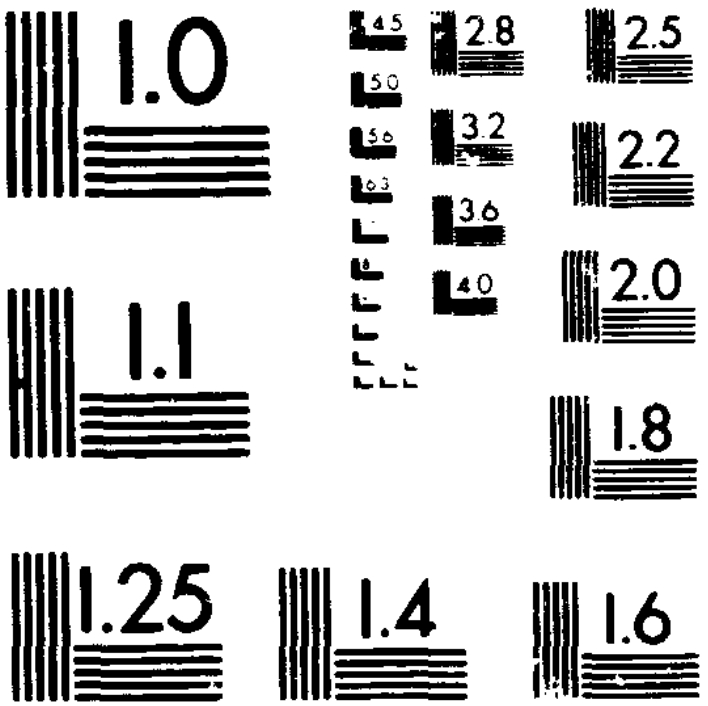


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

FINAL REPORT: E83-10367

Monitoring the Defoliation of Hardwood Forests in Pennsylvania Using Landsat

TM-85362

Compiled and Edited by:

C.L. Dottavio
R.F. Nelson
D.L. Williams

NASA/Goddard Space Flight Center
Earth Resources Branch

(E83-10367) MONITORING THE DEFOLIATION OF HARDWOOD FORESTS IN PENNSYLVANIA USING LANDSAT Final Report (NASA) 104 p
HC A06/MF A01 CSCL 02F G3/43

N83-31067
Unclas
00367

March, 1983

LIST OF PARTICIPANTS

NASA/Goddard Space Flight Center

Dr. C. Lisette Dottavio†
Mr. Ross F. Nelson†
Mr. Darrel L. Williams†

Pennsylvania, Bureau of Forestry

Mr. John Quimby†

Support Assistance

Pennsylvania State University:
Dr. Brian Turner†
Mr. George Baumer
Mr. Fred Luce

Computer Sciences Corporation:
Mr. Mark L. Stauffer†
Ms. Sonia Acrouny Russo†
Mrs. Ruth Kennard

Jet Propulsion Laboratory:
Mr. Nevin Bryant
Mr. Ronald McLeod

†Contributors to text

PREFACE

This report is the final documentation of all research and development activities which were conducted during a 3-1/2 year Joint Research Project (JRP) between NASA/Goddard Space Flight Center and the Pennsylvania Bureau of Forestry/Division of Forest Pest Management. The project was initiated in October 1979 to develop an automated system for gypsy moth defoliation assessment in Pennsylvania using Landsat multispectral scanner data and digital processing techniques.

This report has been structured to conveniently serve the needs of two distinct reader audiences; namely, those interested in a brief, overall summary of accomplishments versus those who desire detailed, quantitative information on how and why certain decisions were made. The overall summary of accomplishments is presented in the first 29 pages of text. At various points within the text, the reader is directed to any one of eight appendices if a more detailed discussion of a specific approach and/or result(s) is desired.

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

BOF/DFPM	- Bureau of Forestry/Division of Forest Pest Management
CIR	- Color Infrared (photography)
DDT	- 1, 1, 1 - Trichloro -2, 2 - di (4-chlorophenyl) ethane
EROS	- Earth Resources Observation System
GRDS	- Ground Reference Data Set
GSFC/ERB	- Goddard Space Flight Center/Earth Resources Branch
IDIMS	- Interactive Digital Image Manipulation System
JPL	- Jet Propulsion Laboratory
JRP	- Joint Research Project
MSS	- Multispectral Scanner
NASA	- National Aeronautics and Space Administration
ORSER	- Office for Remote Sensing of Earth Resources
PSU	- Pennsylvania State University
RVI	- Ratio Vegetation Index
TVI	- Transformed Vegetation Index
USGS	- United States Geological Survey
UTM	- Universal Transverse Mercator
VICAR	- Video Image Communication and Retrieval System

JOINT RESEARCH PROJECT OVERVIEW

Over the last twenty years, the gypsy moth caterpillar (Lymantria dispar) has become one of the most serious threats to the northeastern hardwood forests of the United States. Millions of hectares of woodland throughout New England, New York, New Jersey, Pennsylvania and portions of West Virginia and Maryland have been defoliated during the insect's periodic epidemic population outbreaks.

In the early 1970's, remote sensing scientists identified these major forest disturbances on Landsat multispectral scanner (MSS) imagery. Since that time, research scientists within the Earth Resources Branch of NASA's Goddard Space Flight Center (NASA/GSFC) have been developing image processing techniques that facilitate the use of satellite data to assess forest damage resulting from major insect infestations. These techniques were designed to augment existing surveillance procedures.

The success of these satellite-based studies at Goddard, and the increased threat of gypsy moth defoliation to Pennsylvania forests led to the initiation of a Joint Research Project (JRP) between the GSFC/Earth Resources Branch (GSFC/ERB) and Pennsylvania Bureau of Forestry/Division of Forest Pest Management (BOF/DFPM) in October, 1979. The JRP was designed to develop an automated system for gypsy moth defoliation assessments in Pennsylvania using Landsat multispectral scanner data and digital processing techniques.

The project lasted 3-1/2 years. During the first 2-1/2 years of the project, key elements of the satellite based system were identified, studied and developed by project personnel. The key elements of the operational system included the following:

1. An accurate, cost effective, and timely analysis procedure for defoliation assessment;
2. A statewide data base for storage and retrieval of survey data;
3. An interactive, automated data processing system that allowed timely assessments of defoliation using the selected analysis procedure with the

statewide data base. This processing system was designed such that non-remote sensing personnel could easily use the system with little training.

During the final year of the JRP, this satellite based system was implemented at the Pennsylvania State University for access by DFPM personnel. At that time, foresters and entomologists used this sytem to complete the 1981 defoliation assessment for Centre County and Perry County, Pennsylvania. This activity demonstrated the successful development, implementation, and utility of a satellite-based forest insect defoliation assessment system.

Throughout the JRP, other, smaller scale studies were completed to document the accuracy of satellite-based assessments, cost-benefits, time constraints on satellite-based assessments, and effective data handling procedures. These studies, as well as the key elements of the operational system, are described within this document.

BACKGROUND

Gypsy Moth Defoliation - The Consequences

The gypsy moth caterpillar (Lymantria dispar) is currently one of the most serious forest pests in the northeastern United States. The insect, which is native to Europe and Asia, was introduced to Medford, Massachusetts in 1969 by a French scientist hoping to produce a new variety of silkworm. During this experimentation, several caterpillars escaped and became established in the surrounding woodland. Today, the gypsy moth is widespread throughout New England, New York, New Jersey, Pennsylvania, and portions of West Virginia and Maryland (see Figure 1). Throughout the insect's period of establishment, the gypsy moth has demonstrated the capability to periodically increase its population to epidemic proportions. Currently, the northeasterr U.S. is experiencing one of the largest outbreaks ever recorded.

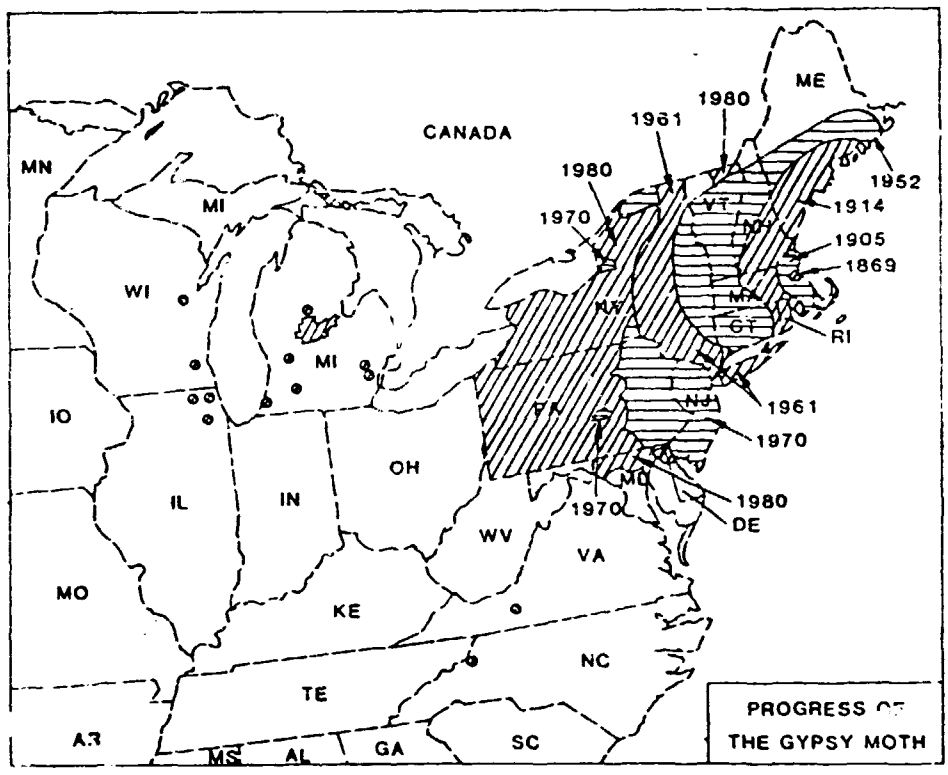


Figure 1. Extent of gypsy moth spread in the northeastern United States (from Marshal, 1981).

Gypsy moth caterpillars damage trees by feeding on foliage. This feeding begins shortly after the caterpillars hatch from their eggs in late April or early May. Defoliation is usually not noticeable until early to mid-June, unless the gypsy moth populations are unusually large. In late June and early July, the heaviest defoliation takes place as the caterpillars reach full size, approximately two inches metric, in their fifth (male) and sixth (female) instars. Where defoliation is extensive, trees may remain bare as late as early August. However, refoliation of hardwood trees that have had 60 percent or more of their foliage consumed usually begins around mid-July, or when the caterpillars pupate. Studies indicate that hardwoods suffering less than 60 percent loss of foliage do not refoliate. The process of refoliation requires the use of stored energy. Repeated attacks deplete the food resources in the tree. As tree vigor declines, death may result due to an attack by organisms or other environmental extremes that ordinarily would not cause tree mortality.

Gypsy moth infestations were first discovered in Pennsylvania in 1932. Major outbreaks did not begin, however, until the mid 1940's. Suppression of the insect activity using aerial applications of DDT was fairly successful at that time. However, in 1963 DDT spraying was abandoned in favor of more environmentally acceptable but less effective insecticides. Since then, there has been a steady increase in the insect's population and range. Figure 2 illustrates this continued rise in gypsy moth populations as reflected in the increasing defoliation during peak years. Presently, insect damage is on an upward swing. During the 1981 summer feeding cycle, federal officials estimated that approximately one million hectares of hardwood forest were defoliated in Pennsylvania (Forest Pest Management Staff, 1982).

The rise in defoliation was also evident in the increase in timber mortality. Between 1970 and 1979, over one million hectares of prime timber land was surveyed

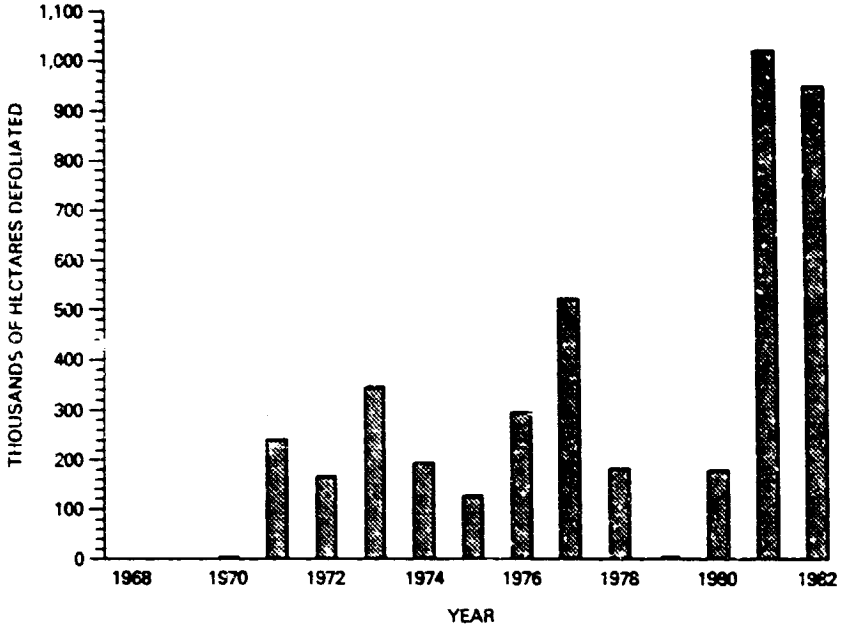


Figure 2. Trend of gypsy moth defoliation in Pennsylvania; thousands of hectares defoliated each year.

in Pennsylvania to estimate the amount of timber lost to gypsy moth damage. The net worth of that timber was estimated to be in excess of 36 million dollars.

Identification of Defoliation . . Test of Satellite Remote Sensing Capabilities

Over the years, state and federal agencies have spent millions of dollars developing pest management programs in an attempt to reduce timber losses resulting from insect damage. These techniques include ground surveys, aerial-based surveys, airphoto interpretation, and satellite-based surveys.

The temporal and synoptic coverage provided by Landsat makes the satellite sensor an ideal survey medium for monitoring widespread phenomena such as insect related damage in forested areas. Hence, considerable research has been directed toward examining the use of Landsat multispectral scanner (MSS) data to monitor gypsy moth defoliation of hardwood forests. Rohde and Moore (1975) reported that

gypsy moth defoliation could be identified on Landsat MSS color composite images using standard photointerpretation techniques. However, the ability to quantify degrees of defoliation was hindered by uncalibrated brightness and tonal changes. Rohde and Moore suggested that digital processing of remotely sensed data might improve mapping accuracy.

Other Landsat-based studies on defoliation assessment included an investigation by Talerico et al. (1978) which described a quantitative photographic approach for delineating various levels of insect defoliation by applying advanced photometric calibration techniques to aerial photography and Landsat imagery. They concluded that Landsat data were not only more economical, but also better than high altitude photography for mapping defoliation.

Remote sensing specialists at NASA's Goddard Space Flight Center (GSFC) have been developing, evaluating and modifying image manipulation and processing techniques since 1975 that facilitate the use of satellite data to assess forest damage from major insect infestations. These research activities resulted in a series of studies conducted in Pennsylvania which demonstrated the utility of Landsat MSS digital data and image processing for gypsy moth defoliation assessment (Williams, 1975; Williams and Stauffer, 1978; Williams et al., 1979; Williams and Ingram, 1981). Each study identified one step in the defoliation assessment process that would improve the identification of forest disturbance classes. Williams (1975) used a supervised classification approach to map areas of heavy and moderate defoliation and healthy forest in eastern Pennsylvania. Classification results were subjectively analyzed and found to be representative of actual ground conditions. Later, Williams and Stauffer (1978) isolated changes in the forest canopy that were related to gypsy moth defoliation by creating a multitemporal Landsat data set containing images acquired before and after infestation. This latter study made use of automated change detection techniques

that essentially eliminated errors of commission with non forest land cover. The authors further improved classification results by applying selected data transformation techniques to the multitemporal Landsat data set (Williams et al., 1979). The selected transformations had originally been developed for estimating agricultural and rangeland standing green biomass (Tucker, 1979). However, Williams et al. (1979) concluded that these same transformations would discriminate heavy defoliation from healthy forest. Areas of moderate defoliation were confused with healthy forest on northwest aspects, but were distinct from healthy forest conditions on southeast facing slopes. This latter study indicated that diverse terrain and topographic conditions typically associated with forest lands cause variations in remotely sensed data, leading to problems in accurately classifying forest cover conditions. In light of this, Williams and Ingram (1981) designed another study which assessed the utility of incorporating high spatial resolution digital terrain data with Landsat MSS data to reduce confusion between spectrally similar forest canopy conditions such as healthy vegetation and moderate defoliation. Their results indicated that these two forest canopy conditions could not be consistently separated from one another even when accounting for any confounding effects on sensor response due to slope orientation. However, their study also confirmed that heavy defoliation is separable from other forest canopy conditions.

RESEARCH AND DEVELOPMENT ACTIVITIES

The NASA-BOF/DFPM Joint Research Project was designed to develop an automated system for conducting annual gypsy moth defoliation surveys in Pennsylvania using Landsat multispectral scanner data and digital processing techniques. The creation of this system involved a number of studies which resulted in the following developments:

1. An effective procedure for defoliation assessment using Landsat digital data;
2. Identification of a temporal window for defoliation assessment;
3. A statewide data base;
4. A data management system to interface image analysis software with the statewide data base; and
5. A cost/benefit analysis of this operational system.

Each of these developments are briefly described in the remaining text. More in-depth discussions of many of the key elements can be found in the Appendices.

Analysis Procedure

Research completed at GSFC indicated that digital analysis of Landsat MSS data for defoliation assessment required a two-step preprocessing procedure that uses multitemporal data sets that represent forest canopy conditions before and after defoliation (see Figure 3). The purpose of this procedure is to create a digital image in which all nonforest cover types have been eliminated or masked-out of a Landsat image that exhibits insect defoliation. By masking out nonforest cover types, confusion between defoliated forest and nonforest is eliminated,* thus preventing errors of commission.

*NOTE: Errors of commission are "eliminated" to the extent of the accuracy with which forest and nonforest cover types can be separated.

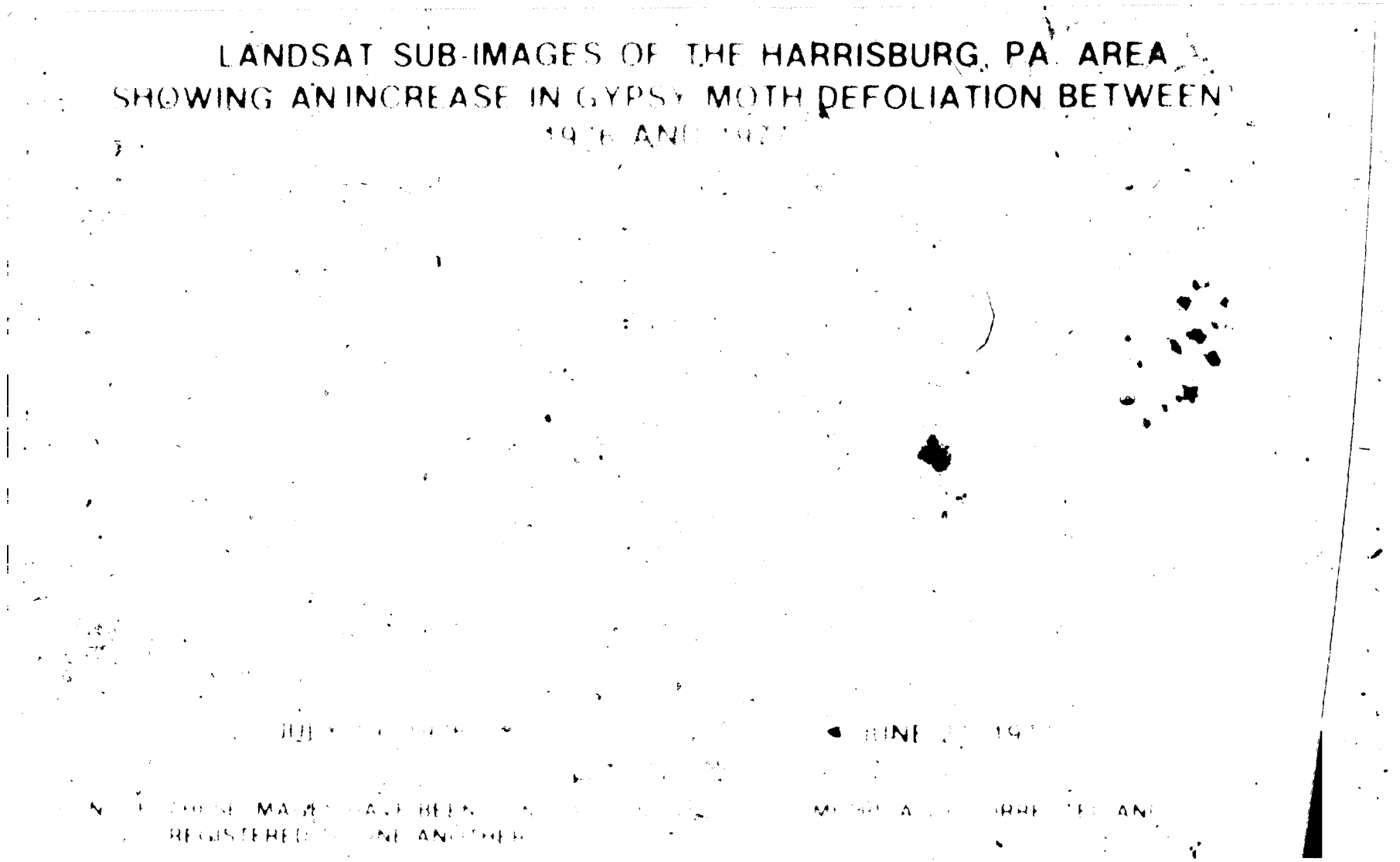


Figure 3. Comparison of Landsat imagery acquired before and after major gypsy moth attack. Harrisburg, Pennsylvania.

The first step of this preprocessing procedure begins by obtaining a Landsat image of a given area during the growing season, but prior to infestation. This image is classified using computer-aided analysis techniques to identify the extent of forest cover versus nonforest cover. In the second step, another Landsat image over the same area that was obtained after insect damage had occurred is digitally registered to and overlaid onto the forest/nonforest classification map derived from step 1. The defoliated Landsat data may be multiplied by the forest/nonforest classification, where 1=forest and 0=nonforest, to produce a masked, defoliated data set. Thus, all nonforest areas in the defoliated image will have a zero value and are ignored (see Fig. 4). Subsequent analyses are applied to the masked, defoliated image for disturbance assessment.

Several analysis procedures are available that could be used to generate the forest/nonforest classification map. Project personnel examined a number of these procedures and identified a two-channel supervised Bayesian classification technique as the simplest, most accurate approach. The selection of the procedure was based not only on accuracy and simplicity, but on ease of updating the forest classification as well. Appendix I describes the study conducted to select this procedure.

Upon completion of the preprocessing, the actual defoliation assessment can be carried out. As was the case with the forest classification map, a number of image manipulation procedures that could be used to conduct the assessment were examined by Goddard analysts. Throughout this research effort, the Pennsylvania Bureau of Forestry/Division of Forest Pest Management provided technical assistance and support information on the location and severity of gypsy moth damage in the state. The data supplied by BOF/DFPM was used to determine the performance of each of the processing procedures examined. A procedure known as the Ratio Vegetation Index was identified as the most appropriate for defoliation

Figure 4. Generation of a masked, defoliated Landsat MSS image by applying a forest/nonforest mask to Landsat data exhibiting insect damage: Doubling Gap, Pennsylvania.

assessment (Nelson, 1981a). A complete description of this activity and research results are given in Appendix II.

The Ratio Vegetation Index (RVI) technique is used to delineate two levels of defoliation (heavy, 60-100% canopy removed; and moderate, 30-60% canopy removed), as well as healthy forest. This index is applied to the masked, defoliated image. The Ratio Vegetation Index is calculated by computing the ratio of the infrared to red response (MSS Band7/MSS Band 5) for each non-zero (i.e., forested) pixel in the masked, defoliated image. Previous work, notably in agricultural applications (Tucker, 1979), had shown that the infrared response increases, the red response decreases, and the infrared to red ratio increases as the amount of green leaf canopy in the sensor's field of view increases. Hence, low ratios in forested areas would indicate a thin (i.e., defoliated) canopy. By comparing ground reference information to the ratio values observed, breakpoints between the various levels of defoliation can be calculated. Once these breakpoints are known, the image may be classified into heavy defoliation, moderate defoliation, and healthy forest. It should be noted, however, that significant confusion exists between healthy and moderately defoliated forest. Figure 5 is a schematic diagram of the defoliation assessment procedure.

The RVI defoliation assessment procedure was used by Pennsylvania BOF/DFPM personnel and Goddard support personnel to complete a 1981 defoliation assessment for one complete Landsat scene (Path 16, Row 32). Pennsylvania BOF/DFPM selected an intensive study site within this scene to compare the estimates of defoliation obtained over that area from several different survey methods: aerial sketchmapping, airphoto interpretation, and Landsat image processing. Table 1 compares the Landsat and aerial sketchmapping defoliation assessments to the airphoto interpreted results. These results are based on the assumption that the

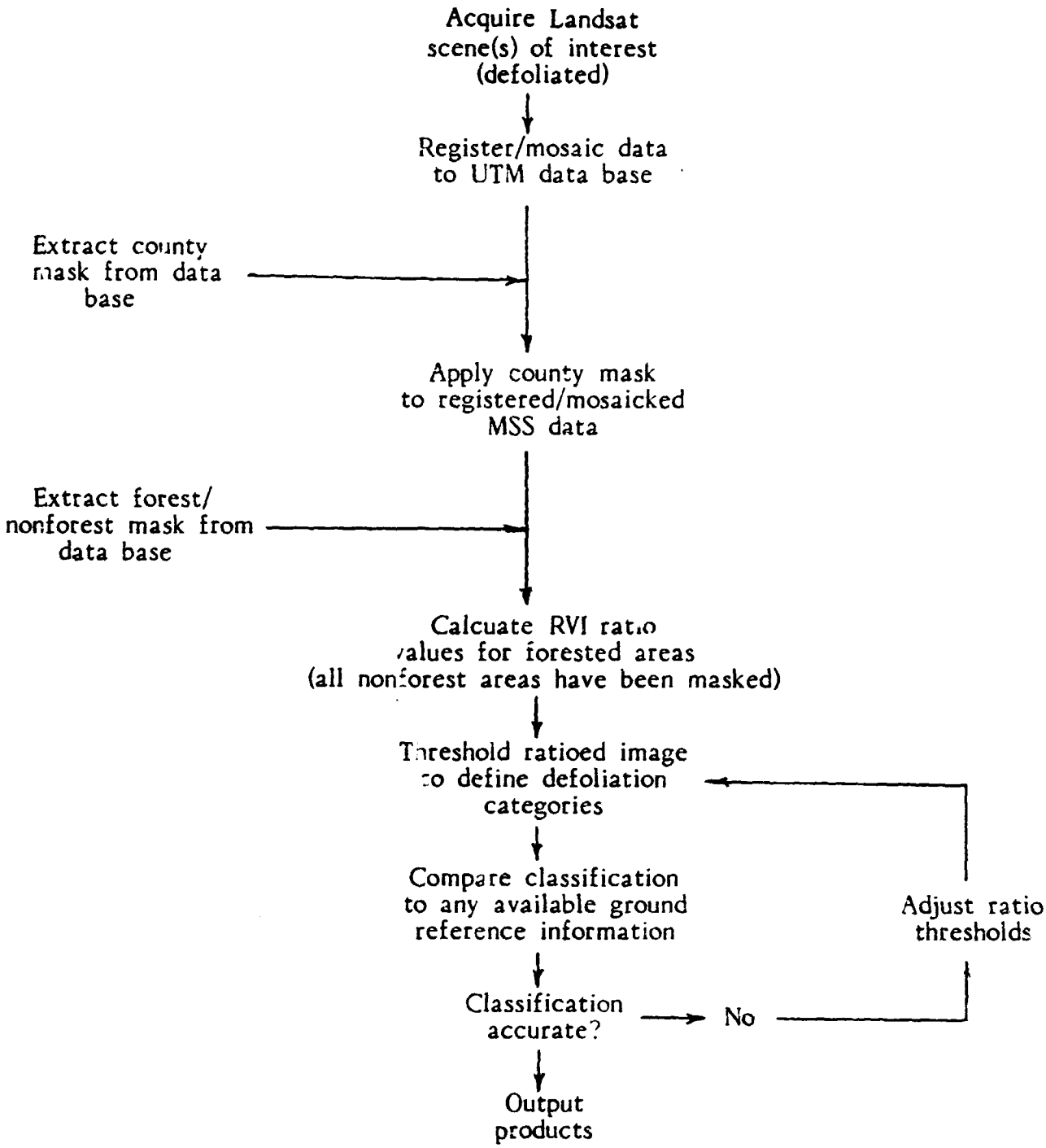


Figure 5. Schematic diagram of the defoliation assessment procedure

airphoto interpreted data most closely reflected the true ground conditions. The airphotos were acquired within hours of the Landsat overpass. The aerial sketchmapping mission was flown within three days of the satellite overpass.

Table 1. A comparison of Landsat and aerial sketchmapping defoliation assessments to airphoto interpreted information for Doubling Gap, Pennsylvania, July, 1981. Two defoliation classes are delineated: heavy defoliation (60-100% canopy removed) and a healthy-moderate defoliation cover type (0-60% canopy removed).

		Landsat		Aerial Sketchmapping	
		Hvy	Hth-Mod	Hvy	Hth-Mod
Airphoto Interpretation	Heavy	77.9	22.1	91.4	8.6
	Hth-Mod	22.5	77.5	43.5	56.6
		Avg:	77.5%	Avg:	74.0%
		Over:	77.7%	Over:	70.1%

A more comprehensive treatment of this investigation may be found in Appendix III.

Selecting the Appropriate Time for Defoliation Assessment

A major concern in developing the Landsat-based defoliation assessment procedure was the acquisition of useful satellite information. Other Landsat studies in the eastern United States encountered problems obtaining cloud-free imagery during the summer months because of climatic conditions. If the defoliation assessment was to depend only on acquiring data at one point during the summer (i.e., peak defoliation), the operational defoliation assessment system would be seriously lacking in flexibility and fundamental utility. Therefore, a study was devised during the JRP to define the temporal limits within which Landsat data might be obtained and still provide useful defoliation information (Nelson, 1981b, see Appendix IV).

The temporal analysis indicated that the effects of gypsy moth defoliation can be assessed over a two month period beginning in early June. However, the optimum

time to delineate defoliation is a two or three week period from late June to early July. Within this temporal window heavily defoliated forest can be successfully separated from moderately defoliated and healthy forest. However, the effects of insect damage can be assessed at time other than peak defoliation, increasing the probability that useful satellite data can be acquired over the defoliation site. It should be noted that the length of the temporal window is fairly consistent from one year to the next, but the beginning or end of the window may shift by one or two weeks depending upon weather and biological conditions.

Pennsylvania Statewide Data Base

The purpose of this JRP was not only to identify and test the most appropriate procedure for satellite-based defoliation assessment, but also to design an operational defoliation assessment system for the entire state of Pennsylvania. Analysis of Landsat data for assessing insect defoliation over an area as extensive as Pennsylvania requires the processing and storage of large volumes of data. Therefore, a system which could accommodate efficient digital processing as well as storage and retrieval of these data needed to be devised.

Early in the JRP project Pennsylvania Bureau of Forestry and NASA/GSFC began to examine alternative methods of handling the large volume of remotely sensed data needed to complete statewide defoliation assessments on a yearly basis. The decision was made to develop a Landsat-derived geographic data base which could be interfaced with analysis software. The data base needed to include the following components:

1. A Landsat digital mosaic of Pennsylvania exhibiting no defoliation and registered to the Universal Transverse Mercator (UTM) projection.
2. A forest resources map (forest/nonforest mask) generated from the Landsat mosaic and registered to the Landsat digital data base.

3. A data layer containing Forest Pest Management District and county boundaries registered to the UTM projection.

During the first year of the Joint Research Project, staff members of the Jet Propulsion Laboratory (JPL) in Pasadena, CA demonstrated the technical feasibility of creating the statewide Landsat digital mosaic. Following this demonstration, JPL generated the Landsat mosaic of Pennsylvania.

The mosaic was created by compiling ten essentially cloud-free, non-defoliated summertime Landsat images over Pennsylvania (see Figure 6). These images were first registered to the Universal Transverse Mercator Projection for the state, which is divided into two UTM zones along the 78th parallel (UTM Zone 17 for western Pennsylvania and UTM Zone 18 for eastern Pennsylvania). The grid (pixel) size of 57 meters was chosen for both zones. After registration, the images were digitally combined (side to side and end to end) to form a Landsat mosaic of Pennsylvania. This mosaic constituted the foundation of the Landsat-derived geographic data base which would be used in subsequent statewide annual assessments.

An evaluation of the registration of the Pennsylvania mosaic was undertaken by GSFC personnel to determine at what level of detail the mosaic accurately reflected map standards (Stauffer and Russo, 1982). The evaluation indicated that the mosaic data could be used in conjunction with small scale (1:250,000) maps. However, misregistrations on the order of approximately three pixels were evident using larger scale (1:24,000) maps. Table 2 presents the average misregistration error (in meters) for each of the eight quadrangles covering the state. In addition, the largest offset found within each quadrangle is listed.

ORIGINAL PAGE
COLOR PHOTOGRAPH

Figure 6. Landsat mosaic of Pennsylvania. These data were obtained during the growing season in 1976-1979, no gypsy moth defoliation was evident.

Table 2. Mosaicked Landsat data to UTM grid misregistration error (in meters) for the entire state (1 pixel = 57 meters).

UTM Zone	Quad	Average Misregistration Error Line	Average Misregistration Error Column	Worst Case Line	Worst Case Column
17	1	37	38	97	74
	2	26	51	68	131
	3	52	93	120	154
	4	49	180	143	604
18	5	62	64	182	268
	6	70	186	342	559
	7	37	36	154	91
	8	55	80	205	410

A more detailed description of the mosaic procedure and registration assessment is given in Appendix V.

The same Landsat images used to generate the Pennsylvania mosaic were also used to generate a forest/nonforest classification map of the state that was input and registered to the data base. The forest classifications were generated by GSFC support personnel with initial assistance from Pennsylvania BOF/DFPM personnel. The procedure outlined in Appendix I was used to classify the Landsat data. A comprehensive evaluation of the statewide forest classification accuracy was completed in a joint effort by BOF/DFPM and Goddard support personnel. The accuracy of the forest/nonforest classification was assessed at random points throughout the state. The Landsat classification identity and the photointerpreted identity of each point were compared. On a point-by-point basis, the overall statewide accuracy was 82%. If 3x3 pixel neighborhoods were considered, the overall accuracy was 90%. A complete description of the accuracy evaluation procedures and results are given in Appendix VI.

Other data layers which were input and registered to the Landsat-derived geographic data base consisted of digitized Pennsylvania county and Forest Pest Management

District boundaries, and USGS 7-1/2 minute topographic map boundaries (see Figure 7). The availability of these boundary overlays enable the data base user to access a subsection of the mosaic without the necessity of retrieving the entire data base. Access to the data base is accomplished by means of a data management front-end system that interfaces the Landsat-derived data base with image analysis software.

A Data Management Front-End System

The Pennsylvania State University, Office for Remote Sensing of Earth Resources (ORSER) developed a data mangement front-end system to interface the Landsat-derived data base with image analysis software. This front-end system provides bookkeeping activities, sets up the image analysis programs for defoliation assessment, and references the data base according to the user's request (Turner, 1981). For example, if an analyst wishes to estimate the extent and severity of insect defoliation for any management district or county within the state of Pennsylvania using the previously described analysis procedure, Landsat data acquired during the gypsy moth defoliation cycle can be registered to the data base. The district or county boundary can then be extracted from the data base to isolate the area of interest. The forest/nonforest classification map can then be extracted and overlaid onto the Landsat data to mask out nonforest cover types and the Ratio Vegetation Index can be applied to this new "masked, defoliated image" to delineate areas of insect disturbance.

All of the image processing jobs previously described are requested via a user friendly, front-end system. This system was developed to allow one not familiar with the different data analysis techniques to interact with complex programs in a conversational manner. A complete description of the capabilities and functions of the data management front-end system is given in Appendix VII.

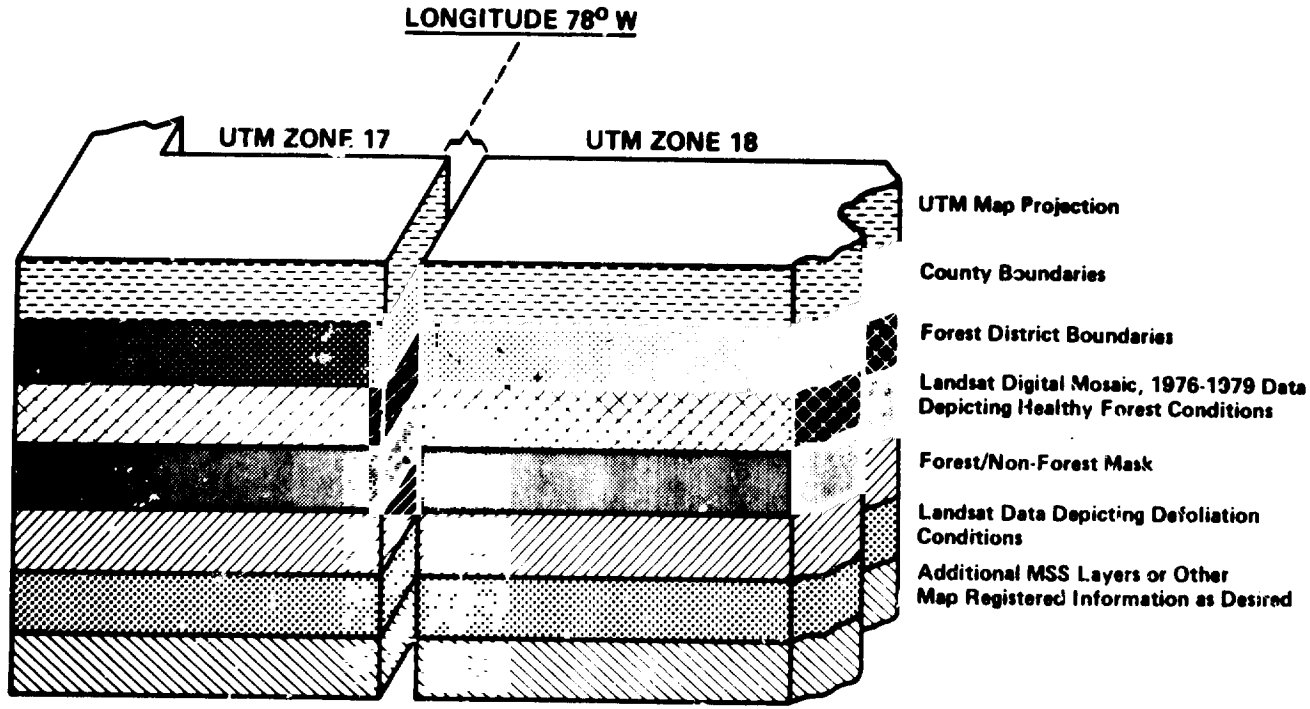


Figure 7. Characteristics of the Pennsylvania statewide data base.

Data Reduction Techniques

The volume of data required to be processed for a statewide defoliation assessment can reach tremendous proportions. Therefore, a study was initiated to evaluate procedures for reducing the amount of data to be processed for the statewide forest mask and subsequent defoliation assessment (Russo and Stauffer, 1982, see Appendix VIII). The study focused on alternative subsampling schemes for data reduction. These schemes included a full resolution data set, a 2 x 2 averaging of pixels, and the selection of every other pixel within every other line. Landsat data acquired over the selected study area was used to generate forest resources maps using a variety of computer-aided analysis techniques.

A comparison among the forest classification performance levels indicated that reducing the Landsat data by averaging or subsampling tended to reduce classification performance. However, the reduction in classification performance which is evident from the 2 x 2 averaging method is relatively insignificant compared to the full resolution scene. Thus, this approach may be a reasonable alternative for reducing the large volume of data required for Landsat-based resource mapping and defoliation assessments, should the need arise.

The Pennsylvania data base as currently implemented on the PSU computer does not utilize averaged or resampled Landsat MSS data. This option is available, however, if future use requires such a constraint.

APPLICATIONS ACTIVITIES

1981 Defoliation Assessment of Centre and Perry Counties

Once the Pennsylvania data base and the user friendly front-end system were on line on the Penn State University computer, tests were run to insure that the systems worked harmoniously. Two counties, Centre and Perry, were assessed to determine severity of defoliation in 1981.

For these tests, "default" defoliation assessments were done. The front-end system allows the analyst to select one of two analysis paths, based on the background and experience of the analyst. If the analyst has a remote sensing background and is familiar with the VICAR and ORSER image processing languages, the analyst may use any number of standard remote sensing technqiues to classify defoliation severity. If however, the analyst has little or no remote sensing background, he may select a "default" pathway where pre-selected job cards are submitted for the digital assessment. In taking this default pathway, the analyst is essentially asking for a "standard" assessment as set forth in this final report (i.e., select the area of interest, apply the forest/nonforest mask, calculate the MSS Band 7/Band 5 ratio, classify the ratioed image, generate the output products). The Band 7/Band 5 ratio breakpoints used in this "standard" assessment are given in Appendix III. See Table III-2 for breakpoints and associated accuracies when the product is compared to airphoto interpretation results. These breakpoints, then, were used to classify the forested areas of Centre and Perry Counties into heavy defoliation, moderate defoliation, and healthy forest.

Cost-Benefit Analysis

The BOF/DFPM kept track of costs associated with obtaining the same type of information via aerial sketchmapping for the two counties. By noting computer costs, a rudimentory cost analysis could be done comparing aerial sketchmapping and

satellite mapping. The cost figures are summarized in Table 3. The costs asseciated with the aerial sketchmapping pertain to money spent to produce the final county-wide defoliation maps. The costs associated with the satellite digital data analysis pertain to the production of tabular statistics, electrostatic printer output (B&W greyscale map), and magnetic tape files which could be used to make a color print of the county-wide defoliation classification. It should be noted that since Centre County straddles the 78th meridian (which effectively divides the Pennsylvania data base into an eastern half and a western half), two separate classifications had to be done. The cost of both classifications (Centre County east and Centre County west) are reflected in the figures below.

Table 3. Cost comparsion (in dollars) for Centre and Perry Counties, aerial sketch-mapping versus defoliation classification using satellite data.

Digital Analysis		Aerial Sketchmapping	
CPU Costs	73.74	Aircraft (18.6 hours)	987.00
Data and Mosaicking (esimated)[1]	1300.00	Miscellaneous (Maps and Travel)	281.00
Wages (3 man hours)	60.00	Wages (83.2 hours)	1397.15
Total	1433.74	Total	2665.15
Cost per Hectare	0.0032		0.0061

[1]This figure is a rough estimate derived as follows:

Estimated cost to mosiac one-half of a Landsat scene	$975.00
Estimated cost of one-half of a Landsat scene	$325.00

Admittedly the cost analysis above can only be used for a rough comparison. The data base was implemented in a research and development mode, hence costs applicable to the operational use of the data base are often difficult to identify. In addition, an assumption implicit above is that the operating equipment (software and hardware) are already in place and functional.

In order to describe a complete cost picture, Table 4 outlines the costs associated with acquiring the hardware, software, and personnel necessary to

implement and maintain the data base. Table 4 is, in essence, a shopping list for those readers who may be considering such a data base, yet do not currently have the facilities.

Table 4. Estimated costs (in dollars) of hardware, software, and personnel necessary to implement and maintain a statewide, satellite-based digital data base.

Hardware	
Purchase a mini computer with peripherals (tape drives, discs, terminals, digitizer)	$500,000
An alterantive:	
Purchase time on exisitng facility, annual budget ($0.20/CPU second, 35 hours)	25,200
Software	
VICAR (leased for 10 years from COSMIC (University of Georgia, Athens)	2,400
ORSER (buy from Pennsylvania State University)	3,000
Data	
1 year (for Pennsylvania) 10 scenes @ $650.00/scene	6,500
Mosaicking Cost	
10 scenes/layer, 2 layers (i.e., healthy and defoliated)	40,000
Analyst	
annual budget	40,000

Table 4 supports the fact that high initial fixed costs are often prohibitive for those who may wish to pursue the digital analysis - data base concept but do not have the facilities. The costs suggest that a multi-user or multi-agency attitude must be cultivated in order to distribute the costs to a wider number of users. The advantage of such a multi-user concept is that the use of a common data base insures interagency format compatibility and facilitates interagency information and data exchange.

Expected Utility of the Defoliation Assessment System

The utility of the Pennsylvania Landsat data base is governed by:

1. the availability of MSS data;
2. the ease with which products may be generated;
3. the accuracy of those products; and
4. the cost of generating the products.

The latter three points are addressed in Appendices VII, III, and in the body of the main report, respectively. These sections explain that satellite defoliation products are generated in a user-friendly environment for which a remote sensing background is not necessary. The satellite products are at least as accurate as aerial sketchmaps, and cost estimates indicate that satellite processing is less expensive.

Were it not for point 1, the facts would suggest that Landsat data analysis should supplant aerial sketchmapping for statewide defoliation assessments. However, the ability to acquire useful MSS data is in question. MSS data must be acquired within a two-month window (see Appendix IV) in order to be useful for satellite defoliation assessments. A given piece of real estate is imaged once every sixteen days by the Landsat 4 MSS, hence one has three, at best four, opportunities to collect useful (i.e., relatively cloud-free) data.

Some estimate of the probability of obtaining useful satellite data may be calculated by looking at historical records. The EROS (Earth Resources Observation System) Data Center in Sioux Falls, South Dakota has archived all Landsat MSS data acquired over the U.S. since the first satellite flew in 1972. Figure 8 presents the results of EROS Data Center archive searches to locate useful MSS data. Useful in this context means a scene acquired between June 1 and August 1 exhibiting cloud coverage less than or equal to 30% and having a data quality rating of at least 2 (on a scale of 8) in bands 5 and 7 (the red and second near infrared bands, respectively). The state is covered by 10 scenes: 5 satellite passes from east to west, 2 scenes per pass north to south. If at least one scene was found which fulfilled the temporal

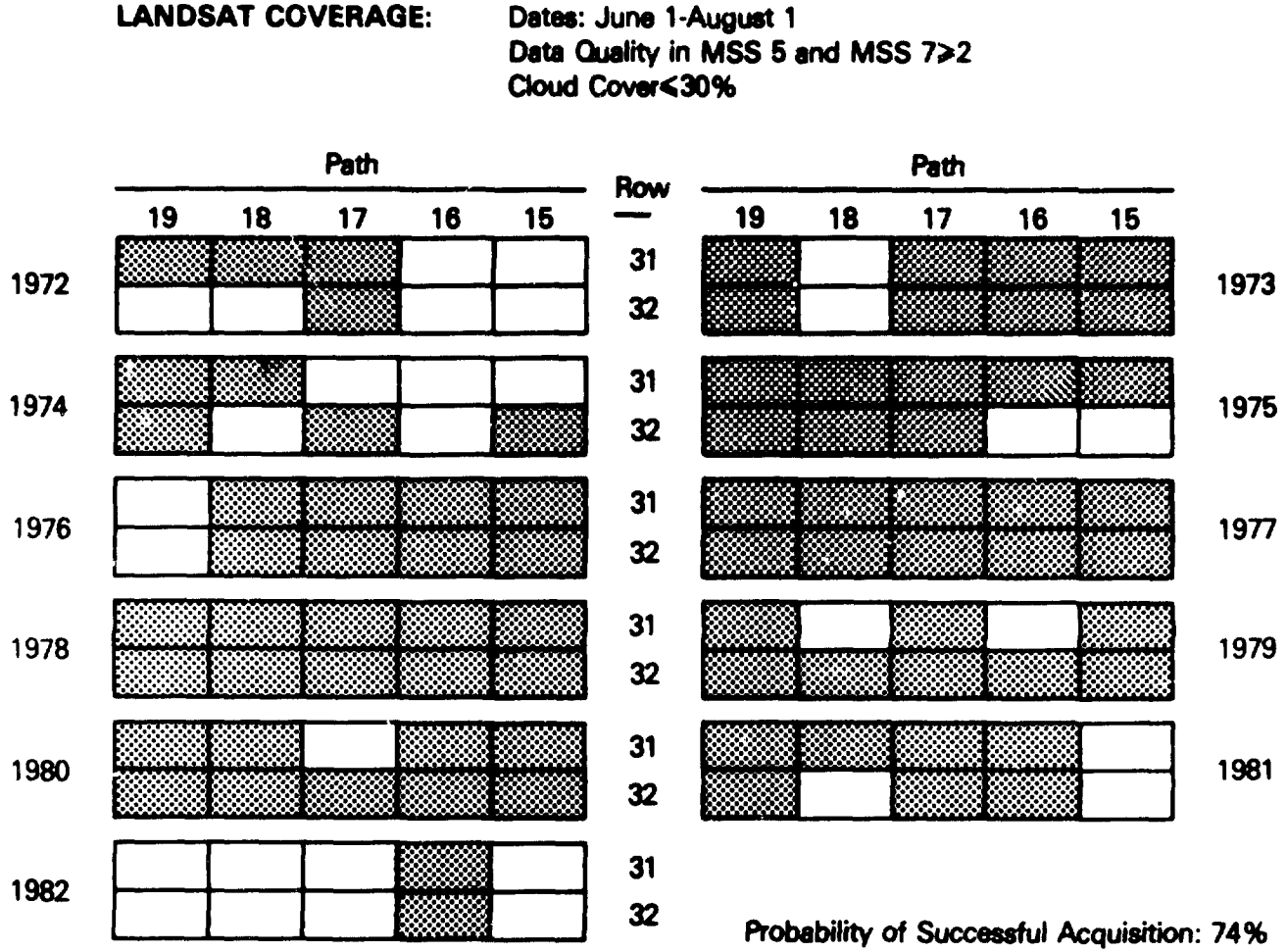


Figure 8. Landsat scenes successfully acquired over Pennsylvania between June 1 and August 1 with less than or equal to 30% cloud cover for the years 1972-1982. Ten scenes cover the state (5 passes, two scenes per pass). Shaded boxes indicate that at least one scene was acquired within the constraints for that given year.

and data quality criteria listed, the path/row was shaded for that year in Figure 8.

Based on this historical search, MSS data were successfully acquired within the defoliation window 74% of the time. It should be noted that two satellites were operating from 1976 to 1980, doubling the probability that useful data could be acquired. Looking at 1972-1975 and 1981 and 1982, when only one satellite was operating, the probability of obtaining data between June 1 and August 1 was 57%. The inability to reliably collect Landsat MSS data dictates that alternate defoliation assessment methods must be available. Therefore, Landsat defoliation assessments will be used to supplement aerial sketchmapping results. Complete conversion to an all-Landsat system cannot be recommended.

SUMMARY

The Department of Environmental Resources does not have the hardware, software, or analysts necessary to perform in-house analysis of LANDSAT digital data, nor is it expected to acquire such a capability in the foreseeable future. Because all of the necessary facilities do exist at the Office of Remote Sensing of Earth Resources (ORSER) at the Pennsylvania State University, it is logical for DFPM to contract with ORSER to do the mosaicking, registrations, and defoliation assessment. DFPM is presently working with ORSER on terms for such an arrangement.

The Division of FPM is most anxious to adopt this technology and integrate it into the present system. As such, the results of the Landsat analyses to detect defoliated areas will be distributed to county and forest district management personnel along with or in place of the aerial sketchmapping figures. Of course, until such time as cloud-free imagery from LANDSAT can be assured, it cannot wholly replace other methods now used to acquire this information.

While it was not specifically addressed in the project, there are other potential uses for LANDSAT data in the Department of Environmental Resources (DER) and the Pennsylvania Department of Agriculture. These include forest type mapping, surface mine monitoring, certain water detection monitoring, and crop monitoring. Because the cost of this technology is relatively high for just one use, it is important to identify other legitimate applications in order to defray the costs of data acquisition and manipulation.

The results of the joint NASA-Bureau of Forestry, Division of Forest Pest Management JRP project have been most encouraging. Simplified digital analysis procedures to produce a statewide Landsat-derived forest resource map and defoliation assessment will enable entomologists to prepare timely surveillance reports and plan for appropriate pest management procedures. The Landsat-derived

geographic data base will facilitate these assessments by allowing quick retrieval of statistics, selected satellite imagery, and defoliation maps. Interactive digital analysis capabilities will facilitate not only the defoliation assessment but also future updating of the forest classification map. Additional information layers can be input to the data base at later dates to enhance its utility to other users. All of these capabilities are possible through the data management front-end system.

BIBLIOGRAPHY

Forest Pest Management Staff. 1982. Gypsy Moth Defoliation in the Northeast, 1981. USDA Forest Service, Broomall, Pennsylvania. 13 pgs.

Marshall, E. 1981. The Summer of the Gypsy Moth. Science, 213 (4511):991-993.

McManus, M.L. 1978. Expanded Gypsy Moth Research and Development Program. Journal of Forestry, Vol. 76, No. 3 pp. 144-149.

Nelson, R. 1981a. ASSESS2 Analysis of Four Methods for Classifying Forest Defoliation (Revised). Goddard Space Flight Center/Earth Resources Branch, Internal Report.

Nelson, R. 1981b. Defining the Temporal Window for Monitoring Forest Canopy Defoliation using Landsat. Proceedings ASP 47th Annual Meeting, Washington, DC, February 22-27, 1981. pp. 367-382.

Rohde, W.G. and H.J. Moore, 1975. Gypsy Moth Defoliation Assessment: Forest Defoliation is Detectable from Satellite Imagery. NASA CR-146626, MF#N76-20601. NASA, Houston. TX.

Russo, S. Adrouny, and M.L. Stauffer. 1982. The Impact of Data Reduction on Forest Classification Accuracy. Computer Sciences Corporation. Contract Report CSC/TM-82/6205, Goddard Space Flight Center, Greenbelt, MD.

Stauffer, M.L. and S. Adrouny Russo. 1982. Characterization of the Registration Accuracy of the Pennsylvania Digital Mosaic. Computer Sciences Corporation. Contract Report CSC/TM-82/6225, Goddard Space Flight Center, Greenbelt, MD.

Talerico, R.L., J.E. Walker, and T.A. Skratt. 1978. Quantifying Gypsy Moth Defoliation. Photogrammetric Engineering and Remote Sensing, Vol. 44, No. 11. pp. 1385-1392.

Tucker, C.J. 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. Remote Sensing of the Environment. Vol. 8, No. 2. pp. 127-150.

Turner, Brian J. 1981. Development of a Data Management Front End for Use with a Landsat Based Information System. Interim Report, Contract No. NAS5-26468, Pennsylvania State University, University Park, PA.

Williams, D.L. 1975. Computer Analysis and Mapping of Gypsy Moth Defoliation Levels in Pennsylvania using Landsat-1 Digital Data. Proceedings of the NASA Earth Resources Survey Symposium, Technical Sessions Presentations. Vol. 1-A, pp. 167-181.

Williams, D.L. and K.J. Ingram. 1981. Integration of Digital Elevation Model Data and Landsat MSS Data to Quantify the Effects of Slope Orientation on the Classification of Forest Canopy Conditions. Proceedings, International Symposium on Machine Processing of Remotely Sensed Data. LARS/Purdue University. pp. 352-362.

Williams, D.L., M.L. Stauffer, and K.C. Leung. 1979. A Forester's Look at the Application of Image Manipulation Techniques to Multitemporal Landsat Data. Proceedings of Fifth Annual Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, IN. pp. 368-375.

Williams, D.L. and M.L. Stauffer. 1978. Monitoring Gypsy Moth Defoliation by Applying Change Detection Techniques to Landsat Imagery. Proceedings of Symposium on Remote Sensing for Vegetation Damage Assessment. American Society of Photogrammetry. pp. 221-229.

APPENDIX I

Selection of Forest/Nonforest Classification Procedure

STUDY OBJECTIVE

The accurate assessment of gypsy moth defoliation is dependent upon the generation of a forest classification that is used in the assessment process to separate forested areas from nonforested areas in the Landsat digital data. Therefore, a loosely defined study was undertaken to identify a simple, cost-effective, and accurate analysis procedure to derive the forest/nonforest mask from Landsat multispectral scanner data. Several analysis procedures were examined including a four channel parallelepiped algorithm using training statistics from four major land categories and several different program parameters and numerous bayesian procedures using a variety of spectral channels, training statistics, land cover categories and program parameters. Two of these procedures, a four channel parallelepiped and a two-channel bayesian, are compared in this Appendix.

DATA AND STUDY SITE DESCRIPTION

The study site selected for this activity is located near Harrisburg, Pennsylvania. Cloud-free Landsat data collected over the study site on July 19, 1976 were obtained to produce the various forest classification maps. A 1741 line by 1286 column subsection of the Landsat scene was extracted from the image for analysis. U.S.G.S. Topographic maps (7.5 minute series, 1:24,000) that corresponded to the study site were used as ground reference data. These were supplemented by available air photos over portions of the study site.

PROCEDURE

Two forest classification maps were generated from the Landsat data using the following analysis procedures.

1. Parellelepiped classification algorithm: Training fields for the four major land cover categories in the study area--forest, soil, agricultural crops and urban

land--were first identified on air photos collected by the Pennsylvania Division of Forest Pest Management. These training fields were then located on the Landsat image to obtain their line and column coordinates. The coordinates were input to a computer program that generates training statistics which describes the spectral response pattern of each class. These statistics were then used in the parallelepiped classification algorithm to produce the forest resources map. The parallelepiped procedure uses the training field statistics to identify the range of spectral values associated with each cover type. Unknown pixels are classified into known cover type classes by comparing the pixel spectral response value to the calculated ranges.

2. Bayesian classification algorithm: The second forest resources map was produced using a Bayesian maximum likelihood classifier. Only forested training fields were identified in the Landsat image. Statistics were derived from these training fields and used to generate a single class (forest), two-channel (MSS_5 and MSS_7) Bayesian classification. The Bayesian classifier produced a probability map which assigns each pixel a value (from 0 to 255) that is proportional to the probability of the pixel belonging to the forest class. Low values connoted nonforest areas, high values connoted forested areas. Pixels of known identity (forest or nonforest) were located in the Landsat data. This probability map was then "density sliced" to produce a forest classification with two classes--forest and nonforest, such that the classification accuracy for the pixels of known identity was maximized.

The Landsat-derived forest resources maps were qualitatively compared to identify the appropriate procedure for generating the Pennsylvania statewide forest classification mask. The selected procedure was given a more rigorous evaluation to determine the actual classification performance.

RESULTS

The parellelepiped and Bayesian forest resources maps were qualitatively compared

to ground reference data to determine classification performance. The performance for both classifiers appeared to be comparable although fewer errors of commission were evident in the Bayesian classification. These results indicated that either procedure would be acceptable for producing the forest classification. Therefore, the selection of the optimum method for analysis was made on the basis of analyst and computer time efficiency. The single class, two-channel Bayesian procedure required only one set of training fields to be located for the forest class. The procedure used only two spectral channels. Hence, analyst time could be minimized and computational costs could be kept down. In addition, the Bayesian procedure allowed greater flexibility than the parallelepiped in that the Bayesian classification could be easily modified by repeating the density slice as more ground reference data became available. These results indicated that a two-channel Bayesian procedure would be most acceptable for generating the statewide forest/nonforest mask.

After the Bayesian forest resources map was selected, the forest/non forest mask was generated from the classified map by rescaling the digital data so that all forest pixels were given a value of 1 and all nonforest pixels were given a value of zero. A 600 line by 500 column section of this product was evaluated to determine how well the mask characterized the actual ground situation.

The 600 x 500 pixel area was registered to corresponding 1:24,000 scale U.S.G.S. topographic maps. One hundred fifteen pixels were randomly selected from each of the two cover categories--forest and nonforest--on the Landsat-derived classification. The location of these pixels was then noted on shade prints obtained from the raw Landsat data and overlaid onto the appropriate topographic map. The actual ground condition of each pixel was then identified as forest or nonforest by noting the point position on the topographic map. Any pixel located on or within one half pixel of a forest/nonforest boundary was noted as a border pixel.

Table 1.1 lists the results of the performance evaluation. The forest/nonforest mask portrayed actual ground conditions accurately in this assessment which estimated performance at the 95% confidence level. The greatest source of error in the mask, border pixels not withstanding, occurred in the forest class. Hence, if the mask has an inherent bias, it is toward identifying nonforest areas as forest.

Table 1.1. Classification performance of the Bayesian two-channel forest/nonforest mask. Numbers indicate actual number of pixels unless otherwise noted.
NOTE: Accuracies given are within ±5% of the true accuracy at the 95% level of confidence.

	Bayesian Forest/Nonforest	
Actual Ground Conditions	Forest	Nonforest
Forest	83	2
Nonforest	10	97
Border (Predominantly forest)	15	5
Border (Predominantly nonforest	7	11

Calculation of Overall Accuracy (including border pixels):

$$\frac{83+15+97+11}{230} = 89.56\%$$

Calculation of Overall Accuracy (excluding border pixels):

$$\frac{83+97}{192} = 93.75\%$$

APPENDIX II

Selection of Defoliation Assessment Procedure

NOTE: Sections of this appendix were extracted from:
Nelson, Ross F. 1981. ASSESS2 Analysis of Four Methods for Classifying Forest Defoliation (Revised). Goddard Space Flight Center/Earth Resources Branch Internal Report, Greenbelt, MD. 11 pgs.

STUDY OBJECTIVE

The purpose of this study was to select the appropriate gypsy moth defoliation assessment procedure using Landsat digital data and computer-aided analysis techniques. Four image processing techniques were examined. These included a supervised classification procedure, two vegetation indices developed initially for agricultural biomass estimation and a data transformation technique designed by the Calspan Corporation.

DATA AND STUDY SITE DESCRIPTION

NASA/GSFC and BOF/DFPM selected a study site for this activity that was located west of Harrisburg, PA. The boundaries of the study site corresponded to the 7-1/2" U.S.G.S., Wertzville Topographic Quadrangle Map. The Wertzville area is located in the Ridge and Valley physiographic region of the Appalachians. The mountains are heavily forested and subject to gypsy moth attack. During the 1977 gypsy moth summer feeding cycle, this area experienced extensive defoliation.

Landsat data over the study site were obtained on June 27, 1977. Cloud cover at that time was minimal. Many large sections of heavily and moderately defoliated forest were noticeable in these data. In addition to the Landsat data, air photos were collected over the Wertzville area within one week of the satellite overpass. Division of Forest Pest Management personnel interpreted these photos to delineate areas of moderate and heavy defoliation. The defoliated area boundaries were transferred onto the 1:24,000 U.S.G.S. map and were later digitized to become a component of the ground reference image used to assess the results of the various image processing techniques examined.

A second nondefoliated Landsat image was required for this activity to generate a forest/nonforest mask (see Appendixes I and VI). The mask was created using a Bayesian thresholding procedure on a Landsat data set collected July 19, 1976. This

data set had been geometrically corrected and resampled to overlay the 1977 data. An accuracy assessment of the forest/nonforest mask indicated that forested pixels were correctly identified 89.6% of the time. This forest/nonforest mask also constituted a component of the ground reference image used to assess the results of the various image processing techniques examined.

PROCEDURES

The following classification approaches were tested to determine which one(s) best delineated gypsy moth defoliation:

1. Bayesian Supervised Classification (BAYES)
2. Ratio Vegetation Index (RVI = $MSS7/MSS5$)
3. Transformed Vegetation Index (TVI = $\sqrt{(MSS7 - MSS5/MSS7 + MSS5) + 0.5}$)
4. Calspan Mathematical Transformation (CALSPAN)

The four image processing techniques selected for this activity had been previously identified in the remote sensing literature by analysts examining the use of Landsat digital data for defoliation assessment (Williams, Stauffer, and Leung 1979). A description of each of these analysis procedures as well as the procedure to generate the Ground Reference Image (GRI) is given below. The results of each of the four classification approaches were compared to the GRI.

GRI - The Landsat-derived forest/nonforest mask and digitized defoliation map derived from aerial photography were registered and combined to produce the Ground Reference Image. Any discrepancies between the forest/nonforest mask and digitized information were rectified in favor of the mask. Hence, if a pixel was identifed as nonforest in the mask, but considered moderately defoliated in the digitized image, its ground reference image identity was nonforest. The decision for adjusting the GRI to match the Landsat classification map was based on the procedure used in photointerpretation.

Analysts would routinely outline broad areas of defoliation on air photos. Occasionally these areas would include small pockets of non-forest. Therefore, airphoto interpretation errors would be likely when examining areas the size of one pixel. By rectifying discrepancies in favor of the Landsat forest/nonforest mask these errors were avoided. The final image product contained four classes:

0 - nonforest
1 - heavily defoliated forest (60-100% canopy removed);
2 - moderately defoliated forest (30-60% canopy removed);
3 - healthy forest (0-30% canopy removed).

BAYES - A supervised classification of the Wertzville study area was completed using the June, 1977 Landsat data exhibiting defoliation. The data were first registered to the July 1976 data (from which the forest/nonforest mask had been produced). The mask was applied to the 1977 imagery to create the masked, defoliated image. Training fields were identified in heavily defoliated, moderately defoliated and healthy forest on the "defoliated image". The location of each of these training fields was obtained from the defoliation map generated by air-photointerpretation. Training statistics were developed from the Landsat data and were then input to a Bayesian classification program to classify the Landsat data into the three previously mentioned forest classes, plus a nonforest category. The final image product contained the same four classes as those listed for the GRI.

RVI - The Ratio Vegetation Index was applied to the same masked, defoliated Landsat data generated for the Bayes test above. The forest ratio values were normalized to a 0-100 scale. The scale was roughly equivalent to crown closures where, low numbers indicated heavy defoliation, high numbers denoted healthy forest, zeros represented nonforest. To determine the

numerical cut-off points between the healthy, moderately defoliated, and heavily defoliated forest, the ratio values in each of the ground reference classes were individually histogrammed to determine their respective frequency distributions. Graphs were drawn and the cut-off points determined. The unmasked portion of this masked image was then classified into the three forest classes based upon the derived cut-off points.

TVI - Processing steps for the TVI defoliation image were identical to those outlined for the RVI procedure. The Transformed Vegetation Index values for each forest pixel of the masked, defoliated image were calculated using the TVI formula. The resulting image was rescaled from 0-100 and histograms were generated to determine the numerical cut-off points for each forest class. The TVI image was then classified into three forest classes plus the nonforest class which resulted from masking.

CALSPAN -MSS Bands 4, 5 and 6 were used in two mathematical transformations formulated by the Calspan Corporation. These transformations were applied to the June 1977, Landsat image only. The calculations resulted in a second image containing 10 classes. The healthy and defoliated areas were individually histogrammed (as above) and their distributions graphed. The cut-off points were determined, the image categorized into 3 classes, and the forest mask was then applied.

Cut-off points for each of the last three procedures are listed in Table II-1.

Table II-1. Cut-off points for defoliation levels and healthy forests calculated from the RVI, TVI and CALSPAN procedures.

	CUT-OFF POINTS			
	Healthy Forest	Moderate Defoliation	Heavy Defoliation	Non-Forest
Ratio Vegetation Index	81-100	37-80	1-36	0
Transformed Vegetation Index	89-100	60-88	1-59	0
Calspan Transformations	1-3	4-6	7-9	0

RESULTS

The Ground Reference Image was compared to the BAYES, RVI, TVI and CALSPAN defoliation assessments. Table II-2 list the per-pixel classification performance of each image processing technique.

Table II-2. Per-pixel classification performance values for the BAYES, RVI, TVI and CALSPAN defoliation assessments (Hthy = healthy forest; Mod = moderately defoliated forest; Hvy = heavily defoliated forest).

Percentage of Pixels Classified into each Category

	BAYES			RVI			TVI			CALSPAN		
	Hlthy	Mod	Hvy	Hlthy	Mod	Hvy	Hlthy	Mod	Hvy	Hlthy	Mod	Hvy
Hlthy	47.6	40.6	11.9	56.6	30.9	12.5	48.8	39.5	11.7	60.9	17.3	21.8
Mod	14.1	74.4	11.6	21.2	64.9	14.0	14.1	72.1	13.8	42.0	28.6	19.3
Hvy	0.4	13.1	86.5	0.1	11.4	88.5	0.1	11.9	88.0	6.6	17.0	76.4
	Overall Accuracy: 51.0%			Overall Accuracy: 58.1%			Overall Accuracy: 51.9%			Overall Accuracy: 58.2%		

Most of the four image processing techniques tended to classify moderately defoliated and heavily defoliated areas correctly at the expense of healthy forest, thus reducing the overall classification accuracy. This reduction may be explained, in part, by the

separability between healthy forest and moderate defoliation. These two cover types have similar responses in each of the four Landsat spectral bands. Consequently most of the misclassified healthy pixels were classified as moderate and vice-versa. Heavily defoliated areas are the most accurately identified for each technique.

In view of the problems encountered because of the spectral similarity between healthy forests and moderately defoliated forests, the selection of an appropriate defoliation assessment technique needed to be based not only on overall classification accuracy, but also the relative performance in each class. For example, although the CALSPAN technique achieved the highest overall accuracy, the low performance value for moderate defoliation (28.6%) made this technique unacceptable for defoliation assessment. Upon examination of per pixel accuracies for each image processing technique for each forest class, the Ratio Vegetation Index procedure was judged to be the most appropriate procedure for defoliation assessments. Although other approaches produced higher accuracies in the healthy and moderately defoliated classes, the RVI procedure was the only analysis technique to classify over 50% of the pixels correctly in all three classes.

REFERENCES

Williams, D.L., M.L. Stauffer, and K.C. Leung. 1979. A Forester's Look at the Application of Image Manipulation Techniques to Multitemporal Landsat Data. Proceedings, 5th Annual Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, IN pg. 368-375.

APPENDIX III

Evaluation of Defoliation Survey Techniques

STUDY OBJECTIVE

The Ratio Vegetation Index (RVI) was selected as the appropriate procedure for digitally analyzing Landsat MSS data to assess levels of defoliation (see Appendix II). Further analysis was necessary to determine the accuracy of computerized defoliation assessment. The specific objectives of this evaluation were twofold:

1. Determine the appropriate ratio threshold values to be used in the RVI classification procedure such that the defoliation assessment accuracy was maximized.

2. Compare the mapping accuracy of several defoliation assessment techniques. The techniques included:

   a. Landsat MSS classification using the RVI;
   b. photointerpreted results using 1:80,000 color infrared airphotos; and
   c. aerial sketchmapping.

STUDY SITE AND DATA DESCRIPTION

The study area, Doubling Gap, Pennsylvania, lies approximately 50 kilometers west of Harrisburg in the ridge and valley region of the Appalachian Mountains. The mountains are heavily forested, the predominant cover type is oak-hickory. The area was heavily defoliated in 1981.

The Landsat-2 satellite collected multispectral scanner data over this region (path 16, row 32) on July 11, 1981. The scene (ID 22362-15035) is cloud free. Landsat MSS data were also collected over this area on July 19, 1976. At that time, no gypsy moth defoliation was noted in the scene (scene ID 2544-15001). This earlier, healthy Landsat data set was registered to the 1981 data; both were ultimately registered to the USGS 7-1/2 minute map base. The 1976 Landsat MSS data were used to produce a forest/nonforest mask in which all nonforest areas were set to zero and all forested pixels equal one.

Color infrared aerial photos (1:80,000) were acquired within hours of the July

11, 1981 Landsat overpass. Pennsylvania Division of Forest Pest Management personnel delineated areas of moderate (30-60% canopy removal), heavy (60-80%), and severe (80-100%) defoliation on the air photos. A Zoom Transfer Scope was used to transfer the photointerpreted information to two USGS 7-1/2 minute topographic maps (Andersonburg and Landisburg). This data was digitized and rasterized to form a 243 line by 372 sample defoliation image (57 meter pixel). The heavy and severe defoliation classes on the maps were digitized as one class-heavy defoliation, 60-100% canopy removed. Hence, the defoliation image (the raster photointerpretation image) consisted of only two classes, moderate defoliation (30-60% canopy removed) and heavy defoliation (60-100% canopy removed). This defoliation image was combined with the Landsat-generated forest/nonforest mask to produce the airphotointerpretation ground reference image. This image contained four classes: 0-nonforest, 1-heavy, 2-moderate defoliation, and 3-healthy forest.

Division of Forest Pest Management personnel collected aerial sketchmapping data over the Doubling Gap area on July 6, 1981. The sketchmappers outlined areas of moderate and heavy defoliation on the Andersonburg and Landisburg 7-1/2' quadrangle maps. The aerial sketchmapping results were digitized and rasterized. The digital sketchmapping results were combined with the Landsat-generated forest/nonforest mask to produce an aerial sketchmap image. The image classes were the same as those found in the airphoto interpretation ground reference image (i.e., 0-3).

To summarize, four data sets (all registered to a 7-1/2 minute map base) were produced from various data sources for further manipulation:

1. 1981 Landsat MSS data depicting defoliation conditions,
2. A forest/nonforest mask derived from an earlier Landsat scene which contained no defoliation;
3. The airphotointerpretation ground reference image, derived from photointerpretation results and the forest/nonforest mask; and

4. The aerial sketchmapping image, derived from aerial sketchmapping results and the forest/nonforest mask.

Throughout the analysis, the airphotointerpretation image was considered to be the truest representation of the actual ground conditions. As such, the airphotointerpretation image served as a ground reference image.

PROCEDURE

The airphotointerpretation image was used to define the appropriate RVI thresholds. Zero/one masks were made for each airphotointerpretation cover type (i.e., healthy forest, moderate, and heavy defoliation). The moderate defoliation mask, for instance, contained ones in moderately defoliated areas and zeros elsewhere. The 7/5 ratio values were computed for the 7/11/81 MSS data. This 1981 RVI image was multiplied by each mask to produce three different, masked, RVI images; i.e., the ratios of healthy forest, moderate defoliation, and heavy defoliation. These three images were histogrammed and the distribution of ratio values were noted for each cover type. The RVI cover type thresholds were defined as the points of intersection of the cover type distribution curves.

The cross-classification problem involving healthy forest and moderate defoliation has been well documented. Attempts to separate these two forest cover types significantly reduce classification accuracy. In this study, two different sets of thresholds were sought, one which most reliably detected heavy defoliation, moderate defoliation, and healthy forest, and the second which most accurately separated heavy defoliation from a healthy-moderate cover type.

Once the optimal thresholds were defined, the transformed (band 7/band 5), 1981 Landsat MSS data were classified (thresholded). The RVI classification was compared to the airphotointerpretation results to determine percent agreement.

STUDY RESULTS

A. Separating Healthy Forest, Moderate Defoliation, and Heavy Defoliation

Figure III-1 depicts the RVI response distribution for the three cover types and the ratio cutoffs between healthy forest, moderate defoliation, and heavy defoliation. The cutoffs follow:

Heavy defoliation: 0.001-3.85 (0 = nonforest);
Moderate defoliation: 3.86-5.10;
Healthy Forest: 5.11- ;
NOTE: 128 greylevels in all four channels.

These thresholds were used to produce a Landsat defoliation assessment image. The Landsat classification was compared to the airphotointerpretation image. The results of that comparison are given in Table III-1.

Table III-1. Comparison of Landsat classification to airphotointerpretation image (assumed closest to actual ground conditions). Table entries are percent agreement.

		Airphotointerpretation Image		
		Heavy Def.	Mod. Def.	Healthy
Landsat Classification	Heavy Def.	72.78	16.20	24.21
	Mod. Def.	25.37	53.42	34.08
	Healthy	1.85	30.38	41.71
	Total percent	100.00	100.00	100.00
	No. of pixels	19789	18212	12955

Average Accuracy: 55.97%
Overall Accuracy: 57.96%

Of concern was the extremely low classification accuracy of the healthy forest class. In order to more accurately classify healthy areas (at the expense of moderate defoliation classification accuracy), the healthy-moderate threshold was dropped to 5.00. At this threshold, the number of pixels classified as healthy most closely matched the number of healthy pixels identified in the airphotointerpretation data (which served as the "ground reference" data set). The altered thresholds and the results stemming from this adjustment are given in Table III-2.

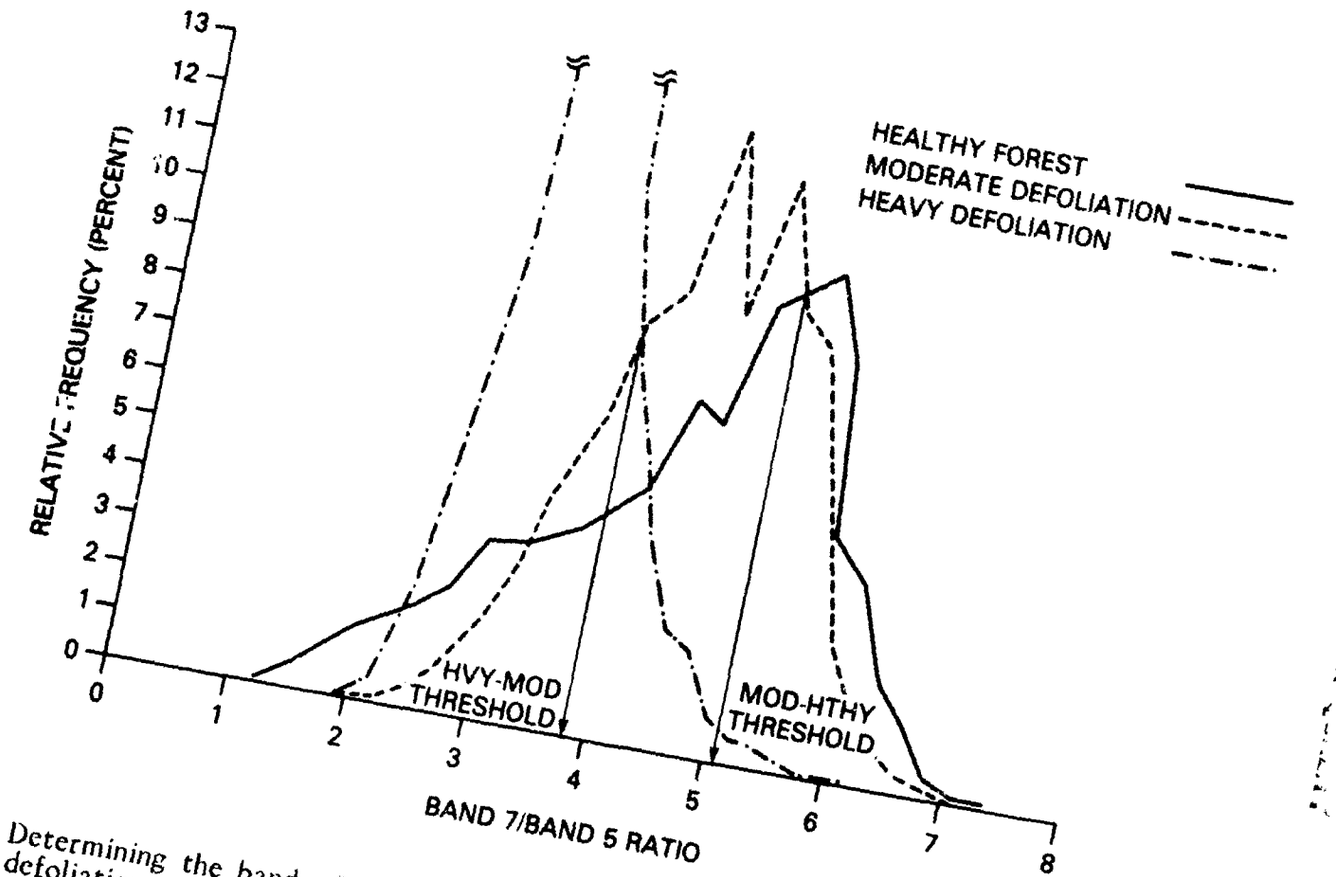


Figure III-1. Determining the band 7/band 5 ratio threshold between heavy defoliation, moderate defoliation, and healthy forest. The relative frequency distribution for each class is shown.

Table III-2. Landsat classification vs. airphotointerpretation results, revised thresholds: heavy defoliation (0.001-3.85), moderate defoliation (3.86-5.00), and healthy forest (5.01-). Table entries are percent agreement.

		Airphotointerpretation Image		
		Heavy Def.	Mod. Def.	Healthy
Landsat classification	Heavy Def.	72.78	16.20	24.21
	Mod. Def.	24.80	48.61	30.13
	Healthy	2.42	35.19	45.66
	Total percent	100.00	100.00	100.00
	No. of pixels	19789	18212	12955

Average Accuracy: 55.68%
Overall Accuracy: 57.25%

The slightly increased healthy forest accuracy and the areal agreement (in terms of number of pixels) may justify the small reduction in the summary accuracies.

Two outstanding characteristics were noted in Tables III-1 and III-2. First, the classification accuracies of the individual classes were decidedly low. Second, as in previous work, cross-classification problems arose between adjacent classes. The misclassification problem was most noticeable between the healthy forest and moderate defoliation cover types. In order to improve classification performance, the healthy-moderate classes were condensed to form a single class. The ability to separate this healthy-moderate class from heavy defoliation is documented in the following section.

B. Separating Heavy Defoliation from Healthy Forest

The operational utility of Landsat may best be realized by using the data to discriminate spectrally separable cover types. The appropriate threshold for delineating heavy defoliation from forest classified as healthy and moderately defoliated is shown in Figure III-2. The thresholds follow:

Heavy Defoliation: 0.001-3.95 (0=nonforest),
Healthy Forest (includes Moderate): 3.96- .

The agreement matrix comparing the Landsat classification with the airphotointerpretation

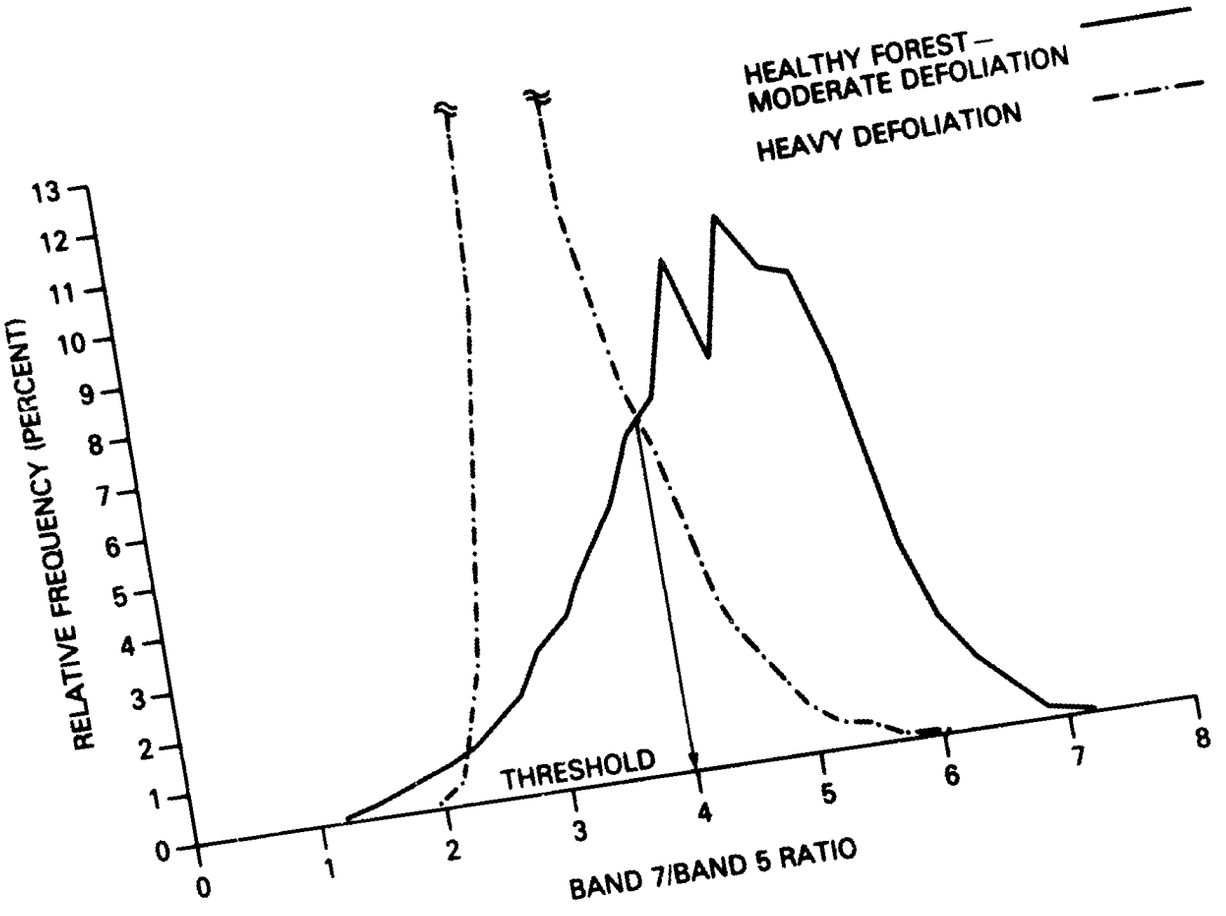


Figure III-2. Determining the band 7/band 5 ratio threshold between heavy defoliation and a healthy forest-moderate defoliation cover type. The relative frequency distribution for each class is shown.

data is given in Table III-3.

Table III-3. Delineating heavy defoliation from healthy forest, Landsat classification vs. photointerpretation. Table entries are percent agreement.

		Airphotointerpretation Image	
		Heavy Def.	Healthy
Landsat Classification	Heavy Def.	77.94	22.48
	Healthy	22.06	77.52
	Total percent	100.00	100.00
	No. of pixels	19789	31167
	Average Accuracy:	77.73%	
	Overall Accuracy:	77.68%	

The increased class accuracies reflect the spectral uniqueness of heavily defoliated forest. Hence the operational utility of the MSS data lies with the separation of two (heavy-healthy) rather than three (heavy-moderate-healthy) forest cover types.

C. Comparison of Aerial Sketchmapping and Airphotointerpretation

An equitable evaluation demanded that alternate methods of assessing insect damage be tested to determine if Landsat data analysis truely was "better". The 1981 skethmapping results were compared to the photointerpretation data (see Table III-4 and III-5).

Table III-4. Aerial sketchmapping vs. airphotointerpretation results, three forest classes. Table entries are percent agreement.

		Airphotointerpretation Image		
		Heavy Def.	Mod. Def.	Healthy
Aerial Sketchmap	Heavy Def.	91.43	62.53	16.42
	Mod. Def.	6.06	20.02	10.98
	Healthy	2.51	17.45	72.60
	Total percent	100.00	100.00	100.00
	No. of pixels	19789	18212	12955
	Average Accuracy:	61.35%		
	Overall Accuracy:	61.12%		

Table III-5. Aerialsketchmapping vs. airphotointerpretation results, two forest classes. Table entires are percent agreement.

		Airphotointerpretation Image	
		Heavy Def.	Healthy
Aerial Sketchmap	Heavy Def.	91.43	43.36
	Healthy	8.57	56.64
	Total percent	100.00	100.00
	No. of pixels	19789	31167
	Average Accuracy:	74.04%	
	Overall Accuracy:	70.15%	

A comparison of Table III-2 and III-4 indicate that aerial sketchmapping delineated the three forest classes more accurately than Landsat. When Tables III-3 and III-5 were compared, it was evident that Landsat did a better job defining two classes (i.e., delineating a healthy-moderate class from heavy defoliation). Aerial sketchmapping seemed to overestimate the amount of heavily defoliated area at the expense of healthy forest.

SUMMARY

The Pennsylvania Landsat data base may be accessed using a user-friendly front-end system designed to accommodate non-remote sensing personnel. Should these people wish to produce a forest defoliation map using Landsat data, "canned" job stems will be available which will specify the necessary processor parameters. In order to produce such a classification, class ratio values must be specified. It is suggested that the thresholds listed in Table III-6 and III-7 be used in the default or "canned" job stems. The thresholds are given in terms of the actual 7/5 ratio value (as used in this study) and in terms of the equivalent byte threshold. The byte thresholds were computed by linearly interpolating the ratio thresholds on a scale of 0 to 255. The largest forested 7/5 ratio value in the Doubling Gap 7/11/81 Landsat imagery was 7.833. This was considered the high end of the 7/5 ratio

scale. Similarly, 255 was considered the high end of the byte scale. Zero marks the low end of both scales. Hence the byte equivalent of a ratio threshold of 5.01 is ((5.01/7.833)256)-1 = 162.74, i.e., 163.

Table III-6. Suggested ratio and byte image thresholds to delineate healthy forest, moderate defoliation, and heavy defoliation.

	Ratio		Byte	
Cover Type	Low Thresh	High Thresh	Low Thresh	High Thresh
Heavy Def.	0.001	3.86	1	125
Mod. Def.	3.86	5.01	125	163
Healthy Forest	5.01		163	255

Table III-7. Suggested ratio and byte image thresholds to delineate healthy forest and heavy defoliation.

	Ratio		Byte	
Cover Type	Low Thresh	High Thresh	Low Thresh	High Thresh
Heavy Def.	0	3.96	0	128
Healthy Forest	3.96		128	255

It is suggested that, if possible, the actual ratio values (leftmost 2 columns, Tables III-6 and 7) be used for thresholding. The ratio values are absolute values, the byte values are on a relative scale, a scale which changes with changes in the maximum forested ratio value. Hence application of byte thresholds to different data sets may yield more inconsistent results.

Analysis of the accuracy of classification has shown that low classification accuracies (below 50%) may be expected for moderate defoliation and healthy forest if three forest classes are delineated. Heavy defoliation, in this study, was classified correctly better than 70% of the time. Aerial sketchmapping produced results which more closely represented photointerpreted ground conditions of the three forest classes, but even using this method moderate defoliation was classified very poorly (~20%).

Landsat data analysis proved more accurate than aerial sketchmapping when

concern lay with only two forest classes, heavy defoliation and healthy forest (moderate defoliation-healthy forest conglomerates). Both cover types were classified correctly better than 70% of the time. In an operational context, delineation of two forest classes may be more realistic.

Preliminary work concerning the temporal stability of the ratios dictates a word of caution. The ratio breakpoints suggested above were derived from July 11, 1981 Landsat data. The application of these ratios to July 30, 1981 data produced a classification in which the extent of moderate defoliation was significantly overestimated. These ratios seem to be dependent on the time of data acquisition. The ratio cutoffs which would produce the most accurate classification will vary from scene to scene. The cutoffs suggested may be used as a "first-cut", but threshold adjustments may be necessary.

APPENDIX IV

Identification of Temporal Window for Defoliation Assessment

NOTE: Sections of this Appendix were extracted from:
Nelson, R.F. 1981. Defining the temporal window for monitoring forest canopy defoliation using Landsat. Proceedings of the Annual Symposium on Remote Sensing ASP-ACSM, Washington, D.C. pp. 367-382.

STUDY OBJECTIVE

An operational defoliation assessment system incorporating Landsat data requires useful satellite information. Previous studies have encountered problems obtaining cloud-free Landsat data during the peak defoliation periods in Pennsylvania. If quality Landsat imagery cannot be obtained during this optimum viewing period, another source of data must be used. The purpose of this activity was to define the temporal limits within which Landsat data might be obtained and still provide useful defoliation information.

BACKGROUND ON GYPSY MOTH POPULATION DYNAMICS

The length of time in which gypsy moth defoliation is discernible on Landsat multispectral imagery is dependent on two factors: (1) the life cycle of the insect, and (2) the response of the forest to infestation. The first factor, the insect's life cycle, actually defines the temporal progression of forest canopy destruction. Gypsy moths overwinter as eggs and larvae emerge in late April or early May. The larvae (caterpillars) begin feeding immediately. As the larvae periodically molt and grow larger, greater quantities of leaves are consumed. The amount of canopy removed by the caterpillars increases to the point where leaf loss may be detected by Landsat. This point in time marks the beginning of the temporal limits within which Landsat data might be obtained and still provide useful defoliation information.

The gypsy moth caterpillar will continue to feed until mid-summer when the insects pupate and transform into adult moths that mate and lay eggs. Hardwoods that have lost more than 60% of their foliage refoliate in July and early August. Hence, the visual effects of defoliation are lessened as the canopy is restored. The ability of the hardwood forest to rejuvenate at least a portion of its canopy precludes the use of Landsat data for defoliation assessment after a certain date. This date identifies the end point of the temporal limits for defoliation assessment.

DATA AND STUDY SITE DESCRIPTION

The study site selected for this activity was located along Bald Eagle Mountain near Williamsport, Pennsylvania. This area is dominated by hardwood forests and is subject to periodic gypsy moth infestations. During the 1977 gypsy moth summer feeding cycle, six relatively cloud-free Landsat images were collected over Williamsport. The Bald Eagle Mountain study site was extracted from each image for the temporal analysis. Each of the extracted image subsections were geometrically registered to one another to insure that identical areas were selected. A description of each Landsat image and study site subsection is listed in Table IV-1.

Table IV-1. Description of Landsat imagery used in temporal analysis.

Date	Scene ID	Subsection Coordinates		Image Quality
		Start Line	Start Sample	
May 22	2851-14532	1830	2850	Clear
June 8	2868-14471	1	275	Clear
June 27	2887-14513	1665	2320	Scattered Cumulus Clouds
July 2	5805-13954	1971	570	Clear
July 14	2904-14450	1920	300	Scattered Cumulus Clouds
August 2	2932-14494	1605	2330	Clear

In addition to the Landsat imagery, 1977 aerial sketch maps over the study area were available. These maps were generated by BOF/DFPM personnel from aerial surveys during which they identified areas of moderate and heavy defoliation. The maps were used to identify twenty-five study blocks of varying sizes located within heavily defoliated (60-100% canopy removal), moderately defoliated (30-60% canopy removal) and healthy forest.

PROCEDURE

The May 22, June 8, July 2 and 14, and August 2 Landsat subsections were registered to the June 27, 1977 sub-image using the General Electric Image 100 (General Electric, 1975) scene registration utility program. Each was registered using 16 control points scattered throughout the study area. The twenty-five study blocks selected from aerial sketch maps were identified on the June 27 sub-image using USGS 1:24,000 topographic maps. Five additional blocks were situated in areas thought to be "constant" reflectors. The five blocks, located in the city of Williamsport, were monitored to evaluate the scene-to-scene variability caused by factors other than those related to vegetation changes.

Ideally, given identical viewing conditions, the reflectance of a spectrally constant landmark should be constant. Urban areas, though not constant, are stable enough to give the analyst an idea of scene variations due to factors such as haze or dust, changing sun angle, and satellite or preprocessing discrepancies. Such indications are useful when assessing seasonal forest changes. The study block information is summarized in Table IV-2.

Table IV-2. Study Sites Selected in the Williamsport - Bald Eagle Mountain Area.

Cover Type	Aspect	Number of Study Blocks	Size of Study Block (Pixels)
Heavy Defoliation	South	7	36
	North	3	36
Moderate Defoliation	South	3	36
	North	6	36
	Flat	1	36
Healthy Forest	South	2	
	Flat	2	9
		1	25
Constant Reflectors	Flat	5	9

The average spectral responses of each study block was determined for each date throughout the 1977 summer. The analysis of the data and the results of the study are given below.

RESULTS

A. Scene-to-Scene Variability - "Constant" Reflectors

The spectral characteristics of the five urban blocks were evaluated to determine if significant MSS response differences existed between dates. The band 5, band 7, and 7/5 ratio responses were tested to see if the between-scenes (between-dates) variability was statistically significant for the constant reflectors. Five of the six dates were evaluated using profile analysis techniques which require contrast computations of the form:

$$C_j = R_j - \frac{1}{n-1} \left(\sum_{i=1}^{n-1} R_i \right)$$

where C_j = contrast calculated for Date j.

R_j = the average response for all urban blocks for that date. The reponse could be the band 5 or band 7 MSS value, or the 7/5 ratio value.

n = number of dates analyzed, five in this case.

$\frac{1}{n-1} \left(\sum_{i=1}^{n-1} R_i \right)$ = the average response of all dates other than date j.

The contrast for a particular date was calculated by subtracting the average response value for that date from the average response values of the remaining dates. The July 14th data contained some cloud cover and therefore could not be analyzed due to the missing data. Pairwise - T statistics were computed for all pairs of contrasts; Hotelling tests were used to determine the significance of the between-dates variability.

The results of the Hotelling tests are noted in Table IV-3. Traditionally, scientists have used the 95% confidence level to accept or reject a null hypothesis. The p value is the probability remaining in the tail of the F distribution (to the right of the calculated F). If p is greater than 0.05, we accept the null hypothesis

that there are no significant response differences between dates when data from constant reflectors are analyzed. If p is less than 0.05, we reject the null hypothesis and conclude that significant differences exist. The results indicated that there were no significant differences between dates for the band 5, band 7, and ratio response values at the 95% confidence level. The p value gives a measure of the significance of the variable in question. As expected, the 7/5 ratio term has the largest p value, indicating that it reduces between-scenes variability.

Table IV-3. Results of the profile analysis of the urban study sites for band 5, band 7, band7/band5 ratioes.

Variable	$F_{Calculated}$	Degrees of Freedom (Denominator)	P
Band 5	50.35	4	0.105
Band 7	13.54	4	0.201
Ratio	8.90	4	0.246

B. Determining the Temporal Limits for Defoliation Assessment

The analysis of constant reflectors indicated that spectral variability among the Landsat subsections would not be caused by the scene temporal differences. Therefore, spectral variability should be caused by changes in ground cover conditions. The spectral response patterns for the 25 moderately defoliated, heavily defoliated and healthy forest study blocks were examined for each date to determine those dates within which these cover types could be spectrally separated from each other.

Figure IV-1 illustrates the spectral characteristics of each forest class for a single date (June 27, 1977) bands 5 and 7. Note that the spectral response pattern for healthy forest and moderate defoliation are nearly identical. In fact, this was the

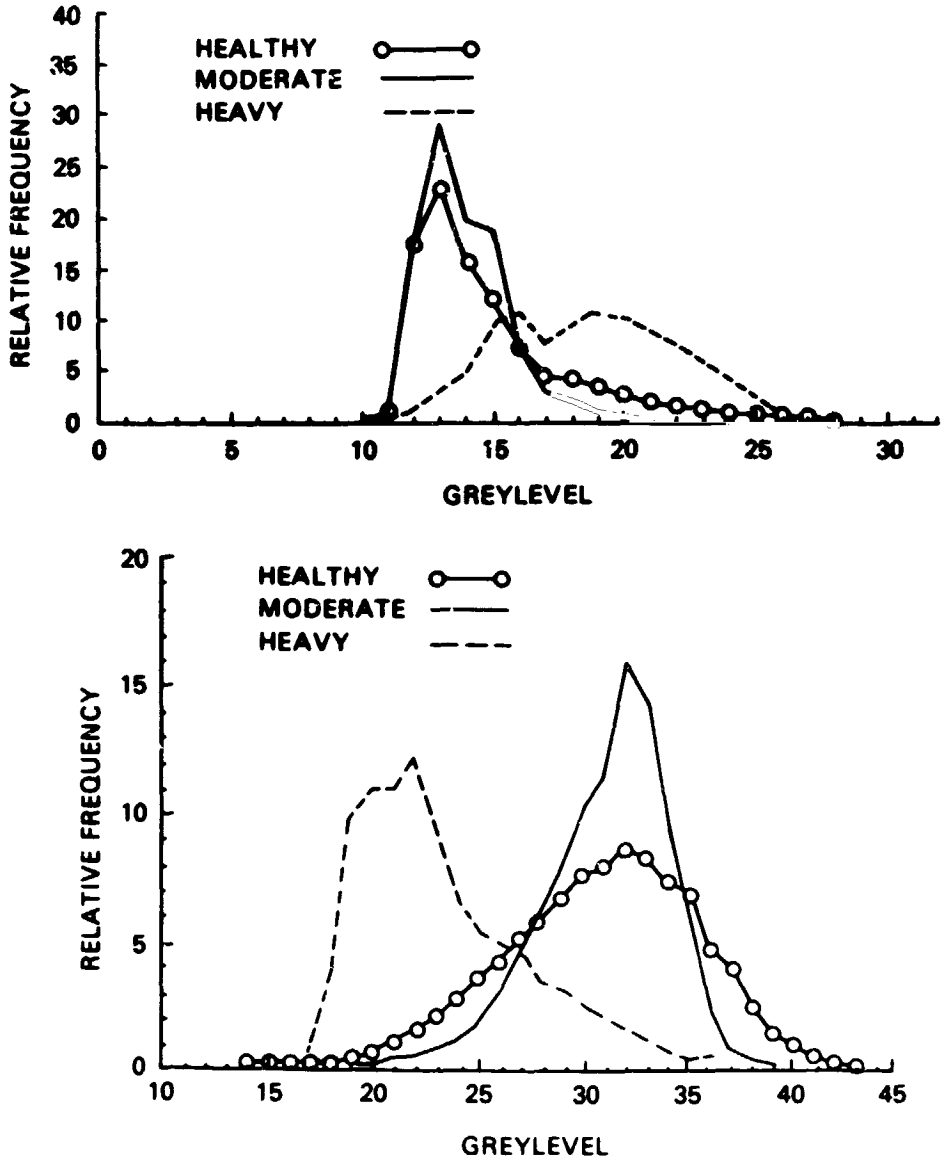


Figure IV-1. The relative frequency of MSS pixel values for moderately defoliated, heavily defoliated, and healthy forest for bands 5 and 7. The Landsat data was acquired on June 27, 1977.

case for all dates examined. These findings explain why previous research had shown that healthy forest and moderately defoliated areas are consistently confused regardless of the technique used to digitally classify the area.

Having determined that healthy forests and moderately defoliated forests are spectrally similar and cannot be reliably separated using Landsat data, the definition of a temporal window for defoliation assessment concentrated on the separability between heavily defoliated forests and other cover types. The relationship between the MSS_7/MSS_5 ratio values calculated from data obtained over moderately and heavily defoliated forest, on north and south facing slopes for each of the six dates in 1977 is shown in Figure IV-2. The graphs show that heavy defoliation can be easily distinquished from moderate defoliation, from June 8, through mid-July, on both north and south facing slopes. The greatest separability between these classes occurred in late June and early July. These dates corresponded to the 1977 peak defoliation period.

The results of this activity indicated that heavily defoliated forests can be reliably defined on Landsat data within a two month window which roughly centers on the period of peak defoliation. The ability to separate healthy forests and moderate defoliation still remains problematic. Williams and Stauffer (1979) suggested that topographic information might help delineate moderate defoliation on south slopes from healthy forest on north slopes. These results indicated that the response distributions of moderately defoliated and healthy areas, regardless of aspect, were so similar that topographic information would do little to diminish the confusion.

REFERENCES

General Electric Co. 1975. Image 100 Users Manual. Ground Systems Department, Space Division, Daytona Beach, Florida.

Williams, D.L., M.L. Stauffer and K.C. Leung. 1979. A Forester's Look at the Application of Image Manipulation Techniques to Multispectral Landsat Data. Proceedings, 5th Machine Processing of Remotely Sensed Data Symposium, Purdue University, West Lafayette, IN. pp. 368-375.

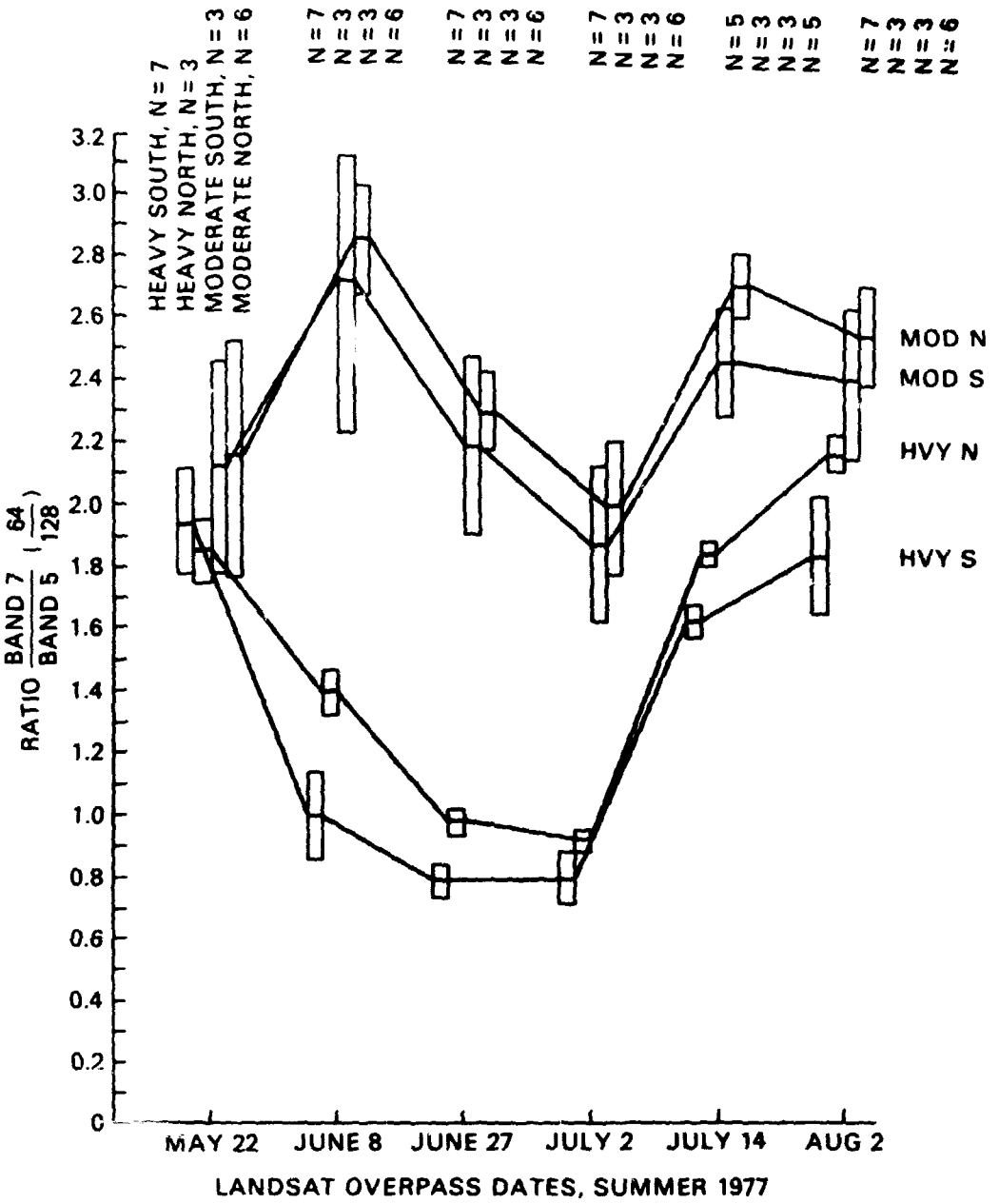


Figure IV-2. Relationship between ratios calculated from data obtained over moderately and heavily defoliated forest, two different aspects, for six dates in 1977. Note: N refers to the number of blocks used to calculate the average and standard deviation. Each block contained 36 pixels. One standard deviation is shown.

APPENDIX V

Landsat Digital Mosaic of Pennsylvania

NOTE: Sections of this appendix were extracted from:
Stauffer, M.L. and S.A. Russo. 1982. Characterization of the Registration Accuracy of the Pennsylvania Digital Mosaic. Computer Science Corporation Contract Report CSC/TM-82/6225, Goddard Space Flight Center, Greenbelt, MD.

OBJECTIVE

The creation of a geometrically corrected Landsat digital mosaic for the State of Pennsylvania was an essential element for the operational defoliation assessment system. This mosaic is the foundation of the Landsat-derived geographic data base and serves as the base data set for all subsequent processing. The Jet Propulsion Laboratory was contracted to generate the Pennsylvania mosaic according to the following criteria:

- Geometrically corrected to the Universal Transverse Mercator Map Projection
- Rotated to North
- Resampled to 57 meter square cells

DEVELOPMENT OF PENNSYLVANIA MOSAIC

The creation of a statewide Landsat digital mosaic for Pennsylvania was broken down into two major activities: (1) a demonstration of JPL capabilities and (2) the actual mosaic generation. In each of these activities, GSFC project personnel were interested in measuring the geometric accuracy of the products (i.e., image to map registration) and the image-to-image registration of the products.

Demonstration of JPL Capabilities

JPL personnel were asked to demonstrate the technical feasibility of creating the statewide Landsat digital mosaic during the first year of the Joint Research Project. During this demonstration phase, JPL digitally joined two adjacent Landsat scenes (north/south pair) acquired over Pennsylvania and reprojected each Landsat frame to UTM with an image raster rotated north to align with the UTM projection. They then registered two coincident Landsat images acquired on a different date to the initial map base imagery. Upon completion of the mosaic and registration, GSFC personnel evaluated the demonstration products to determine if the map projection,

mosaic and registration were adequate for the data base.

Qualitative evaluations of the mosaic and map projection indicated that the products either met or exceeded the standards outlined for the data base system. Seams between the adjacent frames showed no offset. Registration residual values supplied by JPL were less than two pixels. However, there were a number of problems evident with the scene to scene registration. A quantitative evaluation of the registration accuracy using analyst selected control points isolated portions of the mosaic image with offsets ranging from 6 to 10 pixels. Further qualitative evaluations indicated that this misregistration was not limited to isolated sections of the images but that varying degrees of line and sample offset occurred throughout the image.

Since accurate image-to-image registration is critical to the defoliation assessment procedure, it was necessary to determine the cause of these errors and make appropriate corrections. Upon inspection of the registered images, the analyst determined that areas with the largest registration errors contained numerous cumulus clouds which prevented the identification of selected ground control points.

The two problems associated with registration errors, cloud cover and software inadequacies, were corrected by upgrading software and using only cloud-free imagery. After these remedies were identified, JPL was contracted to generate a map-projected Landsat digital mosaic of Pennsylvania.

The Pennsylvania Statewide Mosaic

JPL generated the Pennsylvania Landsat mosaic using the ten scenes listed in Table V.1. GSFC project personnel selected these scenes after a comprehensive review of all summertime Landsat data acquired over the state from 1972 to 1980. The scenes used for the mosaic were selected using the following guidelines:

1. Summertime imagery acquired between May and September
2. Cloud-free data, or maximum cloud cover of 10%
3. No apparent defoliation or other forest disturbance
4. Most recently acquired data which met guidelines 1-3 above
5. Near anniversary coverage (i.e., all scenes from the same month of year, if possible).

Table V.1. Landsat Data Used for the Pennsylvania Mosaic

Path/Row	Scene Id.	Date
15/31	30179-15020	8/22/78
15/32	30098-15013	6/11/78
16/31	21660-15005	8/09/79
16/32	2544-15001	7/19/76
17/31	30478-15123	6/26/79
17/32	30208-15141	9/29/78
18/31	2600-15094	9/13/76
18/32	2600-15100	9/13/76
19/31	21267-15031	7/12/78
19/32	21267-15034	7/12/78

The basic requirements for the mosaic included: (a) registration to the UTM projection so that the image data set could be cross referenced to the United States Geological Survey (USGS) topographic map series; (b) 57m resolution to insure compatibility with future Landsat image products; and (c) rotation to north. The State of Pennsylvania lies within two UTM zones (UTM 17 and UTM 18). Therefore, two separate mosaics, one corresponding to each zone, were required. The evaluation of geodetic accuracy was completed for each zone separately and results will be reported for them as individual case studies.

GSFC received an MSS band 4 mosaic for the western half of Pennsylvania (UTM 17) from JPL in the fall of 1981. Upon receipt of the data, project personnel evaluated the integrity of the image-to-image mosaic to be sure that there were no

discontinuities at the seams between images. In addition, personnel reviewed the results of the geodetic accuracy assessment completed by JPL[1]. Based on this evaluation, JPL personnel were instructed to complete the mosaic for MSS bands 5, 6 and 7. The same procedure for the mosaic of the eastern half of Pennsylvania (UTM 18) was followed in December 1981.

During subsequent processing of the UTM 17 mosaic, project personnel noted several inconsistencies between the registered Landsat data and selected 1:24,000 scale USGS topographic maps. Some of these discrepancies could be attributed to differences between the UTM projection used for the Landsat imagery and the Polyconic projection used in the topographic map generation. These differences should have been remedied by simply offsetting the Landsat image to match the UTM grid lines rather than the borders of the topographic map. However, gross irregularities were still noted in the mosaic data and no consistent offset could be identified to match map and image features.

These problems in registration motivated project personnel to undertake a study to characterize the registration of the JPL mosaic. The initial study was conducted on the data for the western half of the state (UTM 17). Later studies focused on the mosaicked data for the eastern half of Pennsylvania (UTM 18).

Mosaic Geodetic Accuracy Assessment

Project personnel conducted two types of comparisons to evaluate the registration accuracy: a quantitative comparison based on the selection of ground control points and a qualitative comparison based on a visual assessment of the alignment between

[1]Geodetic accuracy was determined by examining registration "residual" values. That is, for a selected point, the deviation between its location in the mosaic and its precise location on the ground is its residual value.

the maps and scaled display products. Each of the comparisons provides unique information regarding the accuracy of the registration.

The procedure selected for the quantitative assessment of the mosaic was to select control points throughout each of the UTM zones and use these control points to register the mosaic to the UTM grid. The control points are used to derive a transformation equation which would be used to "fit" the image data to the UTM grid. If the JPL mosaic were properly registered, the appropriate coefficients of the transformation equation would be 1.00 and 0.00. Deviations from these expected values would indicate that the data are not registered properly and would provide a measure of the misregistration.

The mosaic for UTM Zone 17 has been divided into four quadrants that are roughly equivalent to the following USGS 1:250,000 scale maps.

QUAD 1 - Cleveland, Ohio
QUAD 2 - Canton, Ohio
QUAD 3 - Warren, Pennsylvania
QUAD 4 - Pittsburgh, Pennsylvania

The mosaic for UTM Zone 18 has also been divided into four quadrants that are roughly equivalent to the following USGS 1:250,000 scale maps:

QUAD 5 - Williamsport, Pennsylvania
QUAD 6 - Harrisburg, Pennsylvania
QUAD 7 - Scranton, Pennsylvania
QUAD 8 - Newark, New Jersey

A number of ground control points were chosen in each quadrant using selected 1:24,000 scale USGS topographic maps and their UTM coordinates were identified. The exact location of these points on the mosaic image was determined using a series of Interactive Digital Image Manipulation System (IDIMS) display functions. These locations were then used in an IDIMS registration function to determine the transformation function coefficients and ground control point residual values.

Table V.2 lists the number of control points selected from each quadrant and

a summary of the transformation results for each of the quadrants. In addition, the largest line and sample deviations found are listed as "Worst Case." Note the number of control points for Quads 5-8 are much higher than Quads 1-4. Upon completion of the registration assessment for UTM 17, project personnel felt that a more rigorous and comprehensive selection of control points was warranted. Therefore, the number and distribution of control points for UTM 18 (Quads 5-8) were increased.

Table V.2. Summary of Transformation Results for Quads 1-8. A(2), A(3), B(2), and B(3) are the transformation coefficients of the polynomial. RL is the average line residual (i.e., north/south direction), RS is the average sample residual (i.e., east/west direction. Expected Values for A(2) and B(3) are 1.00; expected Values for A(3) and B(2) are 0.00. The worst case values are the largest line and column misregistration values found for the pixels sampled. RL, RS, and worst case figures are in pixels, 1 pixel = 57 meters. (Taken, in part, from CSC Report #CSC/TM-82/6225.)

Quad	No. Points	A(2)	A(3)	B(2)	B(3)	RL	RS	Worst Case Line	Worst Case Column
1	7	1.00	-0.26e-2	-0.33e-2	0.99	0.65	0.67	1.7	1.3
2	6	1.00	-0.21e-2	0.46e-4	0.97	0.46	0.89	1.2	2.3
3	12	1.00	0.22e-4	0.35e-4	1.00	0.92	1.64	2.1	2.7
4	21	1.00	-0.26e-3	-0.11e-2	1.00	0.86	3.16	2.5	10.6
5	33	1.00	-0.28e-3	0.88e-3	1.00	1.08	1.13	3.2	4.7
6	40	1.00	-0.17e-3	0.15e-3	1.00	1.23	3.26	6.0	9.8
7	26	1.00	0.52e-3	-0.11e-2	1.00	0.65	0.63	2.7	1.6
8	23	1.00	0.11e-2	-0.11e-2	1.00	0.97	1.40	3.6	7.2

The results of the quantitative assessment for UTM 17 suggest that Quads 1 and 2 are closely registered to the UTM projection. However, the evidence suggests that Quads 3 and 4 are not accurately registered. The average line residuals, RL, the average sample residuals, RS, and the worst case errors are low for Quads 1 and 2. In addition, the coefficients for the transformation are acceptable--i.e., near 1.00 and 0.00 for the two quadrants. In Quads 3 and 4, the coefficients for the transformation are also near 1.00 and 0.00. However, the average sample residuals are above 1.0 (1.64 for Quad 3; 3.16 for Quad 4) indicating that registration errors occur in the sample direction. The sample misregistration problem is verified by

the worst case figures.

The results of the quantitative assessment for UTM 18 suggests that only one Quad, No. 7, is accurately registered to the map base. Both line and sample residuals are less than 1.00 for Quad 7. Line and sample residuals for Quads 5, 6, and 8 generally exceed 1.00. Sample residuals are higher than line residuals, suggesting that, as in UTM 17, offsets are generally greater in the sample direction. Again, this fact is illustrated by the large, localized errors found in all four quads. Coefficients for the transformation are near 1.00 and 0.00 for all four quadrants.

Qualitative assessments of the mosaic geodetic accuracy were made by overlaying gray scale computer printouts of the Landsat data onto respective 1:24,000 scale topographic maps. These assessments confirmed the quantitative results taken from residual and transformation coefficient values. Quads 1 and 2 provided acceptable fits to the topographic maps. Quads 4, 5, 6 and 8 exhibited offsets predominantly in the east-west sample direction. Quad 3 appeared to be a border line case, having considerably less offset problems than the four quadrants mentioned previously, but not exhibiting the same level of accuracy as Quads 1 and 2. Contrary to expectations based on the quantitative evaluation, the visual analysis of Quad 7 revealed gross localized errors in selected regions.

A second qualitative assessment was made using 1:250,000 scale U.S.G.S. topographic maps. Registration to the smaller scale maps appeared significantly better than for large scale maps because localized errors were less apparent.

CONCLUSION

Registration errors are more prominent in an east/west, or sample direction. This is also the along track scan of the satellite sensor. The MSS mirror scan is variable throughout the duration of the satellite mission, hence, information on the actual mirror scan velocity profile is inadequate and often inaccurate. This may

account for some of the registration errors. Since the satellite velocity is more stable in the north/south direction, fewer registration errors would be expected in the line direction.

The results of this assessment indicate that a user cannot expect to accurately cross reference points in a map and the mosaic at the single pixel level. However, the registration does appear to be sufficiently accurate to estimate the areal extent and location of defoliation by county or forest pest management district. At this scale the errors in boundary placement on the data are expected to have less impact than attempting to identify local features.

APPENDIX VI

Statewide Forest Classification Assessment

NOTE: Sections of this appendix were extracted from:
Russo, S.A. and M.L. Stauffer. 1983. The Statewide Forest/Nonforest Classification of Pennsylvania Using Landsat MSS Data. Submitted to the Proceedings of the American Society of Photogrammetry, March 1983, Washington, DC.

STUDY OBJECTIVES

A key element in the defoliation assessment procedure is the use of a forest mask to limit the areas searched for defoliation to those regions previously identified as forest. This is done to reduce the potential for misidentifying certain nonforest cover types as defoliation. A forest/nonforest mask was constructed for the entire state using procedures outlined in Appendix I. The purpose of this study was to provide an estimate of the accuracy of the forest classification statewide.

PROCEDURE

The ten Landsat scenes used to produce the statewide forest/nonforest mask are listed in Appendix V, Table V.1. Three image processing systems were used to manipulate these data for the mask:

- VICAR was used to compile statistics and perform classifications (Moik, 1979);
- The Image-100 was used to interactively select training sites (General Electric, 1975).
- The Interactive Digital Image Manipulation System (IDIMS) was used to conduct final checks on processing integrity (ESL, 1981).

Training and Classification

The forest mask was generated using a supervised approach to classification. The training site selection was simplified by the expanse of contiguous forest areas covering much of Pennsylvania and the broad similarity of the hardwood forests. The training sites were assumed to represent the spectral characteristics of the major forest areas. Since the classification was based on only a forest class, training sites were not selected for any nonforest cover types. The training site evaluation procedure insured that no nonforest areas were included. The training sites were evaluated using information obtained from the Image-100 and VICAR processing. The Image-100 was used to interactively select forest training sites and conduct preliminary

ıecks on their validity using frequency histograms. Training site selection depended :avily on the analysts' expertise in Pennsylvania land cover features and their ımiliarity with the appearance of the various vegetation cover types, particularly ›rest and agriculture, on the Landsat false color composite.

Lack of a maximum likelihood classifier and limited ability to handle large ata sets precluded use of the Image-100 for the classifications. Consequently, ırther processing was carried out using VICAR. The training site coordinates obtained n the Image-100 were input to the VICAR STATS program for computing statistics ›r the maximum likelihood classifier. These statistics were also used for deciding he acceptability of training sites. Initially, the statistics for known forest areas /ere acquired. The other training sites were qualitatively compared to these known sites ased on MSS_5 and MSS_7 means and variances. Based on the comparisons, training ites not similar to known forest areas were excluded. Since the utility of the tatewide forest mask depended on its timely availability, more rigorous training site election procedures were not implemented.

The number of training sites per scene varied between 17 and 43, averaging 30, for ı total of 297 statewide. Statistics for the training sites in each scene were consolidated nto a single, composite, forest class. Using the respective sets of forest class ıtatistics, each of the ten Landsat scenes was classified with the VICAR BAYES ›rogram to produce a classification map and a confidence map.

Assessment

An automated comparison of the Landsat-derived confidence maps with the ground reference data set (GRDS) was performed to assess classification results. The objective of the assessment was to determine the confidence map threshold value which resulted in the highest overall agreement between the Landsat-derived forest mask and the GRDS. A secondary benefit of the assessment was an evaluation

of the sensitivity of the forest mask to changes in the threshold value.

To facilitate the classification assessment, the confidence maps were registered to a UTM projection by JPL. Registration allowed specific locations to be identified on both the USGS topographic maps and the corrected confidence maps. The corrected data sets delivered by JPL corresponded to the eight major USGS 1:250,000 maps for Pennsylvania listed in Table VI.1 and are referred to as Quads 1 through 8. The results for UTM 17 and UTM 18 were compiled separately and later combined to produce the results for the statewide assessment.

Table VI.1. JPL Geometrically Corrected Data Sets

Quad	Map Reference	# Lines	# Samples
Quad 1	Cleveland	2000	1500
Quad 2	Canton	3000	1600
Quad 3	Warren	2100	3000
Quad 4	Pittsburgh	3100	3100
Quad 5	Williamsport	2100	3000
Quad 6	Harrisburg	3100	3100
Quad 7	Scranton	2000	3000
Quad 8	Newark	3000	3100

The GRDS consisted of the photointerpreted land cover at a series of random points located throughout the state. The random points were located by first systematically selecting a ten percent sample (86 maps) of the USGS 7.5 minute maps for Pennsylvania. On the basis of standard statistical formulas (Cochran, 1958), the need for 347 sample points, or four points per map, was determined to estimate the amount of forest cover within ± 5 percent with 95 percent confidence assuming 65 percent forest cover for the state. For each of the sampled maps, transparent plots scaled to overlay the 7.5 minute topographic maps were generated and four points were randomly located and transferred to the USGS maps.

The land cover of each sample point was categorized using either 1979 and 1981 Optical Bar Camera (OBC) color infrared (CIR) photography at a nominal scale

of 1:60,000 or 1977 and 1973 black and white aerial photography at a scale of 1:80,000. The photointerpretation identified nine cover types as follows: one forest cover type which included classes such as hardwood, brush, and conifer, and eight nonforest cover types including soil, urban, residential, agriculture, water, cloud, disturbed, and highway, which could be combined to form the nonforest class. At each sample point on the map, the land cover of an approximate single pixel area and a 3-by-3 pixel area was interpreted. For each 9-pixel ground area neighborhood, the number of pixels in each class was tallied. The single point and neighborhood ground reference results are summarized in Table VI.2 and VI.3, respectively.

Table VI.2. Summary of Single Point Ground Reference Interpretations

	# Maps	Forest	Nonforest				
		F	U	R	A	W	D
Quad 1	5	6	-	2	4	-	-
Quad 2	5	11	-	-	4	-	-
Quad 3	13	47	-	-	4	1	-
Quad 4	19	50	2	4	15	1	2
UTM 17	42	114	2	6	27	2	2
Quad 5	13	31	-	1	18	1	1
Quad 6	19	34	-	5	34	-	1
Quad 7	6	17	-	-	7	-	-
Quad 8	6	14	-	-	8	1	1
UTM 18	44	96	0	6	67	2	3
Satewide	86	210	2	12	94	4	5
					117		

Nonforest classes:
U = Urban
A = Agriculture
D = Disturbed
R = Residential
W = Water

Table VI.3. Summary of 3x3 Ground Reference Interpretations

	# Maps	Forest	Nonforest
Quad 1	5	2	0
Quad 2	5	5	3
Quad 3	13	38	1
Quad 4	19	36	11
UTM 17	42	81	15
Quad 5	13	26	13
Quad 6	19	29	28
Quad 7	6	10	1
Quad 8	6	1	6
UTM 18	44	66	48
Statewide	86	147	63

In the single point interpretations, 20 points could not be described due to their location at or near the borders of the map. Of the 327 interpreted points, 210 were identified as forest and 117 were identified as nonforest. This represents a 64/36 percent forest/nonforest distribution. Ninety-four of the nonforest points were identified as agriculture and the remaining 23 were identified as urban, residential, or disturbed.

On the basis of the neighborhood interpretations each pixel was further categorized as border or nonborder, and only nonborder pixels were analyzed. The process required that all nine pixels belong to the same general class in the ground reference data (i.e., either forest or nonforest). For the nonforest designation the procedure required that the nine pixels belong to any of the eight nonforest cover types; a mixture of nonforest types was allowed. The neighborhood photointerpretation procedure resulted in the elimination of 117 border points. Of the remaining 210 points, 147 were identified as forest and 63 were identified as nonforest. This represents a 70/30 percent forest/nonforest distribution.

The procedure for generating the forest mask required that the analyst specify a confidence value which defines the forest/nonforest threshold. At this point, the confidence map was registered to the map so that direct comparisons between the confidence map and the GRDS were possible. An automated procedure was used to compare each of the possible forest masks, corresponding to the 256 confidence map threshold values, against the ground reference data set. This process insured that the optimum threshold value and consequently the most accurate forest mask was produced.

The threshold value selection and forest mask assessment were conducted using both single point and neighborhood comparisons. The criterion used for evaluation was the percent overall agreement between the GRDS and the forest mask. For this calculation the individual nonforest cover types were consolidated into a single nonforest class. The detailed information on nonforest cover types was used only to determine the cover types involved when forest and nonforest were confused in the classification. In the neighborhood comparison, the corresponding 3-by-3 pixel Landsat neighborhood was classed according to whether the majority of the pixels were forest or nonforest.

RESULTS

The results of the threshold selection process are summarized in Table VI.4.

Table VI.4. Single Point and Neighborhood Comparison Results. The maximum percent overall agreement ("overall") confidence map threshold value ("threshold"), and associated percent forest ("F") and nonforest ("NF") agreement for the statewide assessment are listed.

Data Set	Single Point Comparison				Neighborhood Comparison			
	Overall	Threshold	F	NF	Overall	Threshold	F	NF
Statewide	82	120	85	76	90	120	93	85

When evaluated over the entire state, the optimum threshold value was 120 for both the single point and neighborhood comparisons. As expected, the overall agreement for the neighborhood comparison (90 percent) was higher than for the single point comparison (82 percent) simply because of the problems typically associated with multispectral classifications of border areas and the photointerpretation of boundary points. Higher agreement figures could probably be achieved using more rigorous classification procedures; however, the potential costs were considered to outweigh the benefits.

In the process of selecting the optimum confidence map threshold value, the information necessary to evaluate the sensitivity of the forest mask to changes in the threshold value was obtained. Figure VI.1 is a graph of the forest, nonforest, and overall percent agreement versus threshold value for the neighborhood comparisons. The trend for the overall agreement indicates that the maximum agreement is obtained over a narrow range of threshold values. This trend emphasizes the need for judicious selection of the threshold and the importance of using the reference data to guide the selection process.

SUMMARY

The use of the Bayesian classification confidence map is an effective tool for conducting single class classifications. For classifications with higher accuracy requirements, training techniques involving a more detailed breakdown of land cover classes and more thorough ground comparisons are recommended. A simpler procedure yielding comparable accuracies may be possible. For example, it may be feasible to use MSS_7/MSS_5 ratio values in much the same fashion as the confidence map and associated threshold values to obtain the forest/nonforest mask.

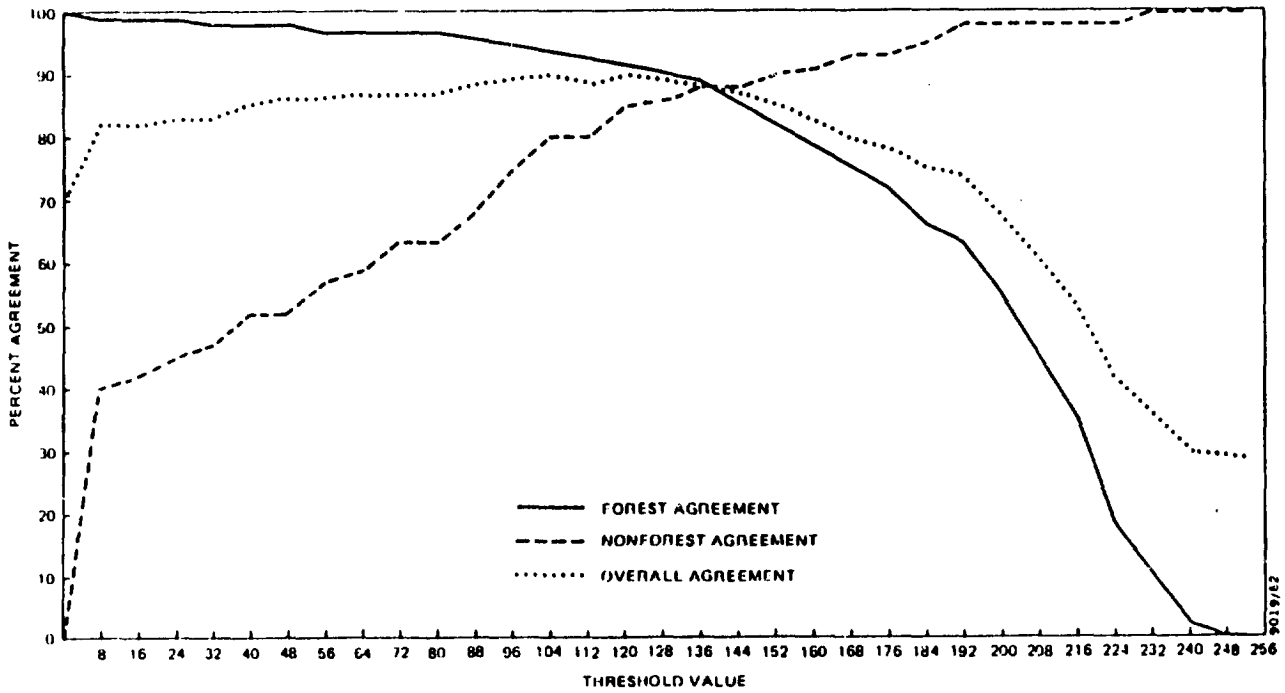


Figure VI.1. Percent Agreement vs. Threshold Value for Statewide Data Set, Neighborhood Comparison.

REFERENCES

Cochran, H.G. 1953. Sampling Techniques, 2nd Ed.. John Wiley and Sons, NY, pg. 75.

ESL. 1981. IDIMS Functional Guide, Vols. I and II. Technical Manual ESL-TM 705, Sunnyvale, CA.

General Electric Co. 1975. Image 100 Users Manual. Ground Systems Department, Space Division, Daytona Beach, Florida.

Moik, H. 1979. SMIPS/VICAR: Application Program Description. Goddard Space Flight Center, Greenbelt, MD.

APPENDIX VII

Data Management Front-End System

Note: Sections of this appendix were extracted from:

Turner, Brian J. 1981. Development of a Data Base Management Front-End for Use with a Landsat Based Information System. Interim Report. Contract No. NAS5-26468, Pennsylvania State University, University Park, PA.

OBJECTIVE

The Pennsylvania State University, Office for Remote Sensing of Earth Resources (ORSER) is an interdisciplinary organization with expertise in forestry, soils, engineering and remote sensing. Because of their staff's familiarity with the gypsy moth in Pennsylvania, and their remote sensing capabilities, ORSER was requested to develop a data management front-end system that would permit access to the data base which incorporated Landsat and ancillary data covering the entire state of Pennsylvania. This front-end system would be specifically designed to facilitate annual defoliation assessments by interfacing image analysis software with components of the data base required for the assessments. Specifically, the following capabilities were required:

1. Access to and storage of information within the Landsat-derived geographic data base;
2. Facilitate registration of new Landsat and ancillary data to the data base;
3. Subset the data base into user defined geographic areas;
4. Assist the analyst in performing defoliation assessments via a user friendly executive that produces and submits user-defined image analysis programs; and
5. Tabulation of defoliation assessment results.

FRONT-END SYSTEM DEVELOPMENT

A user friendly system has been set up using the INTERACT Executive File available at the University Computational Center at University Park, PA. This file allows a non-programmer to request a job for extracting a specified section of the data base and then allows the analyst to process that section using the ORSER software (Turner et al, 1978). The user conversationally requests counties, forest districts, Pest Locator Grid units or quad sheets, then gives the name or code of

the requested area. The EXEC program locates the MSS data and the boundary information and sets up a program that will write the MSS data within the boundary to disk or tape. This executive feature makes the data base and front-end system appear simplistic, when, in fact, the workings of this interface are extremely complex.

Storage and Retrieval

The most critical and extensive procedures developed under this contract were the archival and retrieval techniques. The Landsat mosaic and forest mask data are stored in the ORSER Data Base Format. This is a band-interleaved-by-line format in which all of the pixels for one band of a scan line are stored as one logical record on a tape. Scan lines are then organized in ascending order and grouped into tape files containing a specified number of lines. Header information on the files is stored so that selected portions of the mosaic or mask can be accessed without reading the entire tape. Along with the Landsat cellular data base layers, there are data layers that consist of sets of UTM coordinates that describe county and forest district boundaries As part of the front-end system, there is an index that relates each boundary to its corresponding file on the Landsat data tape. Other boundaries can easily be added to the data base as long as the coordinates are in the UTM projection. Landsat data that are registered to the original mosaic must first be converted to the ORSER Data Base Format before they can be stored in the data base or accessed by the front-end system.

Registration to Data Base

Registration of additional Landsat data to the data base may be done using the data management front-end system. The software necessary to register and mosaic new Landsat scenes to the data base was developed at the Jet Propulsion Laboratory. These programs were incorporated into the VICAR (Video Image Communications and Retrieval) image processing language (Moik, 1979), which may be assessed by the front-end system. All of the VICAR image processing functions are available

to the user, however, the primary reason behind implementing the VICAR software was to drive the registration and mosaicking functions. In order to use any of the VICAR programs, the user must be familiar with that language. The front end sets up the appropriate job cards so that the job can be submitted to the computer. The user must type in the image processing control statements.

Image to image registration and mosaicking require not only the selection of identifiable points within corresponding Landsat images, but also the selection of tie points to adjacent scenes. The procedures, then, may require considerable analyst interaction and they require the knowledge of a relatively user-hostile image processing language, VICAR. Hence although these procedures may be accessed by the front end, only experienced image processing analysts should attempt to add new layers of Landsat data to the data base.

Subsetting the Data Base

The Landsat-derived geographic data base can be subset in the ORSER Data Base Format using a specialized program SUBDB. The output from this program is in raw data format that can be used for subsequent analysis by any of the ORSER image analysis programs.

When a user requests a specific county or district boundary the SUBDB program automatically reads the file that contains the UTM coordinates for the specified area. The program then converts these coordinates into starting and stopping points within each scan line of the Landsat mosaic. The program also computes the maximum and minimum line and column numbers that will be needed. The SUBDB determines which file in the data base to start with based on the minimum line number. The program starts with this file directly and processes sequentially from there. The data are then reformatted into the raw data format while replacing all pixels that lie outside the specified area with null pixels. The raw data set is

written onto an output tape and is ready for subsequent processing.

The Defoliation Assessment

Defoliation assessments for any county within Pennsylvania can be generated using the Landsat-derived geographic data base and management front-end system. The front-end system, in addition to containing a series of prompts for the user, also contains a "set-up" index and a catalog of ORSER image analysis job controls. These features work together so that when a user requests any analysis program, the control cards are automatically organized and submitted to the main frame computer with only a minimum of prompts for the user. For example, to create the "defoliated forest image" which is needed to apply the Ratio Vegetation Index, the user would go through the following sequence of steps:

1. Request current Landsat data within county under investigation. The front-end automatically calls program SUBDB to retrieve that county from the data base and puts the Landsat data into raw image format.

2. Request forest/nonforest classification within county under investigation. The front-end again calls upon the program SUBDB to retrieve that county from the statewide forest resource map registered to the data base. The forest/non-forest classification is a character map in compressed format.

3. Request the program MASK. The front-end system sets up the control card listing for MASK, a program that will mask out all the nonforest pixels within the Landsat data set acquired during Step 1 using the forest/nonforest mask acquired during Step 2.

The MASK program requires two input data sets. The first must be in the ORSER raw data format. The second must be in the ORSER compressed map format. Both of these formats are described in the ORSER User's Guide. The program reads the raw data (Landsat) and the character map (forest/nonforest

classification) and sets the value in all channels of the raw data to zero for any pixel having a blank as its character value in the character map. It then writes this data set out in the ORSER raw data format which can then be read by any of the ORSER analysis programs that read raw data, such as a Ratio Vegetation Index program.

4. The user continues the defoliation assessment by requesting the RATIO program. This program calculates the MSS_7/MSS_5 ratio for each pixel within the image to facilitate delineation of different forest defoliation classes.

5. The results of the RATIO program can be displayed on a line printer, VERSATEC plotter, or tabulated. Programs have been written to accommodate the analysts request for any of these display products.

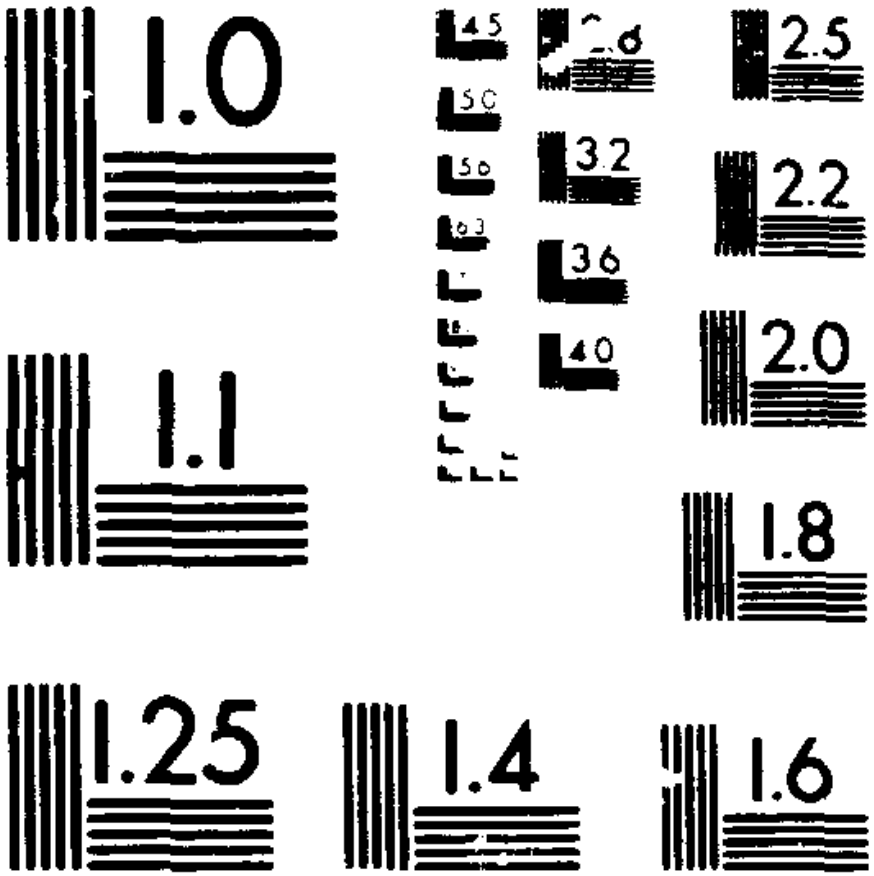
Steps 1-4 may be done "automatically" if the user wants to produce a standardized defoliation assessment. A default option has been installed in the front-end such that the user only has to specify the area of interest. A job is then submitted which extracts the area of interest, applies the forest/nonforest mask and classifies that image using the 7/5 ratios. A second program must be submitted to produce desired output products.

Tabulation of Defoliation Assessment Results

As with the actual assessment procedure, a program can be requested using the front-end system that will tabulate the number of pixels in each forest category and print these values for the user.

CONCLUSIONS

A data management front-end system has been developed and implemented on the Penn State University computer. The front-end allows users to interface with the Landsat-based information system in a user-friendly environment. Software has



been developed to adapt existing ORSER and VICAR programs to the peculiar needs of the Landsat mosaic data base as supplied by JPL. Archival and retrieval techniques have been developed to efficiently handle this dat- base and make it compatible with the requirements of the Pennsylvania Bureau of Forestry.

REFERENCES

Moik, H.G. 1979. SMIPS/VICAR: Application Program Description. Goddard Space Flight Center, Greenbelt, MD.

Turner, B.J., G.M. Baumer, and W.L. Myers. 1982. The ORSER Remote Sensing Analysis System: A Users Manual. Research Publication 109/OR, Office for Remote Sensing of Earth Resources, Pennsylvania State University, University Park, PA. 265 pgs.

C-2

APPENDIX VIII

Data Reduction Techniques

Note: Sections of this appendix were extracted from:
Russo. S.A. and M.L. Stauffer. 1982. The Impact of Data Reduction on Forest Classification Accuracy. Computer Sciences Corporation Contract Report CSC/TM-32/6205, Goddard Space Flight Center, Greenbelt, MD.

STUDY OBJECTIVE

Processing of the large volume of Landsat multispectral scanner data for the Pennsylvania statewide forest classification map necessitated that several factors be considered to insure that an accurate product be generated cost effectively and efficiently. There existed trade offs among processing requirements, analyst involvement, and classification performance that needed to be addressed within the context of GSFC objectives. Efficient processing was important simply because of the volume of data that needed to be analyzed (10 full Landsat scenes). The accuracy of the forest classification was critical because defoliation assessments were dependent upon the initial identification of forest cover types.

Several data reduction techniques were examined by JRP project personnel to determine if the required accuracy of the forest classification map could be maintained while reducing computer processing time. These techniques fell into two general categories:

1. reduction of spectral channels,
2. subsampling the data

DESCRIPTION OF STUDY SITE AND DATA

The area selected for this study is located in central Pennsylvania northwest of Harrisburg, Pennsylvania. The area corresponds to the USGS 1:24,000 Wertzville topographic quadrangle and lies within the Ridge and Valley Province. The area contains cover types typical of the state including extensive oak-hickory forest, agricultural lands, small woodlots, and rural communities.

Cloud-free Landsat data collected July 19, 1976 (Scene No. 2544-15005) was selected for use in this study. The data was chosen because of its availability and the absence of major forest disturbances such as gypsy moth defoliation. Several supporting data sets were also available for this study site. These included USGS

topographic maps compiled in 1952 and updated in 1973 and color infrared aerial photographs collected August 13, 1980. The Landsat data were registered to the 1:24,000 Wertzville Topographic Map.

PROCEDURE

This study evaluated two data reduction procedures: reducing the number of spectral channels processed; and reducing the number of pixels processed. Specifically, the following procedures were examined.

1. Channel Reduction - A comparison was made between using the full complement of spectral channels (MSS4, 0.5-0.6 μm; MSS5, 0.6 - 0.7 μm; MSS6, 0.7 - 0.8 μm; MSS7, 0.8 - 1.1 μm) and using two spectral channels (MSS5 and MSS7), to identify forest cover types in the Wertzville area. Numerous studies have shown that MSS5 and MSS7 are the most important channels for vegetation identification, therefore, these channels were considered the appropriate choice for data reduction. The reduction was accomplished by only processing the selected bands and did not require any special preprocessing.

2. Pixel Reduction - A comparison was made between full resolution data (100% pixels) and reducing the number of pixels by 75% to identify forest cover types in the Wertzville area. Two techniques were used to achieve the 75% reduction:

   a. selection of every second line and pixel

   b. computation of the average value for successive 2 x 2 pixel windows.

Subsampling the data on the 2 x 2 grid required some preprocessing.

The channel reduction and pixel reduction techniques were combined such that six data sets were generated (see Table VIII-1). Using a supervised Bayesian classification procedure on each of the data sets, forest/nonforest resource maps were generated.

Following the Bayesian classification, each product was evaluated to determine the classification performance. One hundred ninety-nine points were randomly

selected from the generated maps. The location of these points on the ground was determined by overlaying the Landsat-derived forest/nonforest classification onto the Wertzville topographic quadrangle map. Using the quadrangle location, each point was located on the aerial photography using a Bausch and Loom Zoom Transfer Scope. The cover type was noted as either forest or nonforest. In addition, the land cover class of a 3 x 3 pixel neighborhood was noted for each of the 199 points. The ground cover type for each point and neighborhood was compared to the Landsat-derived classifications to determine how well each data set listed in Table VIII-1 represented actual ground conditions. Neighborhood comparisons were considered necessary to minimize the impact of registration errors on the accuracy assessment.

Table VIII-1. Listing and abbreviations of data sets used to examine the impact of data reduction techniques on forest/nonforest classification.

		Number of Channels Used	
		4	2
Resolution	Full Resolution	FR-4	FR-2
	2x2 Subsample	SS-4	SS-2
	2x2 Average	AV-4	AV-2

RESULTS

Performance evaluation results for each of the six data sets are given in Tables VIII-2 and VIII-3 (single point comparisons and neighborhood comparisons, respectively). The results of this study suggest that feasible, cost-effective alternatives to the use of a 4-channel full resolution data set for forest/nonforest classification exist.

The use of MSS_5 and MSS_7 with the full resolution (i.e., every pixel) Landsat data allowed a 50 percent reduction in the volume of data to be processed with little change in classification performance relative to the 4-hannel forest/nonforest classification. Data reduction

by pixel subsampling or averaging also reduced data volume with only a moderate impact on classification of forest and nonforest. Therefore, the use of any of these techniques could be considered appropriate, based on the requirements of the activity underway. For example, if the primary concern is the delineation of large contiguous areas of forest, a reduction of pixel resolution might be acceptable. On the other hand, if smaller woodlots need to be identified, the analyst might choose to maintain the full resolution data set with 2 channels of data. Based on the results of this study and the study described in Appendix I, the two-channel full resolution (FR-2) Bayesian classification procedure was selected to generate the statewide forest/nonforest classification map of Pennyslvania.

Table VIII-2. Performance evaluation for Landsat-derived forest/nonforest classifications. Percentages based on single pixel comparisons.

	Percent Agreement Between Classification and Ground Reference Data		
Data Set Used	Overall	Forest	Nonforest
FR-4	89	89	90
FR-2	88	88	88
SS-4	83	81	85
SS-2	83	81	85
AV-4	86	86	85
AV-2	84	84	85

Table VIII-3. Performance evaluation for Landsat-derived forest/nonforest classifications. Percentages based on neighborhood comparisons.

Data Set Used	Percent Agreement Between Classification and Ground Reference Data		
	Overall	Forest	Nonforest
FR-4	100	100	100
FR-2	100	100	100
SS-4	92	91	96
SS-2	91	89	96
AV-4	93	92	95
AV-2	93	92	95

END

DATE

FILMED