THE ORSER LANDSAT DATA BASE OF PENNSYLVANIA

B. J. Turner
Office for Remote Sensing of Earth Resources
The Pennsylvania State University
University Park, Pennsylvania 16802, U.S.A.

and

D. L. Williams
NASA/Goddard Space Flight Center
Earth Resources Branch/923
Greenbelt, Maryland 20771, U.S.A.

1. INTRODUCTION

A mosaicked Landsat data base for Pennsylvania has recently been installed at the Computation Center of The Pennsylvania State University. Initially constructed by Penn State's Office for Remote Sensing of Earth Resources (ORSER) for the purpose of assisting in state-wide mapping of gypsy moth defoliation, the data base will be available to a variety of potential users. It will provide geometrically correct Landsat data accessible by political, jurisdictional, or arbitrary boundaries.

2. FOREST DEFOLIATION ASSESSMENT PROJECT

Each year, state and federal agencies spend millions of dollars developing programs to prevent the spread of the gypsy moth caterpillar (Lymantria dispar), which has defoliated millions of hectares of hardwood forest. Since the caterpillar was introduced in the United States in 1869, (in an effort to produce a new variety of silkworm) the gypsy moth has become established throughout most of the northeast, and south to West Virginia and Maryland. Gypsy moth populations have periodically increased to epidemic proportions. Currently one of the largest recorded outbreaks seriously infested nearly 4 million hectares (10 million acres) during the 1981 summer feeding cycle, and projections for 1982 are even higher.

Integrated pest management programs, developed to prevent the insect's spread, depend largely on accurate, timely, and efficient methods of detecting and mapping incipient forest canopy damage. Ground surveys, aerial sketchmapping, and photointerpretation have been used to detect the damage, but the expense and subjectivity of these methods have led to a search for more efficient and accurate techniques. In view of the wide areas of damage, it has also become desirable to standardize the methods used among the various state agencies.

Researchers began to look for a new survey technique which could provide timely, accurate, and standardized assessments at a reasonable cost. By the mid-1970's, after Landsat multispectral scanner (MSS) data became widely available, research began to indicate that Landsat data had potential for monitoring

widespread forest disturbances such as infestations of the gypsy moth and other insect species. The standardized spectral, spatial, and temporal coverage of Landsat data sets, and the synoptic coverage provided, seemed to be ideally suited as a survey medium. ORSER and NASA (National Aeronautics and Space Administration) at Goddard Space Flight Center (GSFC) in Greenbelt, Maryland, were among the early participants in such research.

In order to demonstrate the usefulness of satellite remotely sensed data for monitoring insect defoliation of hardwood forests in Pennsylvania, a joint research project was initiated between NASA/GSFC and the Pennsylvania Bureau of Forestry, Division of Forest Pest Management (DFPM). A framework for automated assessment of defoliation using Landsat MSS data was provided by the earlier GSFC work (Williams and Stauffer, 1978; Williams et al., 1979; Nelson, 1981).

The procedure for defoliation assessment requires four steps:

Creation of a healthy forest classification mask: Prior to insect infestation, data for a cloud-free summer Landsat image over the study site are obtained. This image is classified into two categories, using digital analysis techniques: forest and non-forest. Pixels classified as forest are assigned the value of 1 and all other pixels in the scene are assigned the value of 0. The resultant image is called the "1/0 forest/non-forest mask."

Application of the forest/non-forest mask to the image showing defoliation: An image of the study site obtained at the peak of defoliation or shortly thereafter is digitally registered to the 1/0 forest/non-forest mask. This registered image is then multiplied by the forest/non-forest mask to produce a "defoliated forest image," in which areas in the scene which show forest have been isolated from other cover types.

Application of the ratio vegetation index to assess forest disturbance: The ratio vegetation index (RVI) is the ratio of the infrared to the red spectral response (MSS band 7/band 5) for each pixel within the image. The RVI is applied to the defoliation image, creating a new image, the "assessment image," in which low ratio values indicate heavy defoliation and high values indicate healthy forest. Because of previous application of the mask, zeros indicate non-forest.

Separation of defoliation levels: Aerial surveys or other ground reference data are compared to the assessment image to determine the numerical levels separating healthy, moderately defoliated, and heavily defoliated forests. It is important to note that the key requirement in this procedure is the ability to register several different images to a common reference base. Such a common reference base has been created for the state of Pennsylvania by the Office for Remote Sensing of Earth Resources, at The Pennsylvania State University.

3. CREATION OF THE PENNSYLVANIA DATA BASE

The Pennsylvania legislature has mandated that the state's Division of Forest Pest Management conduct annual assessments of insect-related damage to forests throughout the state. Yearly statistics must be compiled to study trends in insect population dynamics, as well as for planning management alternatives. Although a wealth of information has been acquired over the years, it is of

limited use because it exists in various hard copy formats (e.g., maps, aerial photographs) which do not lend themselves to computer storage and retrieval, and because the non-standardized format of these products, and the subjectivity of analysis procedures used to generate them, makes meaningful trend analysis almost impossible. Landsat, on the other hand, offers a standardized MSS data source which has been collected for over 10 years. The information is in digital format, which can be processed quantitatively and repeatedly, and both the original data and the derived results can be readily stored, retrieved, and compared by computer. However, the size of the state, and the corresponding volume of data required for accurate defoliation assessment presented a unique challenge. Not only was it necessary to store and retrieve the data, but extensive digital image processing was required, as well as a means to compare and assess the output products from such processing.

In the course of the joint project between NASA/GSFC and DFPM, various methods were considered for handling the large volume of Landsat data required to conduct defoliation assessments on an annual basis. It was decided to develop a Landsat-derived, multilayered, geographic data base which could be interfaced with image analysis software. This data base had to contain a minimum of three layers:

1) a Landsat digital mosaic of Pennsylvania exhibiting no defoliation and registered to the Universal Transverse Mercator (UTM) map projection, rotated to north, and resampled to 57 meter square cells (the cell size of future Landsat data);

2) a forest resources map (forest/non-forest mask) derived from the Landsat data in the first layer; and

3) digitized Forest Pest Mangement District boundaries and county boundaries registered to the Landsat mosaic.

The capability to add additional data layers, such as the most recent Landsat data depicting defoliation, was also required.

Fortunately, the ability to retrieve, digitally process, and store Landsat MSS data sets was already available at the Office for Remote Sensing of Earth Resources (ORSER), located at The Pennsylvania State University. Thus, it was decided to develop and house the Pennsylvania Landsat data base on the IBM 370/3081 computer at the University's Computation Center. ORSER agreed to develop or acquire, upgrade, and implement all software necessary to create and manipulate the data base.

4. CREATION OF THE MOSAIC

The Pennsylvania mosaic of Landsat data acquired prior to defoliation would provide the foundation for all subsequent procedures in operating a defoliation assessment system. Because of their demonstrated capabilities in generating Landsat mosaics of California and Arizona (Zobrist and Bryant, 1979), NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California was asked to generate the initial mosaic. The mosaicking procedures required the use of the VICAR/IBIS software system developed at JPL, as well as additional mosaicking software which has been incorporated into the VICAR system.

Mosaicking begins with the selection of several ground control points within each image frame. Seam control points on adjacent frames are then selected by automatic correlation analysis. These are adjusted by a distortion model for each frame, based on the ground control points. Seam points are then reconciled by averaging their mapped locations in adjacent frames. Finally, the processed Landsat data are "cut" at the mapped seam boundary to produce the mosaic piece and the pieces are "sewn" together (Zobrist et al., unpublished manuscript). The control points selected for one Landsat spectral band can be applied to the other three bands and the same geometric correction performed.

5. REFORMATTING THE DATA BASE

As supplied to ORSER, the magnetic tapes containing the mosaic were in band-sequential VICAR format. That is, each file contained data for a quadrangle of one degree of latitude by two degrees of longitude. Eight such quadrangles were necessary to cover the whole state. Unfortunately, this format was not suitable for the Penn State computing environment, where it is much less expensive to locate the beginning of a file on a tape than it is to read individual records. Thus, it was more efficient to store the data in long-line records, with relatively few records per file, than in large quadrangles of data. It was also more convenient to store the data in a form similar to the ORSER raw data (RD) format, a modified band-interleaved-by-line format, than in the band-sequential format.

The ORSER data base (DB) format, like the RD format, is also a band-interleaved-by-line format. Here all the pixels for one band of a scan line are stored as one logical record and the scan lines are organized in ascending order, just as in the RD format. Scan lines are grouped into files containing 500 lines. Thus, 12 files, containing 500 lines each, are used for each half of the data set. Header information on the files is stored within the program so that only the files containing data within the area of interest need be read. This reduces the computer time required to access an area that may be several thousand scan lines down the data base.

Three programs were needed to reformat the half-state data from the 16 VICAR files into the ORSER DB format: SEW reads up to four VICAR-format files of adjacent areas and concatenates them to form one VICAR file. This is done for each of the four bands. INT reads VICAR files and generates band-interleaved-by-line files. It is run on bands 4 and 5 together and then on bands 6 and 7 together. DBGN then reads these two files, interleaves them, and breaks them down into 12 files of 500 scan lines each. To check the results, band 7 of the complete data set was displayed on a Versatec electrostatic printer (Figs. 1 and 2). The three reformatting programs can also be used to add information to the data base, such as extra bands of Landsat data or data for adjacent geographic areas.

In addition to the grid-cell formatted Landsat data, the data base consists of sets of coordinates, stored on separate tape files, describing irregular areas, such as the county and forest district boundaries currently in the system. An index in the front-end system relates each county and forest district name to its corresponding file on the tape. Additional boundaries (watersheds, for instance) can be added to the system, as long as their coordinates are in the UTM projection.

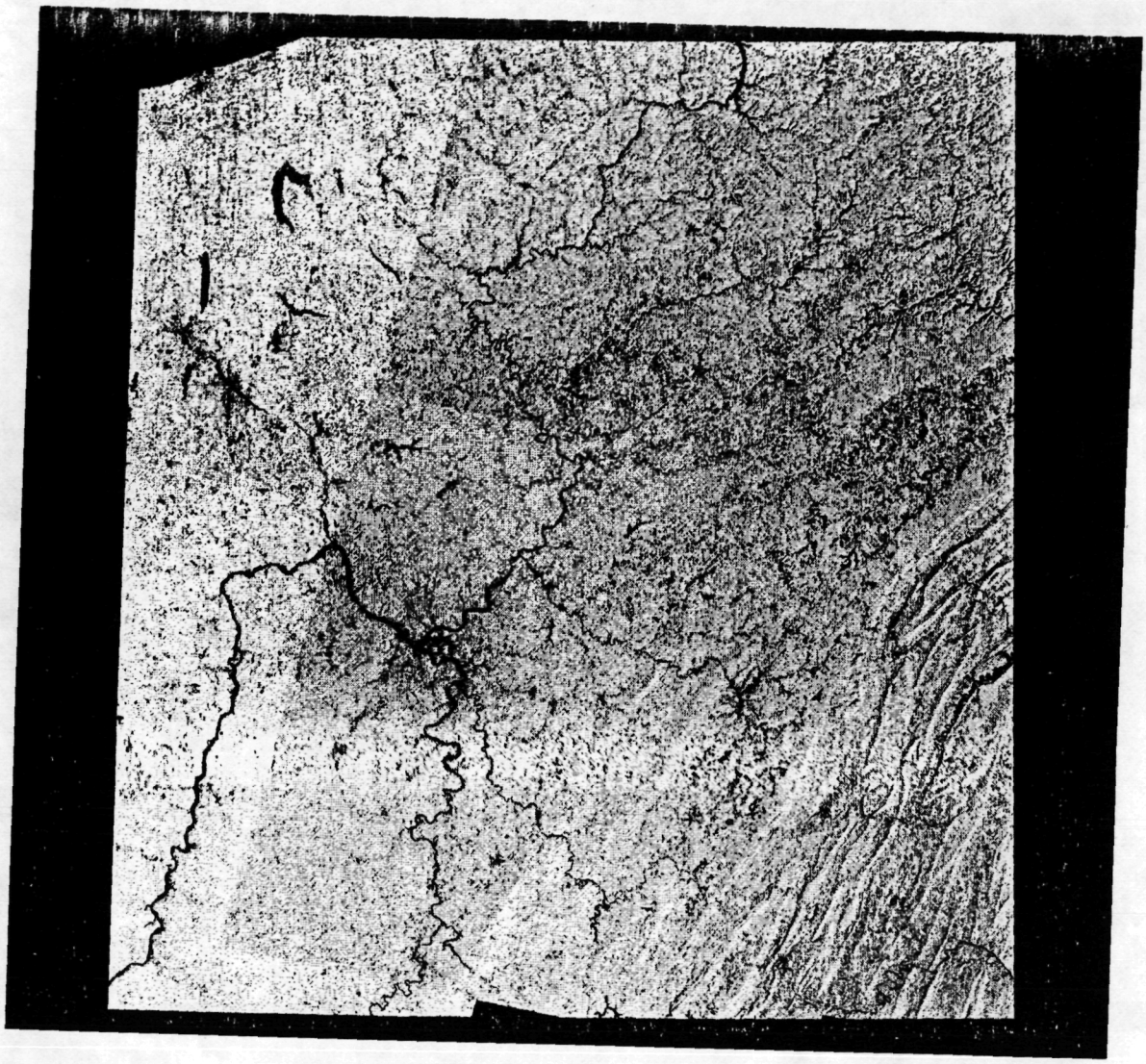

Figure 1. Band 7 electrostatic printer display of the western half of the Pennsylvania mosaic (UTM 17).

Figure 2. Band 7 electrostatic printer display of the eastern half of the Pennsylvania mosaic (UTM 18).

6. SUBSETTING FROM THE DATA BASE

The SUBDB program was written to subset an irregularly bounded area from a data set in the ORSER DB format and output it in the RD format for subsequent analysis by any of the ORSER programs. The program may read a file containing the coordinates describing a predefined polygon, or coordinates may be entered directly in UTM meters or as line-and-element numbers. These coordinates are converted to start and stop points within each scanline. The program then determines which file to start with, processing sequentially from that file. The data are reformatted and all pixels lying outside the defined polygon are replaced with zeros (null pixels). The new data set, now in the ORSER RD format, is written to tape and can be processed by any ORSER program that handles raw data. An example of a county data set extracted in this manner from the data base and displayed on an electrostatic plotter using the NMAP program is shown in Fig. 3. In order to extract the UTM coordinates of counties and forest districts supplied on tape from GSFC, the PIOS program was written. It converts UTM coordinates to line-and-element numbers, producing input to the SUBDB program.

7. DEVELOPMENT OF THE FRONT-END

In most cases, using the data base involves moving large data sets between storage media and the computer. Because such transfers require manipulating job control language (JCL)--a process unfamiliar to many potential users--a user-friendly front-end processor was developed to set up jobs. At the University, where the large IBM 370/3081 operates in batch mode, the best way to develop such a front-end was to use the EXECUTE facility of the INTERACT (also known as MENTEXT or WYLBUR) system introduced at the University two years ago.

The INTERACT system is designed for program development, remote job processing, and document composition. Responding to commands from local or remote terminals, it interprets these, performs the requested processing, prompts for further information, and provides error messages where appropriate (Cullinane Corp., 1980). Using the EXECUTE facility of INTERACT, which provides a complete programming language, the user can construct an EXECUTE (EXEC) file, containing an executable series of instructions. Such files are commonly used by non-programming personnel to perform operating system functions, and are particularly useful for handling frequently-used functions involving data manipulation. The INTERACT front-end for the ORSER system has proven very useful for users at the University (Turner et al., 1982).

To operate the Pennsylvania data base, an EXEC file was set up as a major subset of the existing ORSER EXEC file. After entering the ORSER EXEC file, the data base user responds to the first prompt by typing in "DATABASE." A series of prompts then permits the user to select the county, forest district, grid cell, quadrangle area, or irregular polygon desired; asks for the name, number, or coordinates of the specified area; and asks for the band numbers required, and whether the output is to be put on tape or disc. By typing in "HELP" to any of these prompts, the user is supplied with further explanation of the reply appropriate to that prompt. The result of the interaction described above is an active file containing the JCL and selected options needed to execute the SUBDB program. When used directly to run the job, the required data subset will be stored on the requested medium in ORSER RD format, ready to be processed by any

ELK COUNTY SUBSET

NMAP (BRIGHTNESS MAP)
OF ALL FOUR BANDS

EIGHT UNIFORM PERCENT CLASSES

Scale 1:450,000

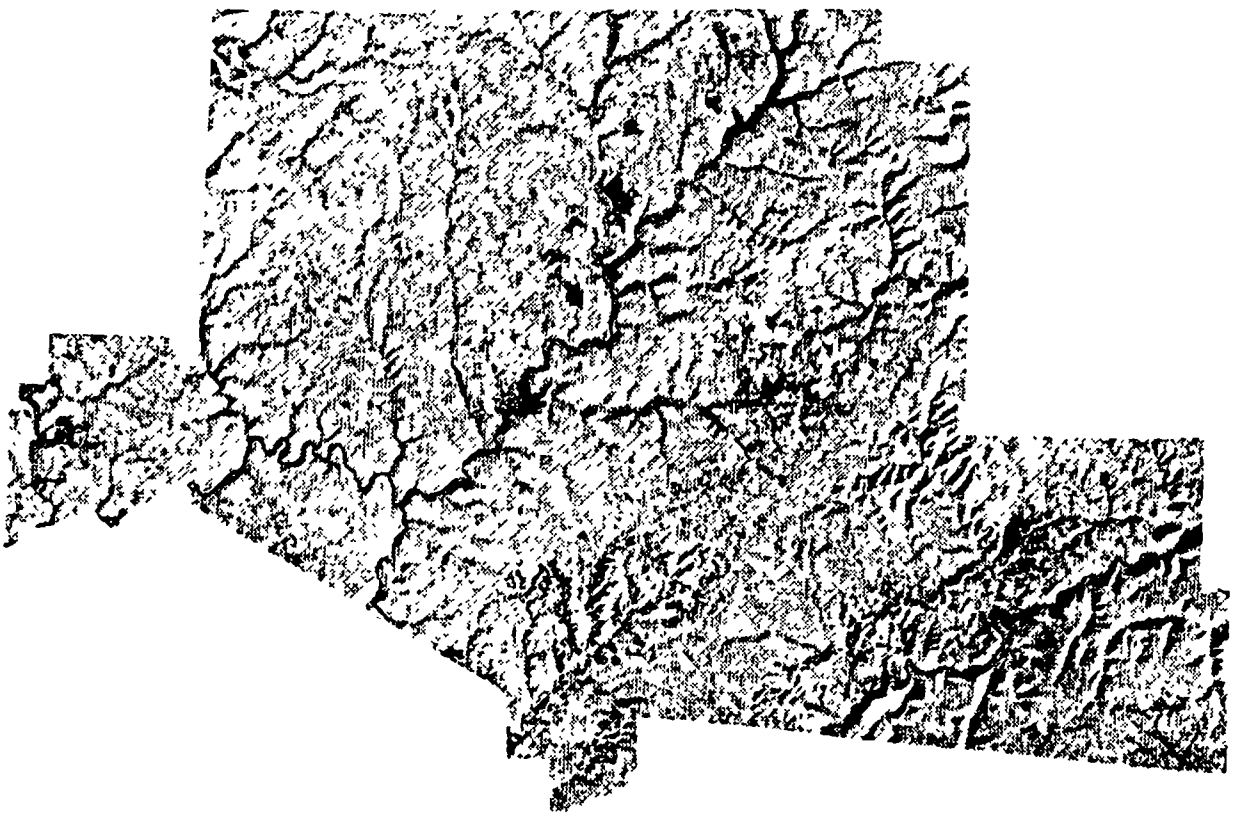

Figure 3. Electrostatic plotter print of the Elk County data set.

of the appropriate ORSER programs. (An example session with the EXEC file is given in the Appendix.)

Subset data sets are not currently cataloged within the EXEC file. At this stage, it has been sufficient to store large data sets on tapes cataloged through the ORSER tape library system (Turner et al., 1982), and to regenerate small subsets when needed. However, additional data layers, such as the recently-added binary forest/non-forest mask, may soon create a need for a cataloging system.

8. CONSTRUCTION OF ADDITIONAL DATA LAYERS

In addition to the forest/non-forest mask mentioned above, a Pennsylvania mosaic of summer 1981 data is nearing completion and will be registered to the data base. The western half of the mosaic (UTM 17) is being constructed at JPL, while the eastern half (UTM 18) is being constructed at ORSER. For this purpose, ORSER obtained the VICAR/IBIS software and additional mosaicking software modules from JPL, and implemented these at the Computation Center for access through the ORSER EXEC file.

The 1981 mosaic is constructed in a fashion similar to the original mosaic, except that each scene is registered to data base control points rather than to ground control points. After a week's training by a JPL representative, and the correction of some minor errors in the JPL procedures, a mosaic was produced which exactly overlaid the data base mosaic with the exception of one small area. This area was subsequently found to have too few control points. Reinstallation of some points with marginally acceptable correlations, and a repeat of the process, resulted in an exact fit.

9. COST ESTIMATES

The direct cost of producing a half-state (six-frame) mosaic of approximately 5250 lines and 6100 elements is approximately $8,000. This estimate includes approximately $3,000 for computer costs (at University rates) but excludes the cost of the data. Although this is a significant investment, such a mosaic has the advantage of being current and geographically registered to past data. In this form, subsequent processing of this data set is significantly reduced.

10. APPLICATIONS

The primary application of the layered mosaic is for state-wide annual assessments of defoliation of Pennsylvania forests. It is anticipated, however, that the data base will be of value to many land management and monitoring agencies throughout the state. Among the many potential applications, the following are suggested.

1. Monitoring forest resources: Much of the two-thirds of Pennsylvania covered in forest is approaching commercial maturity. Large scale changes in these forests are occurring because of harvest, mineral and fuel exploration, insect attacks, and competition from other land uses. Using the Landsat data base as the mid-date in a three-date analysis, ORSER is attempting to determine optimum change-detection procedures.

2. Soil mapping: Digitized soil maps can easily be overlaid on the data base for comparison without further rectification. The value of Landsat data for improving existing soil maps in Pennsylvania is under investigation.

3. Updating existing data bases: ORSER has developed techniques for interfacing the Landsat data base (or data derived from it) with existing geographic information systems (GIS's). The user defines a grid or polygon pattern, such as the grid-cell pattern of an existing GIS. Classified Landsat data are then extracted through this pattern and the area statistics are summarized by polygons (Irish and Myers, in preparation). Since most current land-use data bases are at the same map projection as the Landsat data base, further expensive geometric correction can be avoided.

4. Adding existing digitized information: Several types of digitized data are currently available in either raster form (e.g., digital terrain data), or in line or polygon form (e.g., roads, jurisdictional boundaries). Many of these data sets are already stored at the University Computation Center, and could easily be added as layers to the data base, if desirable.

5. Construction of small-area land cover maps: Because the significant tasks of geometric correction, and often of defining boundaries, are unnecessary when using the data base, the initial cost of these operations is spread over many projects. As a result, the cost of generating land cover maps for small geographic areas, such as watersheds and townships, is substantially reduced.

11. SUMMARY

The Office for Remote Sensing of Earth Resources at The Pennsylvania State University, working through a contract funded by NASA, has acquired a Landsat digital mosaic data base of the state of Pennsylvania in the UTM map projection. ORSER has also acquired the software and expertise to construct additional Pennsylvania mosaics and register them to the data base. In cooperation with personnel from the Jet Propulsion Laboratory, a state-wide summer 1981 mosaic has been constructed and registered to the data base to demonstrate the use of such data for assessment of gypsy moth defoliation. A user-friendly front-end system which permits storage, interrogation, retrieval, and manipulation of subsets of the data base and associated ancillary data, has also been developed. Thus, defoliation assessments in the state will be facilitated by the capability to quickly retrieve selected satellite imagery, and generate defoliation maps and associated statistics. In addition, the existing forest resource base map can be continually updated, enabling forest entomologists to prepare timely surveillance reports and pest management plans.

There are wide applications for the data base which, together with various ancillary data sets, can provide geographically consistent information from many sources suitable for a variety of purposes, both in research and applied fields. We anticipate that the data base will be a key source of land-use and resource data for the state.

REFERENCES

Cullinane Corporation. 1980. INTERACT User's Guide.

Irish, R. R., and W. L. Myers. In preparation. An Interface for Remote Sensing Digital Image Systems and Geographic Information Systems. To be presented at the National Conference on Energy Resource Management, September 1982.

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

Turner, B. J., G. M. Baumer, and W. L. Myers. 1982. The ORSER Remote Sensing Analysis System: A User's Manual. Research Publication 109/OR. Institute for Research on Land and Water Resources, The Pennsylvania State University.

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

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, Fifth Annual Symposium on Machine Processing of Remotely Sensed Data. pp. 368-375.

Zobrist, A. L., and N. A. Bryant. 1979. Map Characteristics of Landsat Mosaics. Proceedings, American Society of Photogrammetry Annual Convention. pp. 260-273.

Zobrist, A. L., N. A. Bryant, and R. G. McLeod. Unpublished manuscript. Technology for Large Digital Mosaics of Landsat Data.

```
? exec fro $men.u41000.gmb.lib#orsergo on cat clr


WELCOME TO THE ORSER SYSTEM.
OK TO CLEAR ACTIVE FILE? ok
ENTER PROGRAM NAME OR 'HELP' FOR A DETAILED LIST OF INSTRUCTIONS.
ENTER 'LISTTAPES' TO LIST WORKING TAPES (RS TAPES) ASSIGNED TO YOU.
ENTER 'POLYGON' TO EXECUTE ANY ORSER POLYGON PROGRAM.
ENTER 'EXIT' TO EXIT THIS EXEC FILE.
ENTER 'DATABASE' TO ACCESS THE PENNSYLVANIA LANDSAT DATABASE.
-->database

WELCOME TO THE PENNSYLVANIA LANDSAT DATABASE.

LANDSAT DATA CAN BE RETRIEVED BY COUNTY NAME (C), BY FOREST DISTRICT (D),
USER DEFINED POLYGON (U), OR BY PEST LOCATER GRID CELLS (P).
ENTER THE TYPE OF AREA TO BE RETRIEVED (C/D/U/P) OR TYPE 'HELP' FOR
MORE INFORMATION.
-->c

*** THE PENNSYLVANIA LANDSAT DATABASE ***

      ACCESSING AREA BY COUNTY NAME

ENTER THE COUNTY NAME. ONLY ONE COUNTY CAN BE ACCESSED AT A TIME.
ENTER 'HELP' FOR MORE INFORMATION.
-->elk

IS OUTPUT ON TAPE OR DISK?  (T/D)
-->t
ENTER LAST NAME AND FIRST INITIAL SEPARATED BY ONE BLANK.
-->baumer g
ENTER OUTPUT TAPE NAME
-->rs0114
1000 COMMANDS EXECUTED WITH NO TYPING -- CHECK FOR LOOP

* END OF COUNTY ACCESS METHOD *


ENTER JOB PARAMETER OPTION NUMBER(S) OR 'HELP' FOR A LIST
OF OPTIONS.  TO EXIT EXEC FILE, HIT RETURN.
-->

** ACTIVE FILE NOW CONTAINS STEM FOR RUNNING THE DATABASE PROGRAM **
FOR INFORMATION ON RUNNING THE PROGRAM, ENTER 'HELP', OR HIT
RETURN TO EXIT.
-->

*** END OF ORSER EXEC FILE ***
```