7
ROMANIA	91.1	91.1	88.9	94.4	98.8	100.0	100.0	100.0	100.0	86.1
S. AFRICA	90.0	93.3	88.3	93.3	99.3	100.0	100.0	100.0	100.0	90.0
PHILIPPINES	91.7	96.7	91.7	91.7	97.8	100.0	96.7	98.3	99.8	83.3
THAILAND	91.7	96.7	96.7	100.0	99.8	100.0	100.0	100.0	100.0	93.3
TURKEY	93.3	80.0	91.7	96.7	98.5	100.0	100.0	100.0	100.0	78.3
US - A	80.0	80.0	80.4	90.4	95.4	98.3	98.3	99.1	97.9	73.5
US - B	93.0	93.0	97.2	98.6	98.9	100.0	100.0	100.0	100.0	88.7
US - C	84.6	87.9	95.3	97.2	96.1	97.2	97.2	100.0	99.6	85.5
US - D	77.0	84.7	86.7	85.7	94.8	96.9	96.9	98.0	96.5	74.0
USSR - LATVIA	96.7	100.0	97.8	98.9	99.9	100.0	100.0	100.0	100.0	97.8
USSR - UKRAINE	96.7	96.7	100.0	100.0	99.8	100.0	100.0	100.0	100.0	97.8
USSR - TRANS-VOLGA	97.8	95.6	97.2	100.0	99.8	100.0	100.0	100.0	100.0	97.8
USSR - VOLGA-URAL	98.3	96.7	98.3	98.9	99.9	100.0	100.0	100.0	100.0	93.3
USSR - SIBERIA	94.4	97.8	93.3	100.0	99.2	97.8	100.0	100.0	99.7	86.7
YUGOSLAVIA	86.7	95.6	97.2	97.8	98.0	100.0	100.0	100.0	99.7	79.4

TABLE 5

MEAN AND STANDARD DEVIATION FOR THE PERCENT OF SAMPLES MISSED*

COUNTRY/REGION	2 LANDSAT-D VEHICLES MEAN	2 LANDSAT-D VEHICLES STD. DEV.	3 LANDSAT-D VEHICLES MEAN	3 LANDSAT-D VEHICLES STD. DEV.
EGYPT	0	0	0	0
AUSTRALIA	0.17	0.74	0	0
CANADA	0.17	0.74	0	0
THAILAND	0.17	0.74	0	0
USSR - LATVIA	0.06	0.25	0	0
USSR- VOLTA-URAL	0.06	0.25	0	0
USSR - UKRAINE	0.17	0.41	0	0
USSR - TRANS-VOLGA	0.17	0.54	0	0
USSR - SIBERIA	0.78	1.30	.32	.84
BRAZIL - SOUTH	0.09	0.42	0	0
BRAZIL - NORTH	0.50	1.22	0	0
JAPAN	0.33	1.02	0	0
ARGENTINA	0.44	0.66	0	0
MEXICO	0.44	0.66	0	0
S. AFRICA	0.67	1.74	0	0
ROMANIA	1.17	1.22	0	0
ITALY	1.26	1.10	.13	.35
FRANCE	1.50	1.41	.16	.42
TURKEY	1.50	2.29	0	0
PAKISTAN	1.67	2.54	0	0
YUGOSLAVIA	2.00	2.06	.32	.54
PHILIPPINES	2.25	2.61	.64	.87
INDONESIA	3.44	2.52	.64	.87
USA - REGION B	1.13	1.18	0	0
USA - REGION C	3.88	1.15	.40	.50
USA - REGION A	4.65	2.26	2.11	.99
USA - REGION D	5.15	2.26	3.50	1.16
CHINA - SOUTH	1.25	1.70	0	0
CHINA - NORTH	2.64	3.57	.83	.83
CHINA - CENTRAL	8.42	2.27	1.90	1.15
BANGLADESH	5.66	5.72	.32	.54
INDIA - PUNJAB	0.78	1.57	0	0
INDIA - COASTAL	2.16	1.51	.48	.59
INDIA - BILASPUR	5.19	3.67	3.69	3.77
INDIA - GANGES	8.07	6.61	1.14	.81
INDIA - CENTRAL	8.30	4.40	2.43	1.97
WORLD	3.01	0.50	.82	.30

*FAILURE TO OBTAIN 1 CLOUD FREE OBSERVATION DURING THE DESIGNATED TIME FOR EACH SAMPLE SEGMENT CONSTITUTES A MISS.

Effect of Increased Swath Width

Comparison runs for a 20 percent increase in swath width were made for candidate systems 1 and 2. The percent of samples acquired, the increase in samples acquired due to increased swath width, and the reduction in the number of samples missed for each case are tabulated below in Table 6.

TABLE 6

EFFECT OF 20% INCREASED SWATH WIDTH

No. of Satellites	Alt. KM	Samples Acq. Percent 185 KM Swath	Samples Acq. Percent 222 KM Swath	Samples Acq. % Increase	Samples Missed % Reduction
1	704	87.1	92.3	6.0	40.3
2	704	97.0	97.9	0.9	30.0
1	1485	83.1	88.9	7.0	34.3
2	1485	96.5	97.8	1.3	37.1

The effect of increased swath width is shown in Figure 4 for one satellite and in Figure 5 for two satellites. Figures 4 and 5 give a quick appreciation for the benefits of increased cloud cover acceptance and swath width. Two satellites with increased swath and cloud cover acceptance perform slightly better than the nominal three satellite case.

FIGURE 4

EFFECT OF INCREASED SWATH AND HIGHER CLOUD COVER ACCEPTANCE FOR ONE LANDSAT-D SATELLITE

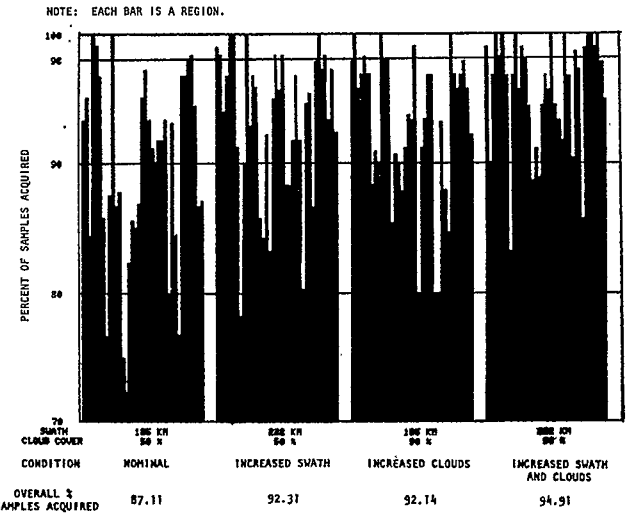
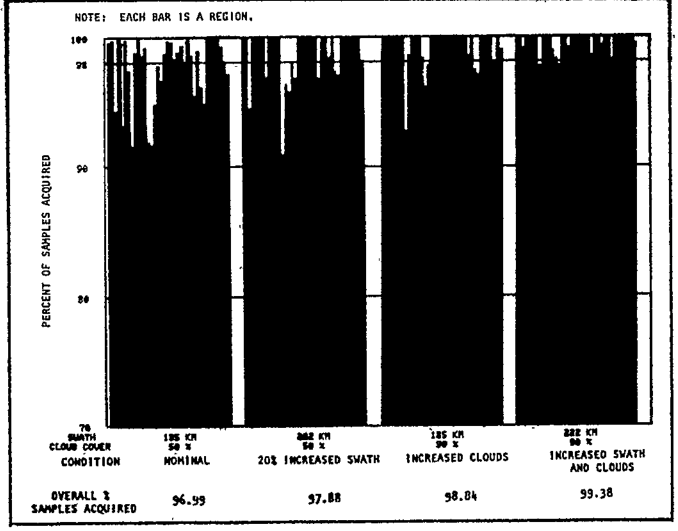


FIGURE 5

EFFECT OF INCREASED SWATH AND HIGHER CLOUD COVER ACCEPTANCE FOR TWO LANDSAT-D SATELLITES


NOTE: EACH BAR IS A REGION.

Effect of Increased Cloud Cover Acceptance

The effect of accepting scenes with up to 90% cloud cover was investigated for one and two Landsat-D satellites. The results from the four cases compared are as shown below in Table 7.

TABLE 7

No. of Satellites	Max. Cloud Cover Accept. (%)	Samples Acquired (%)	Samples Acquired % Increase	Samples Missed % Reduction
1	50	87.11	-	-
1	90	92.14	5.8	29.0
2	50	96.99	-	-
2	90	98.84	1.9	61.5

The effect of increased cloud cover on the processing load is shown below in Table 8.

TABLE 8

Cloud Category Accepted	Max. Cloud Cover Accept. %	Samples Acquired %	NO. OF SCENES TO BE PROCESSED Total For Year	Peak Day	Average For Peak Period
2	30	94.09	4790	47	36 (26)
3	50	96.99	6383	59	48 (34)
4	90	98.84	9317	85	70 (50)

5 Day Work Week
(7 Day Work Week)

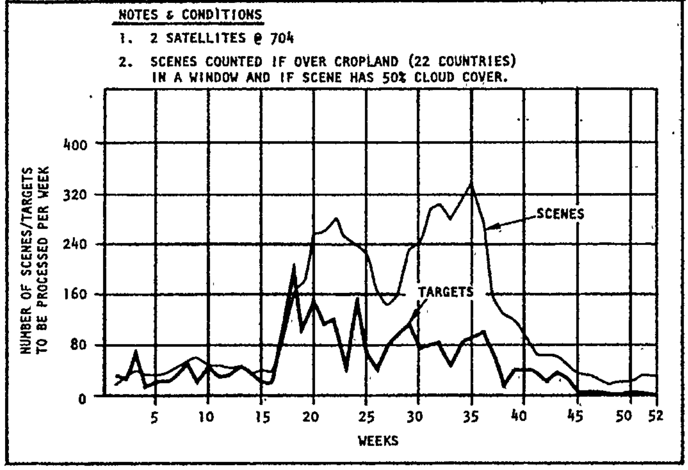
The average number of scenes to be processed per day during the peak growing season from May to September is based on processing only one observation per window. This was chosen to measure the sensitivity of obtaining samples versus the number of satellites.

Yearly Loading for Agricultural Usage

The sizing and configuration of an operational processing system is dependent upon the time line of data acquisition. Two convenient measures of data volume are acquired scenes and the number of sample segments. A scene represents the 38,027,776 pixels as would comprise an area 185 KM sq. at 30 meters spatial resolution. Each pixel represents effectively 5.0625* bytes of data. A sample segment represents 1600 pixels in a desired area one KM square with 44% excess pixels to allow for registration error. A yearly plot representing weekly scene and target acquisition is presented in Figure 6. This data was obtained for the condition of two satellites in a 704 KM orbit, 185 KM swath, an acceptance criteria of 50% or less cloud cover on a scene, and the acceptance of only one observation per target per window. The targets plotted are the simulation targets with a base number of 30 per single crop region. Each 30 targets represents a larger number of samples in an operational system.

FIGURE 6

WEEKLY SCENE AND TARGET ACQUISITION FOR ONE YEAR


NOTES & CONDITIONS
1. 2 SATELLITES @ 704
2. SCENES COUNTED IF OVER CROPLAND (22 COUNTRIES) IN A WINDOW AND IF SCENE HAS 50% CLOUD COVER.

---

*Spectral resolution is 256 levels of one byte/band for five bands at 30 meters spatial resolution plus one band at 120 meters resolution or $(30/120)^2 = .0625$.

The actual number will depend on the crop densities for the region. A figure of 1,000 samples is a reasonable average and was used in determining processing loads.

Processing System Data Volumes

The total number of pixels per year to be subjected to the differing processes was determined for the different candidate systems. This information is shown in Figure 7. What is readily apparent is the small percentage, less than 0.1% of agricultural information in the acquired data. This factor indicates a need to extract the useful information from the data at the earliest practical time in the sequence of agricultural data processing. The two columns of numbers, one headed "Extract Agricultural Samples" which are all of the samples in the accepted scenes, and the other column headed "Edit Cloud Samples," which are only those that are in the clear part of the accepted scenes. The latter column is the yearly total in millions of pixels that can be expected for the different acquisition systems represented by each row in the figure.

FIGURE 7

A COMPARISON OF SATELLITE EFFECTS ON DATA VOLUME FOR PROCESSING

CONDITION NO. OF SAT.	ALTITUDE IN KM		ACQUIRE IMAGE DATA USING SATELLITES	EDIT CLOUDY SCENES *	PREPROCESS SCENE DATA	EXTRACT AGRICULTURAL SAMPLES	EDIT CLOUDY SAMPLES	PROCESS NEW INFORMATION	INFORMATION AS % OF ACQUIRED DATA	PERCENT OF DESIRED SAMPLES ACQUIRED
1	704	185	189568		119331	424.7	320.4	136.0	.0717	87.11
2	704	185	381609		242731	854.1	646.0	151.2	.0396	96.99
3	704	185	563724		358450	1273.3	958.7	154.8	.0275	99.15
1	1485	185	164166		102789	365.9	281.9	130.0	.0792	83.12
2	1485	185	327952		207251	731.4	548.7	150.6	.0459	96.51
1	704	222	215199		134086	499.0	392.2	144.3	.0671	92.48
2	704	222	429258		271443	1012.7	755.7	152.6	.0355	97.78
1	1485	222	267995		168113	442.8	344.6	140.4	.0524	88.88
2	704 & 743	185	375106		239005	828.7	628.5	150.8	.0402	96.65
2	704 & 725	185	387427		243416	890.2	668.2	137.3	.0354	87.97

*MAXIMUM OF 50 PERCENT CLOUD COVER IN A SCENE ACCEPTED FOR PREPROCESSING.

While the usefulness of multiple observations of the same sample segment is acknowledged, this study placed a premium on the value of the first observation of a sample segment during each of the specified windows. Those observations are represented in the column headed "Process New Information."

Figure 8 includes the data volumes within the agricultural processing system when oversampling is used. The oversampling numbers were determined using the regionally weighted data obtained by simulation. The number of extra samples required for each region was determined based upon the mean plus 1.65 sigma of those samples missed in each region.

FIGURE 8

A COMPARISON ON EDITING EFFECTS ON DATA VOLUME FOR PROCESSING

CONDITION NO. OF SAT.	ALTITUDE & SWATH IN KM	ACQUIRE IMAGE DATA USING SATELLITES	EDIT CLOUDY SCENES *	PREPROCESS SCENE DATA	% OS	EXTRACT AGRICULTURAL SAMPLES	EDIT CLOUDY SAMPLES	PROCESS NEW INFORMATION	INFORMATION AS % OF ACQUIRED DATA	PERCENT OF DESIRED SAMPLES ACQUIRED
1	704 Alt. 185 Swath	189568	50	119331	0	424.7	320.4	136.0	.0717	87.11
					33.5	567.0	427.8	156.0	.0823	100.00
			90	176259	0	622.1	382.8	143.8	.0759	92.14
					18.1	734.7	452.1	156.0	.0823	100.00
2	704 Alt. 185 Swath	381609	30	182153	0	640.4	518.9	146.8	.0385	94.20
					14.8	735.2	595.7	156.0	.0409	100.00
			50	242731	0	854.1	646.0	151.2	.0396	96.99
					8.2	924.1	699.0	156.0	.0409	100.00
			90	354305	0	1246.2	746.9	154.2	.0404	98.84
					3.2	1286.1	770.7	156.0	.0409	100.00
3	704 Alt. 185 Swath	563724	50	358450	0	1273.3	958.7	154.8	.0275	99.15
					2.2	1301.7	980.1	156.0	.0277	100.00

* MAXIMUM PERCENTAGE OF CLOUD COVER IN A SCENE ACCEPTED FOR PREPROCESSING.

OS INDICATES OVERSAMPLING TO OBTAIN 100% OF DESIRED SAMPLES 95% OF THE TIME.

An obvious way to reduce the data volume is to reject images of less than some established quality, as determined by the percentage of cloud cover in the scene. This is the currently accepted operational approach. While it reduces the amount of unproductive activity of preprocessing data that will ultimately be rejected as unusable, it has the detrimental quality of also rejecting some highly desirable data. For the data shown in Figure 6. the acceptance criteria was established at 50%. The subsequent numbers represent data that could be expected in the 50% clear portion of the scenes accepted for preprocessing. Thus, the two columns of numbers; one headed by "Extract Agricultural Samples" which are all of the samples in the accepted scenes, and the other column headed "Edit Cloud Samples," which are only those that are in the clear part of the accepted scenes. The latter column is the yearly total in millions of pixels that can be expected for the different acquisition

Scene Cloud Editing

The concerns of bias resulting from oversampling leads to a desire to minimize the oversampling requirement. The oversampling percentages are indicated for each line in Figure 8. The dramatic reduction in the percentages is evident when the samples are extracted from up to 90% cloudy scenes. This indicates a very real trade-off in processing costs versus either additional satellites or oversampling bias. The processing penalty is in the data volume to be preprocessed. Unfortunately, the limits on the amount of oversampling permitted without excessive bias are not currently available to guide the assessment of the worth of the additional processing. Fortunately, as the processing is moved closer to the source, including the ultimate of onboard the vehicles, the concept of editing cloudy scenes vanishes with the entire scene concept.

Comparison of Processing Costs

The previous discussion focused on processing variations in the maximum percent of cloud cover acceptable in a scene for preprocessing and the extraction of agricultural samples. Some dollar estimates are presented in this paragraph as a means for comparing the relative costs of each approach. The basis for comparison is the peak processing required during any two week period. Peak load impacts the processing system sizing and cost. The peak two week load was determined for the two 704 KM satellite acquisition period and the subsequent loads in successive downstream processes were determined for three cloud editing strategies.

The rational used was to develop a system concept for the nominal case and normalize the gross order of magnitude cost on a per megapixel basis. The acceptance of 50% cloud covered scenes for processes were determined for three cloud editing strategies. The acceptance of 50% cloud covered scenes for preprocessing was considered nominal.

The relative processing costs for the 30, 50 and 90 percent cloud cover acceptance is illustrated in Figure 9. The agricultural processing is shown for the situation of processing all usable information through the models and only one sample set per window. In all cases the samples are square 1.2 KM on a side.

The relative performance of each strategy is measured as a percentage of targeted samples obtained. For a performance improvement from 94 to 99 percent, the relative processing cost increases from 236K to 517K. This provides a strong argument to evaluate the efficacy of oversampling.

The relative processing costs for each of the editing strategies and oversampling are portrayed in Figure 10. The results of processing oversamples do not appreciably affect the processing cost. By the measure of performance used, each strategy is equal. The biasing effects of oversampling needs to be further investigated before a conclusion on the most desirable editing strategy can be stated.

FIGURE 9

DATA FLOW & COST COMPARISON

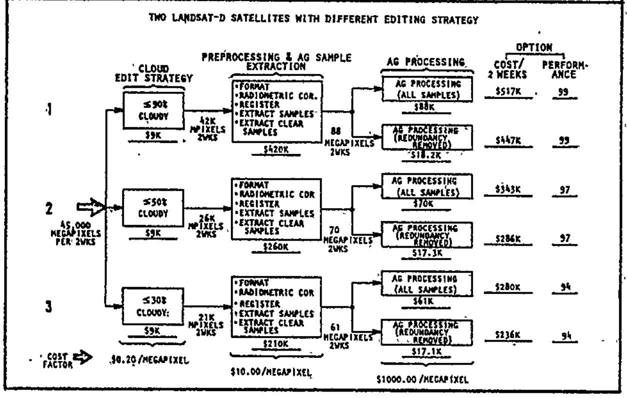
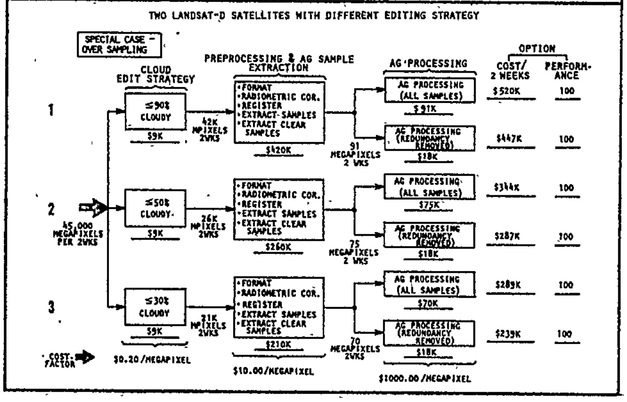


FIGURE 10

OVERSAMPLED DATA FLOW & COST COMPARISON



SUMMARY OF CONCLUSIONS

This study supports the conclusion that the objective of Global Crop Production Forecasting is realistically achievable. It also provides some guidelines for achieving the goal. The following conclusions apply to meeting the data acquisition requirements:

- Use 2 satellites of the Landsat D Type.
- Use equal spacing of the satellites to achieve effective 8 day coverage.
- Dedicate one of the satellites to the agricultural objective.
- Increase The Thematic Mapper sensor swath width by approximately 20 percent on the dedicated satellite.
- Employ oversampling techniques if the bias errors can be accommodated.
- Extract the samples before any scene data is rejected.

Additionally, some approaches to achieving the objective with one satellite are possible. They are:

- Increase the sensor swath significantly.
- Apply on-board sample extraction
- Use a floating sample approach
- Selectively sacrifice accuracy and coverage requirements for selected problem areas.

REFERENCES

1. General Electric Space Division, "Global Crop Production Forecasting - An Analysis of the Data System Problems and Their Solutions," Final Report, Huntsville, Alabama, GE Document No. 78HV031, May 8, 1978.

2. Castruccio, Dr. Peter, "Final Report on Office of Application Objective Analysis" No. ECO-3, 1978.

3. Downs, Dr. Sanford W., Paul A. Larsen, and Dietwald A. Gerstner, "Analysis of Data Systems Requirements for Global Crop Production Forecasting in the 1985 Time Frame," NASA Technical Paper 1164, February 1978.

4. Hooper, James W. and Dale W. Rowe, "Data System Dynamic Simulation - A Total System for Data System Design Assessment and Trade Studies," Record of Proceedings, The Eleventh Annual Simulation Symposium, Tampa, Florida, March 15-17, 1978.