Computer-Aided Boundary Delineation of Agricultural Lands

Thomas D. Cheng, TGS Technology, Inc., Moffett Field, California
Gary L. Angelici, Sterling Federal Systems, Inc., Palo Alto, California
Robert E. Slye, Ames Research Center, Moffett Field, California
Matt Ma, TGS Technology, Inc., Moffett Field, California

November 1989

NASA

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

COMPUTER-AIDED BOUNDARY DELINEATION OF AGRICULTURAL LANDS

Thomas D. Cheng,* Gary L. Angelici,† Robert E. Slye, and Matt Ma‡

Ames Research Center

SUMMARY

The National Agricultural Statistics Service of the United States Department of Agriculture (USDA) presently uses labor-intensive aerial photographic interpretation techniques to divide large geographical areas into manageable-sized units for estimating domestic crop and livestock production. Prototype software, the computer-aided stratification (CAS) system, was developed to automate the procedure, and currently runs on a Sun-based image processing system. With a background display of Landsat Thematic Mapper and United States Geological Survey Digital Line Graph data, the operator uses a cursor to delineate agricultural areas, called sampling units, which are assigned to strata of land-use and land-cover types. The resultant stratified sampling units are used as input into subsequent USDA sampling procedures. As a test, three counties in Missouri were chosen for application of the CAS procedures. Subsequent analysis indicates that CAS was five times faster in creating sampling units than the manual techniques were.

INTRODUCTION

The National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) has the primary responsibility of providing statistics for domestic crop and livestock production. These statistics are derived from data collected though a variety of sampling techniques and surveys. The computer-aided stratification (CAS) system has been developed at NASA Ames Research Center (ARC) to automate the stratification and sampling-unit delineation portion of sample survey procedures that are currently performed manually. The CAS stratifies sampling units by land-use and land-cover type, using image processing hardware and software. The system provides coverage areas and boundaries of stratified sampling units which are used as inputs for the subsequent sampling procedures from which agricultural statistics are developed.

The introductory section of this report includes a description and history of the "area sampling frame" as currently used by NASS. It also describes the basic objectives of the research toward automation of boundary delineation procedures for agricultural lands, system integration, and applications.

The next section describes the techniques developed to automate the NASS area sampling frame procedures and the environment in which they operate. The computer hardware and software

---

*Research Scientist, TGS Technology, Inc., Ames Research Center, Moffett Field, California.

†Scientific Programmer, Sterling Federal Systems, Inc., Ames Research Center, Moffett Field, California.

‡Scientific Programmer, TGS Technology, Inc., Ames Research Center, Moffett Field, California.

configuration of the CAS system is presented first, followed by the CAS procedures and generic functions involved in the construction of area sampling frames.

A detailed account of the trial use of the procedures in three counties in Missouri is presented in the third section. The steps developed in the second section are presented as they will be applied by NASS in an operational mode in the computerized environment.

In the final section, the automated procedure is evaluated, current developments are described, and ideas for improving the CAS system are offered.

Background

General– NASS started using remotely sensed data, in the form of aerial photography, in the 1950s to aid in the construction of state area sampling frames. In the 1970s, NASS (then SRS, the Statistics Reporting Service) experimented with Landsat sensor digital data as an input in the development of a crop estimator (Sigman et al. 1977; Craig et al. 1978). The experience gained from this research led to large-area applications. The AgRISTARS program (AgRISTARS 1981; 1982; 1983) from 1979 through 1986 further propelled the USDA toward using the Landsat sensor data to estimate crop production over large areas in homogeneous regions. The goal of the experiments with Landsat data was to provide quick estimates of crop acreages, with significantly smaller sampling errors than those of estimates generated with procedures that used ground sample data only (Allen and Hanuschak 1988).

An area sampling frame is a construct that serves as a sampling vehicle for large-area surveys. It is compiled by dividing the area to be surveyed into contiguous parcels, each of which is easily locatable on the ground and is suitable for a sample-based survey. The area sampling frame for a state or county consists of a collection of all parcels of land in an area of interest framed by either political or natural boundaries. The AFS of NASS in Fairfax, Virginia, is responsible for the development of NASS area sampling frames.

Area frame procedure history– Iowa State University began the construction of area frames for use in agricultural surveys, in 1938. NASS investigated the use of area sampling frames on a research basis in the mid-1950s and made them operational for the United States in 1965. Modifications made during the 1960s and 1970s involved chiefly the sampling method. Improvements during the 1980s were mainly in the form of increased computerization and the use of satellite sensor imagery. This area sampling frame survey program has grown over the years and is now conducted in 48 states, with approximately 16,000 segments being visited in the field by data collection personnel for the major agricultural survey conducted in June each year (Cotter and Nealon 1987).

Area sampling frame– The current manual procedure employed by NASS to develop and edit sampling frames was designed to ensure a statistically valid sample and operational efficiency.

Some of the key steps in the procedure, from Cotter and Nealon (1987), are as follows:

1. Stratification: the technique of dividing the land in an area, for example, a county or a state, into land-use and land-cover groups (known as strata) based on photo-interpretation processes. The strata are defined mostly by the intensity of agricultural activities or cultivation. This step includes preconstruction analysis; procurement of stratification materials, for example, Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM) imagery 1:250,000-scaled false-color composite prints; and land-use/land-cover stratification into primary sampling units (PSUs). PSUs are groups of six to ten ultimate sampling units or segments bounded by permanent features of the landscape, such as roads and rivers.

2. Digitization: PSUs are electronically digitized for area measurements that are used to allocate the samples proportionally.

3. Multi-step sampling: PSUs are delineated from the strata based on land use, and a random sample of PSUs is chosen for further breakdown into segments. After the PSU has been delineated into segments, a segment is randomly chosen and is visited in the June Agricultural Survey to collect information concerning, for example, crop types and coverage areas.

4. Analysis and quality assurance: the land-use strata definitions, the number of PSUs, the size of the segments, etc. are periodically analyzed to eliminate errors of commission and omission and, thus, to improve the accuracy of the statistics produced.

NASS software development– In an attempt to improve the accuracy of crop estimates and reduce the cost of generating statistics, NASS developed the EDITOR image processing system in the late 1970s. EDITOR was written at the University of Illinois in association with NASS and ARC.

EDITOR used the sample segments from the area frame procedure in conjunction with Landsat MSS data and conventional clustering and classification algorithms to generate large-scale crop estimates (Hanuschak and Morrissey 1977). Although manipulation of Landsat MSS digital data was the primary function of EDITOR, the software contained no capability to display the data or to perform any interactive image processing functions.

In the early 1980s, NASS decided to rewrite the EDITOR software so that it could be operated on a number of different computer systems. The new software, written largely at ARC, was coded in a portable and modular way in the Pascal and C languages. The new system was called PEDITOR, for portable EDITOR, and was completed in 1985 (Angelici et al. 1986). It was tested during the California Cooperative Remote Sensing Project from 1985 to 1987. The PEDITOR system has since been transferred from the original DEC-10 system and the Motorola 68000-based workstations, the Micro Image Display Analysis System (MIDAS) (Erickson et al. 1983; Hofman et al. 1983), to Sun 2/120 and Sun 3/260 systems under the Sun OS (UNIX) operating system (Hlavka and Sheffner 1988).

The implementation and testing of PEDITOR demonstrated the feasibility of performing large-area crop estimation on a microprocessor-based workstation (Hlavka and Sheffner 1988). As a follow-on to the PEDITOR project, NASS funded ARC to add software to PEDITOR that could allow for the display of digital image data. Additional modules for display were written in 1986–87, and as a result, NASS decided to explore automating the area sampling frame procedures using display of digital data as the

3

driver. Two general approaches were discussed: to automate the existing procedures, or to develop a totally new method. The two agencies decided to take the first approach as a pilot study that will eventually lead to the second. Thus, in 1987, ARC and NASS started implementing prototype software to computerize the stratification of the area sampling frame techniques.

Objectives

The existing NASS manual procedures for creating and editing area sampling frames is slow and labor-intensive. The "automated" procedure attempts to perform area frame functions with display hardware and software on a microprocessor-based workstation. The goals of the current work on the automated procedure are to complete the software and implement an operational system by 1991.

This study concentrates on the automation of the stratification and sampling unit delineation procedures, and has three basic objectives (Hanuschak et al. 1987):

1. Select or implement a software package that will supply the tools necessary for compiling and editing area frames and producing stratified sampling unit boundaries using Landsat TM and U.S. Geological Survey (USGS) Digital Line Graph (DLG) data. The software will operate on a Sun-based image display workstation.

2. Design a procedure for boundary compilation and editing based on on-screen image/photo interpretation of Landsat TM digital data and DLG data. The procedure must be user friendly, that is, the hardware and software systems must be easy to operate so that the procedure can be followed efficiently by a nontechnical staff.

3. Implement the CAS procedures in an operational environment.

TECHNICAL APPROACH

Overview

Computer-aided stratification is essentially a photographic interpretation technique using a color display and on-screen delineation of stratified sampling units, that is, PSUs. The objective is to display TM imagery and DLG data using an image processing system as an aid in the stratification process. The hardware and software chosen to implement the procedures will be discussed in this section.

Hardware System

Computer– Both NASS-AFS and ARC acquired similar hardware systems to enable parallel software development and testing; the software was developed at ARC and tested on the same basic system at NASS-AFS. The CAS software development team selected a Sun 3/260 workstation as its computing engine. The system at the NASS-AFS site is a 25-MHz, MC68020-based computer performing at three

4

MIPS (million instructions per second) of computing power, with a 20-MHz MC68881 floating-point coprocessor, and 8 megabytes (Mb) of random access memory (RAM), running under Sun OS 4.0. The system uses a Wyse amber monitor with a 24-line by 80-column display. (ARC's Sun 3/260 system uses a 19-in. monochrome high-resolution monitor as the console with 1152 by 900 pixels, and has a 372-Mb disk and a 60-Mb cartridge tape drive.) A 1.4-gigabyte (Gb) internal disk system and a 1/4-in. cartridge tape drive with 2.2-Gb capacity are the input-output (I/O) devices for the NASS-AFS system. A 19.2-kbaud modem serves as a communication link between the ARC system at Moffett Field, California, and the NASS-TRS system at Washington, D.C.

Display– A Raster Technologies Model One/25 (RT 1/25) was used initially as the CAS display system. The device is interfaced with the Sun 3/260 computer via a DR11W circuit board plugged into the VMEbus, and uses a C-language-based software driver. The RT 1/25 has 8-bit color depth for each of the colors, red, green, and blue (RGB), offering a total of 24 bits per picture element (pixel), with a display window of 512 by 512 pixels. In addition to the 24-bit raster display, there are two overlay planes which can be independently used for vector data. An RGB monitor, which displays up to 16.7 million colors, and a 13-button cursor on a X-Y digitizer tablet are interfaced to the system via an I/O port.

Software Environment

The prototype CAS software was written as a part of the PEDITOR system for ease of development and testing. PEDITOR consists of more than 100 Pascal-based functional programs (Hlavka and Sheffner 1988). Both Sun 3 systems at ARC and Fairfax have an identical copy of the software. Most of the programs are compiled under Silicon Valley Software's Pascal compiler, except for the hardware-system-dependent programs, which are written in C language.

The CAS programs have been placed in the PEDITOR menu structure under the menu selection of "Area Frame Development." There are 20 Pascal-based functional programs specifically written as a prototype for the CAS (table I) (Hanuschak et al. 1987) (see appendix, this report). During the testing phase of software development, these programs were modified as needed by ARC personnel according to requests and comments from the NASS-AFS staff. The programs can be categorized into three groups: displaying or manipulating TM and DLG data, creating and editing polygons, and system or file management. Because data files generated by CAS are in PEDITOR format, they can be read and processed by other PEDITOR functions designed to produce crop estimates. In addition, a status file, called "_statefile," serves as a CAS program-to-program communication file to retain pertinent CAS processing information, including image and screen coordinates of the display, type of file, latitude and longitude of the scene, and the pixel size of the image.

Procedures

To emulate the current manual stratification and sampling-unit delineation procedures as performed by NASS, the CAS system uses on-screen digitization techniques. The work under the CAS system will be performed on a graphic workstation display using digital imagery, instead of on false-color composite prints. The new procedure begins by entering the PEDITOR system. CAS is accessed through the "Area

Frame Development" menu selection of PEDITOR (fig. 1). The general work flow involves displaying a TM image, overlaying it with DLG data, delineating PSU boundaries, editing, and then saving the results. The following is a description of the basic steps involved in the CAS procedures.

Display TM image– The program *rtdisp* is invoked, reads three of seven TM channel windows from a disk file, and displays the superimposed RGB images onto the system color display screen in a user-selected order. The display window is 512 by 512 pixels, with an 8-bit color depth for each of the RGB images. The program supports TM data in three different formats: band sequential (BSQ), band inter-leaved by line (BIL), or pixel interleaved by line (PIL). The program *mapima* adjusts the range of brightness for each of the RGB images in order to enhance certain land-use features. A subfunction of *rtdisp* redisplays a new window of the image with the same RGB parameters that were set in the previous display, and enters new window center coordinates into the status file.

Register DLG data– The DLG data are used as a reference for the subsequent polygon digitization process in which the polygons (or PSUs) are delineated on permanent features, such as roads or rivers. A set of programs prepares the DLG data for their display. The program *dlgscn* scans each file on a USGS DLG tape and stores the tape's attribute codes into a file. The *tapdlg* program creates disk files, in PEDITOR format, of those DLG files on the USGS tape that are desired by the user. The *adddlg* program reads selected PEDITOR-format DLG files and combines them to create a single larger file. The *dspdlg* program then reads the DLG file, refers to the status file for image information, and displays the DLG lines for requested attributes, such as transportation (roads) or hydrography (water features), onto the overlay planes of the display registered approximately to the image. The *regdlg* program allows users to precisely register DLG data to any target image using a set of control points (or tiepoints). The least squares transformation generated by *regdlg* is subsequently used in *dspdlg* to adjust and overlay the registered DLG lines onto the image.

PSU delineation– The computer function *poly* allows the user to digitize or create an arbitrarily shaped polygon, and to modify or edit previously created polygons. The function currently uses a multi-button cursor on a digitizing tablet. The operator performs the delineation by typing in an identifying stratum and PSU number, delineating the area of interest, and closing the polygon. As the polygon is closed, the program annotates the longest edge of the polygon with a stratum-PSU label. The area of the digitized area in acres (converted to square kilometers for this report) and square miles is automatically computed based on the scale factor listed in the active status file. The boundaries and other associated information are written into a polygon file for future processing by subsequent CAS or PEDITOR soft-ware programs.

Checking and editing– A set of subcommands within the *poly* function allows the operator to iden-tify any ill-defined polygons, that is, those that overlap other polygons or those with gaps between them-selves and adjacent polygons. The subcommands also allow the operator to correct polygon problems by splitting polygons, moving or adding node points on the boundaries, and deleting entire polygons.

APPLICATION OF CAS PROCEDURES

As a test of the CAS software, the NASS-AFS personnel applied the CAS procedures to a three-county area in Missouri. PSU polygonal data file products, including a PEDITOR-format file of PSU boundaries and a listing of area statistics, were generated for each of the three counties.

Study Area

The three Missouri counties Livingston, Linn, and Macon (fig. 2) were chosen as the study area for the test application of the CAS procedures. They were chosen partially because of the availability of USGS DLG data, which served as a reference data plane for the area frame construction procedures.

The state of Missouri is located in one of the most important agricultural regions of the country. The state can be divided into six agricultural regions: Northern Plains, Eastern Missouri, Osage Plains region, Springfield Plateau, Ozarks, and Mississippi Alluvial Plains. The three study counties are located in the Northern Plains region, where the farms tend to be larger than elsewhere in the state. To have an overall understanding of the agricultural (and other land-use) composition of the counties, all operators of the CAS software must be familiar with the county crop statistics. The following is a brief description of the agricultural composition for each county.

Livingston County has the smallest area of the three counties, 1,373 km^2 (339,200 acres), and 55.8% (766 km^2 (189,248 acres)) of the total area is agricultural land. The important crops in order of planting areas in 1987 were soybeans, hay and wheat, sorghum, corn, and oats.

Linn County comprises 1,611 km^2 (398,080 acres), and 50.4% (812 km^2 (200,640 acres)) of the total area is agricultural land. The important crops in order of planting areas in 1987 were soybeans, hay, corn, wheat, sorghum, and oats.

The area of Macon County is 2,067 km^2 (510,760 acres), and 39.3% (812 km^2 (200,704 acres)) of the total area is agricultural land. The important crops in order of planting areas in 1987 were soybeans, hay, corn, wheat, sorghum, and oats.

Data Coverage

TM data– TM data collected by the Landsat satellite were used for the analysis of the three-county area. Livingston, Linn, and Macon counties fall into the nominal coverage area of three TM scenes: Path 25 Row 32 (P25R32), P26R32, and P26R33. Six quadrangles (quarter scenes, or quads) of TM imagery among the available scenes, selected with regard to the cloud coverage and the optimum time for the 1987 growing season, were acquired for the analysis. Quad 3 of the first scene (acquired 25 July 1987, 0% cloud coverage, Scene ID: Y5124116131X0), quad 4 of the second (acquired 17 August 1987, 10% cloud coverage, Scene ID: Y5126416203X0), and all four quads of the third (acquired 17 August 1987, 10% cloud coverage, Scene ID: Y5126416203X0) were purchased for the study. Livingston and

Linn counties were fully covered, but Macon County was only partially covered; as a result, the southern tip of Macon county was excluded from the analysis.

DLG data– The USGS DLG data covered the study area in 7.5' quadrangles (or tiles). Two data categories, transportation and hydrography, were acquired for this study. After running the *dlgscn* and *tapdlg* CAS programs, PEDITOR-format DLG files were created for each 7.5' quadrangle. The DLG files for the two categories were joined into two contiguous files for each county using the CAS function *adddlg*. The DLG files for both categories consisted of 24, 15, and 20 tiles to cover Livingston, Linn, and Macon counties, respectively. Some of the tiles were used twice for adjacent counties, and the DLG files were joined so as to cover an area larger than the county in question. However, a narrow horizontal strip of northern Linn county was not covered.

CAS Operation

Each of the three counties was independently stratified twice, by two different operators. This allowed the comparison of within-county stratum areas and PSU sizes. For comparison, eight strata codes and definitions (table II), corresponding to a PSU size tolerance guide, were set up to provide a guideline for the study. The acquired TM quads were joined to make a large window, covering three counties. The DLG transportation data were registered with the corresponding TM imagery. This required the selection of 37 tiepoints for all the counties, to overlay on the two data sets. Linn County was chosen to demonstrate the procedures.

Delineating PSUs– The operator displayed a 512- by 512-pixel window of the TM image for Linn County, starting with the county's northeastern corner. She then adjusted the color to enhance vegetation areas, agricultural or cultivated areas, and nonagricultural areas. Each enhancement was then saved in a color table for future use. The corresponding DLG transportation or hydrography vector data of the county was displayed on one overlay plane as a reference.

The operator then delineated PSUs, or polygons of uniform land-use and land-cover type bounded by permanent features, according to predetermined strata definitions, using the polygon function *poly* on a second overlay plane with the cursor and the X–Y digitizer tablet. The periodic magnification of a selected area by a factor of 2, 4, or 8, using a button-controlled zoom function, aided the operator in the definition of the PSU boundaries. In addition, the DLG data can be turned on and off independently to allow the operator to see permanent features, such as roads or rivers, which might be obscured by the DLG lines (fig. 3). The corresponding Linn County statistical data on crop types and acreages also aided the operator in identifying a PSU and in digitizing the boundaries properly. Each polygon, or PSU, was labeled with its strata code as the prefix, and with a sequential number unique within the county as the suffix. For example, 11-23 meant that this PSU was categorized as Stratum 11, "75% or more cultivated land," and was the 23rd PSU delineated in the county. The label and a calculated coverage area in acres and square miles (converted to square kilometers for this report) was displayed on the operator's text monitor and written into a computer file for future use.

After completing one 512- by 512-pixel displayed area, the operator placed the cursor on the edge of the display window and proceeded to display an adjacent area. The new cursor location was marked as the center of the new display window, and the process continued until the entire county was digitized.

Editing and saving– After completing the polygon delineation for Linn County, the operator redisplayed the PSU boundaries on the screen and used the polygon function to edit those boundaries. Polygons that overlapped or had gaps between themselves and other polygons were corrected, and polygon points were deleted, added, or moved, on different occasions. The output file of PSU boundaries was retained as a permanent record of the "frame." Editing of previously defined polygons because of changes in land-use patterns can be done at any time using the editing capabilities of the polygon delineation software.

CAS Products

The CAS procedure was applied to all three study counties, and the products for each county consisted of a polygon file containing the PSU boundaries and a list of the labels and areas of the PSUs.

Polygon file– The polygon file (fig. 3) will be used to further divide selected PSUs into smaller areas called segments. More detailed information will be collected during field visits for the segments in the selected PSUs.

PSU list– The tabular list of PSUs displays the acreage of each delineated PSU. A subcommand of *poly* automatically computes the area in square miles and in acres when the polygon is closed and labeled. Using CAS procedures, one operator delineated 85, 88, and 107 PSUs for Livingston, Linn, and Macon Counties (table III) with areas of 1,371, 1,583, and 1,981 km^2, respectively. Comparison of these area coverages with the crop statistics discussed earlier in this section shows that there are 0.0%, 1.8%, and 4.5% differences for Livingston, Linn, and Macon counties. The higher discrepancy in the Macon County coverage area was caused by the exclusion of a small part of the county, a result of the lack of TM coverage. Area files for all counties are used to randomly select PSUs for further breakdown into segments.

CONCLUSIONS AND DISCUSSION

Conclusions

At the conclusion of this study, the CAS procedures were evaluated relative to the manual method of PSU delineation, and a number of advantages of using the CAS method of automated stratification were identified.

1. The manual labor required to construct area frames was reduced substantially. For example, the construction time required for one operator to complete Macon County was 2.5 days (including digitizing time) using the CAS system, and 12.5 days plus 2.5 days for digitizing using the conventional manual method. The PSU delineation, performed manually by transferring boundaries between map and image media, can be done electronically. Much of the manual work is eliminated.

2. The ability to update an area frame is improved with the CAS system. Because the PSU boundaries can be easily edited based on the changes of the land-use pattern, the area frames can be updated with greater frequency.

3. Precision of the surveys may be enhanced. The CAS capability of viewing Landsat and map attribute data together may enable the operator to more accurately assign the PSUs to strata.

Current Development

This study tested CAS prototype software functions and procedures by applying them to three Missouri counties. The success of the system can be measured by its ability to attain the three objectives specified in the Introduction. The CAS system fulfills the first objective and the first half of the second objective by designing, implementing, and testing the prototype software. The second half of the second objective relates to the user interface of the software, and the ARC team has proposed a bit-mapped graphic user interface that will increase the utility and friendliness of the system.

In order to fulfill the third objective, the implementation of the CAS system in an operational environment, the ARC team and AFS personnel have jointly completed an investigation for a lower cost, UNIX-based image/graphics workstation. The workstations can be networked into a multidisplay system in an operational environment. These workstations will be running CAS software, using the current Sun3 system as a data server, and configured with minimum disk space for the operating system and CAS executables.

Hewlett-Packard 9000 Model 360 TurboSRX color workstations are being procured at both NASS and ARC as of this writing. The HP workstation uses an MC68030 25-MHz processor with a computing power of 5 MIPS. It also has 8 Mb of RAM, a 161-Mb disk drive, Ethernet, 19-in. noninterlaced color console and display monitor (ARC has a 16-in. noninterlaced color console for its HP workstation), and other peripherals. The TurboSRX graphics subsystem has 24-bit, 1024- by 1280-pixel color display memory, four overlay planes (one reserved, three usable), and comes with an HP Starbase Graphics Library, upon which the HP version of CAS display is to be implemented. The workstation is an HP-UX based system, an HP version of UNIX. A Pascal compiler was also included in the purchase for the PEDITOR and CAS procedures.

Current image processing hardware allows the CAS system to expand its display capability from a 512- by 512-pixel display with two overlay planes to a 1024- by 1024-pixel or larger display with up to eight overlay planes. As the cost–performance ratio decreases, the CAS software will be able to run on more powerful display systems, thus enabling the development of additional software to automate more of the presently manually oriented area frame techniques in subsequent studies.

Discussion

Hardware improvement– The addition of new hardware would improve the performance and enhance the usefulness of the CAS system. For example, an optical disk drive will increase the on-line storage space available for image display. The latest rewritable, removable optical disk drive system has

a nominal storage capacity of 650 Mb, which is more than the number of megabytes consumed by two full TM scenes. This would allow random access to TM data at a moderate speed. The computer's disk becomes a high-speed scratch pad for image windows, which would be especially useful in the CAS system. An optical electronic scanner will enable the system to incorporate hard-copy maps into the system in electronic form. Topographic maps, county highway maps, aerial photographs, and other hard-copy products will be available in electronic form as additional aids for the PSU delineation. A color printer will aid in conducting field surveys by producing high-resolution hard-copy prints of either raw or processed satellite images overlaid with PSU or segment boundaries, or transportation, political, or other useful boundaries.

Software improvement– Several improvements to the CAS software system would both enhance existing functions and enable new capabilities. Additional functions are needed to subdivide the randomly selected PSUs into segments, and then further to randomly select field segments that will be visited to establish ground truth (personal communication with J. Cotter of NASS, USDA, 1989). In order to take advantage of the availability of various map inputs via image scanner, geographical information system (GIS) capabilities could be incorporated into the software. The GIS can also utilize field data, as well as digital elevation model (DEM) and DLG data. The CAS software user interface can be improved to streamline operational procedures and increase productivity.

Future system– The CAS effort is a pilot study which seeks to automate the stratification and sampling-unit delineation portion of the NASS-AFS area sampling frame techniques. The basic steps of the process were not changed in the CAS procedures; they were merely replaced by computerized procedures. However, to take full advantage of the current computer technology, a more fundamental design is needed to greatly improve the throughput of the CAS process. The subsequent stage of the area frame techniques, the sampling procedures, should also be considered in a future automated system.

A scenario for a future automated CAS system (personal communication with M. Holko of Soil Conservation Service, USDA, 1988), which may be proposed incrementally as future projects, begins with the classification of a Landsat TM image, full or partial scene, which may cover multiple counties to produce a class map/image. This resultant image will be processed through ground-truth verification procedures to combine spectral classes into land-use types or PSUs. This type-image will then be processed using edge-finding, edge-tracing, vector-conversion, and other algorithms to produce boundaries for all the PSUs. A quality control procedure will be designed and implemented to verify the accuracy of the automated boundary delineation process. This hypothetical process greatly reduces the cost of labor-intensive photograph interpretation now a part of the stratification procedures, and produces high-quality, vectorized boundaries for PSUs. These boundaries can be used for PSU selection or as a basis of further sampling procedures.

REFERENCES

AgRISTARS: Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing Annual Report -- Fiscal Year 1980. AP-J0-04111, NASA LBJ Space Center, Houston, TX, 1981.

AgRISTARS: Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing Annual Report -- Fiscal Year 1981. AP-J2-04225, NASA LBJ Space Center, Houston, TX, 1982.

AgRISTARS: Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing Annual Report -- Fiscal Year 1982. AP-J2-0393, NASA LBJ Space Center, Houston, TX, 1983.

Allen, J. D.; and Hanuschak, G. A.: The Remote Sensing Applications Program of the National Agricultural Statistics Service: 1980 - 1987. SRB Staff Report Number SRB-88-08, National Agricultural Statistics Service, U. S. Department of Agriculture, Washington, DC, 1988.

Angelici, G. L.; Slye, R.; Ozga, M.; and Ritter, P.: PEDITOR—A Portable Image Processing System. Proceedings of IGARSS '86 Symposium, Zurich, Sept. 1986, ESP SP254, ESA, c/o ESTEC, Noordwijk, The Netherlands, 1986, pp. 265-269.

Cotter, J. J.; and Nealon, J.: Area Frame Design for Agricultural Surveys. National Agricultural Statistics Service, U. S. Department of Agriculture, Washington, DC, 1987.

Cotter, J. J.: A Tutorial on Using the Area Frame Development Functions in PEDITOR for Computer-Aided Stratification. Second Annual NASS/NASA Technical Meetings: Computer-Aided Stratification: A NASA EOCAP Award Program, Jan. 1989, NASA Ames Research Center, Moffett Field, CA, National Agricultural Statistics Service, U. S. Department of Agriculture, Washington, DC, 1989.

Craig, M. E.; Sigman, R. S.; and Cardenas, M.: Area Estimates by Landsat: Kansas 1976 Winter Wheat. U.S. Department of Agriculture, Economics Statistics and Co-op Service, Aug. 1978.

Erickson, W. K.; Hofman, L. B.; and Donovan, W. E.: The Microcomputer Workstation: An Alternate Hardware Architecture for Remotely Sensed Image Analysis. Eighth William T. Pecora Symposium, EROS Data Center, Sioux Falls, SD, Augustana Research Inst., Sioux Falls, SD, 1983.

Hanuschak, G. A.; and Morrissey, K. M.: Pilot Study of the Potential Contributions of Landsat Data in the Construction of Area Sampling Frames. Research and Development Branch, Research Division, Statistical Reporting Service, U.S. Department of Agriculture, Washington, DC, 1977.

Hanuschak, G. A.; Carney, B.; Cotter, J.; and Sheffner, E.: Compiling and Editing Area Sampling Frames Using Digital Data for Land Use Analysis and Boundary Definition. A proposal in response to NASA Research Announcement 87-OSSA-6, submitted to the NASA Earth Observation Commercial Application Program, FY88 through 91, SRB/NASS/USDA, Washington, DC, 1987.

Hlavka, C. A.; and Sheffner, E. J.: The California Cooperative Remote Sensing Project: Final Report. NASA TM-100073, 1988.

Hofman, L. B.; Erickson, W. K.; and Donovan, W. E.: MIDAS: A Microcomputer-Based Image Display and Analysis System with Full Landsat Frame Capabilities. Proceedings of the Seventeenth International Symposium on Remote Sensing of Environment, Ann Arbor, MI, May 1983, ERIM, Ann Arbor, MI, 1983, pp. 157-172.

Sigman, R. S.; Gleason, C. P.; Hanuschak, G. A.; and Starbuch, R. R.: Stratified Acreage Estimates in the Illinois Crop-Acreage Experiment. Proceedings of the 1977 Machine Processing of Remotely Sensed Data Symposium, IEEE, Piscataway, NJ, 1977, pp. 80-89.

APPENDIX

COMPUTER-AIDED STRATIFICATION PROCEDURE:
INDIVIDUAL FUNCTION DESCRIPTIONS

Add DLG Files, *adddlg*: combines selected PEDITOR DLG files to create a single (larger) PEDITOR DLG file. When several DLG files must be used for a large geographic area, *adddlg* enables the user to display all of the required DLG lines in a single execution of the *dspdlg* program.

Alter Display Status, *setdst*: allows the user to set certain entries in the display status file such as calibration file name, control point file name, scene latitude and longitude, and pixel size.

Button Assignment, *prmenu*: reinitializes all basic macros and displays the user-selectable key assignments. This function is intended to reset the key assignments to their proper configuration (table IV).

Categorized Color Mapping, *modcm*: modifies the color map of an image using predefined color values. Each image value (or range of values) may be assigned a color from a list of available colors. The program allows pseudocoloring for a single band or categorized image.

Clear Display, *eraspl*: enables the erasure of all data from user-selected image or graphics display planes on a raster graphics display device. At the program's only prompt, the user may select a plane to be cleared by typing a one-letter abbreviation.

Control Point Selections, *pickcp*: picks control points for the registration of two data planes (e.g., scene-to-scene registration). The *pickcp* function displays the control point row and column and the number of current control points on the screen. The control points chosen may be written to a file for use by other programs.

Digital Line Graph Tape Read, *tapdlg*: reads a selected file from the USGS optional format DLG tape and creates a PEDITOR-format DLG file. The header information plus the available line-attribute codes in the file (obtained from execution of the *dlgscn* program) are written to the user's terminal in an easily readable format. The *tapdlg* function extracts the DLG line data from the file for user-selected attribute codes and writes them to the output PEDITOR-format DLG file.

DLG File Display, *dspdlg*: displays DLG vectors as input from a PEDITOR DLG file on a raster graphics display device. This function displays DLG data in three modes: (1) viewing of a complete DLG file; (2) the display of a rough registration to some image data of a selected subset of the DLG data file; and (3) the display of more precisely registered DLG data using additional transformation data.

Execute System Commands, *runsys*: invokes a system command of the host computer without getting out of the CAS software.

Image Statistics, *imstat*: calculates means and standard deviations of the data image memory for rectangular areas.

Initialize Display Device, *rtinit*: initializes the RT 1/25 and invokes the basic macros and key assignments that are known to all display modules. Because of different cursors, the basic key assignments on the systems at Ames and Fairfax are different (table IV).

Load Image to Display, *rtdisp*: reads a multichannel image file from disk and displays it on the RT 1/25. Large images may be sampled to display the entire image, or 512- by 512-pixel subwindows may be displayed at full resolution. Up to three channels may be selected for display in the red, green, and blue image planes.

Mapping Functions, *mapima*: changes the color balance of a displayed image by adjusting the mapping functions of the appropriate red, green, and blue planes. Options include linear or piecewise linear mapping, equiprobable mapping, and the dynamic mapping function which allows the user to control the contrast and brightness of the displayed image by moving the cursor. The program also provides the capability to save and recall the color mapping table.

Polygon Functions, *poly*: allows users to create, edit, and save strata overlays to the image in vector format. Polygons are created by marking points on the image with the display device cursor. Other program subcommands provide for checking of digitized strata, calculating and listing of polygon areas, and calibrating of the overlay file to a geographic data base.

Register DLG to Image, *regdlg*: allows users to register DLG data to any target image using a raster graphics display device. Tiepoints are chosen by the user with the display device cursor, a control point file is generated, and a least squares transformation is calculated. The resultant transformation parameters are used in the *dspdlg* program to overlay registered DLG lines onto the image.

Scan DLG Tape, *dlgscn*: scans each file on an entire USGS optional format DLG tape and saves DLG header and line-attribute information in two text files. The unique major/minor attribute codes found in the file are written after the header information in the DLG tape log file and in the DLG attribute file to be read by the *tapdlg* program.

Segment/Polygon File Display, *segdsp*: displays a segment network file or a polygon file on a raster graphics display device. The user has the option of registering the data to the image stored in the image planes of the display device. The user is able to assign a color to a single field or to a range of fields. Field coloring can be done either with the cursor or from the terminal.

Show Display Status, *showds*: shows the display status file, *_statefile*, which contains information on such parameters as image-plane routing, coordinates of the display window, coordinates of the window file relative to the Landsat scene, sampling rate, and pixel size.

Thematic Mapper Tape Read, *tmread*: reads Landsat TM multiple-channel data from multiple-tape volumes and writes the data into a PEDITOR-format file. Although *tmread* was implemented for the CAS system, it is presently located under the "Tape Functions" item in the PEDITOR menu.

Quit, *quit*: terminates Area Frame Development functions, and goes to the PEDITOR menu.

TABLE I.– CAS FUNCTIONAL PROGRAMS[a]

Function[b]	Module	Date	Author
Add DLG Files	*adddlg*	09/09/87	Angelici
Alter Display Status	*setdst*	12/23/86	Slye
Key Assignment	*prmenu*	12/09/86	Ma
Categorized Color Mapping	*modcm*	11/27/88	Slye
Clear Display	*eraspl*	12/16/88	Angelici
Control Point Selections	*pickcp*	09/22/87	Slye
Digital Line Graph Tape Read	*tapdlg*	12/06/88	Angelici
DLG File Display	*dspdlg*	12/15/88	Angelici
Execute System Commands	*runsys*	12/04/86	Slye
Image Statistics	*imstat*	03/20/87	Sheffner
Initialize Display Device	*rtinit*	10/28/86	Ma, Slye
Load Image to Display	*rtdisp*	12/05/88	Slye
Mapping Functions	*mapima*	05/15/87	Slye
Polygon Functions	*poly*	11/14/88	Ma, Slye
Register DLG to Image	*regdlg*	12/19/88	Angelici
Scan DLG Tape	*dlgscn*	10/08/86	Angelici
Segment/Polygon File Display	*segdsp*	02/19/87	Sheffner
Show Display Status	*showds*	12/11/86	Slye
Thematic Mapper Tape Read	*tmread*	10/20/88	Slye
Quit	*quit*		

[a]These Pascal-based programs were written in the 3-yr period from 1986 through 1988 as a prototype for the CAS system; they have been constantly modified for more efficient processing.

[b]Described in the appendix.

TABLE II.– USDA PRIMARY SAMPLING UNIT (PSU) AND STRATA DEFINITION

Stratum	Definition	Area, km^2		
		Minimum	Desired	Maximum
11	75% or more cultivated land	2.6	15.6 to 20.8	31.1
12	50-74% cultivated land	2.6	15.6 to 20.8	31.1
20	15-49% cultivated	2.6	15.6 to 20.8	31.1
31	Ag-urban, less than 15% cultivated	0.6	2.6 to 5.2	7.8
32	Residential-commercial land	0.3	1.3 to 2.6	2.6
40	Range and pasture, less then 15%	5.2	31.1 to 41.4	62.2
50	Nonagricultural land	2.6	None	None
62	Water	2.6	None	None

TABLE III.– LAND USE PROFILE

Stratum	Livingston PSU[a]	Livingston Area, acres (km²)	Linn PSU[a]	Linn Area, acres (km²)	Macon PSU[a]	Macon Area, acres (km²)	Total, acres (km²)
11	39 (15.5)	150,860.0 (610.4)	43 (19.7)	210,304.3 (851.1)	19 (18.1)	86,154.8 (348.6)	447,319.1 (1,810.2)
12	23 (14.2)	81,715.7 (330.5)	34 (17.6)	148,575.4 (601.1)	16 (14.5)	58,168.7 (235.2)	288,459.8 (1,167.3)
20	18 (21.5)	96,051.6 (388.5)	8 (14.2)	28,457.2 (115.0)	52 (21.0)	270,883.1 (1,096.1)	395,391.9 (1,599.8)
31	1 (8.0)	2,031.8 (8.0)	2 (3.6)	1,844.8 (7.3)	2 (1.7)	838.1 (3.4)	4,714.7 (18.9)
32		--		--	1 (0.8)	199.8 (0.8)	199.8 (0.8)
40	4 (8.3)	8,198.8 (33.2)		--	15 (18.6)	70,008.8 (283.1)	78,207.6 (316.2)
50		--	1 (7.5)	1,919.8 (7.5)	1 (4.4)	1,141.3 (4.4)	3,061.1 (12.2)
62		--		--	1 (8.3)	2,085.5 (8.3)	2,085.5 (8.3)
Total PSUs Total acres Total km² Agric. acres[b] Agric. km²	85	338,858.1 (1,371.1) 328,627.3 (1,329.7)	88	391,101.8 (1,582.5) 387,336.9 (1,567.3)	107	489,480.5 (1,980.8) 415,206.6 (1,680.0)	1,219,439.5 (4,934.7) 1,131,170.8 (4,577.0)
							Average
Agric./total acres ratio (%) Avg PSU acres Avg PSU km²		96.9 4,458.6 (17.9)		99.0 4,444.3 (17.9)		84.8 4,574.5 (18.4)	92.8 4,499.8 (18.1)

[a]Number of PSUs (average PSU size, km²).

[b]Agricultural or cultivated area.

18

TABLE IV.– KEY ASSIGNMENTS FOR VARIOUS FUNCTIONS ON AMES AND FAIRFAX SYSTEMS

Function	Key assignment	
	Ames	Fairfax
Zoom	1	1
Pan	2	4
Toggle cursor color	3	C
Toggle image planes (route)	4	D
Toggle color of graphics plane 0	5	E
Toggle color of graphics plane 1	6	F
Red	7	6
Green	8	7
Blue	9	5
Toggle on/off of graphics plane 0	*	A
Toggle on/off of graphics plane 1	0	B
Quit	#	2
Select	Grey	3

PEDITOR OPENING MENU

Aggregation Functions
Area Frame Development
Binary File Display
Cray Functions
Display Functions
Estimation Functions
Ground Truth Functions
Identify a File
Mask Functions
Plot Functions
Registration and Digitization
  Functions
Segment Functions
Statistical Funtions
Window Functions
Quit

①

Area Frame Development

Add DLG Files
Alter/Show Display Status
Categorized Color Mapping
Clear Display
Digital Line Graph Tape Read
DLG file Display
Execute System Commands
Image Statistics
Initialize Display Device
Load Image to Display
Mapping Functions
Polygon Functions
Register DLG to Image
Scan DLG Tape
Segment/Polygon File Display
Show Display Status
Thmatic Mapper Tape Read
Quit

②

Loading Images

Display an Image
Map Image with Colormap
Redisplay Current Image
Set Parameters
Use Other Directories
Quit

④

Polygon Commands

Area Recomputation
Calibrate Polygon Data
Check Digitized Strata
Clear Current Overlay Plane
Delete a Polygon
Enter Digitizing Mode
File List Polygons
Get Other Directories
Initialize Everything
Labels On/Off
LH (List Holes)
List Polygons
Merge Polygons
Move Overlay to Image
Pixel Size
Read Polygon File
Refresh the Screen
Relabel a Polygon
Resequence Polygon Labels
Save Increment
Set Parameters for Output
Stratum Display
Write Polygon File
Quit

③

Mapping Functions

Blank Display Plane
Change Routing
Dynamic Mapping
Equiprobable Mapping
Histogram
Initialize Mapping
Linear Mapping
Piecewise Linear Mapping
Recall Mapping
Save Mapping
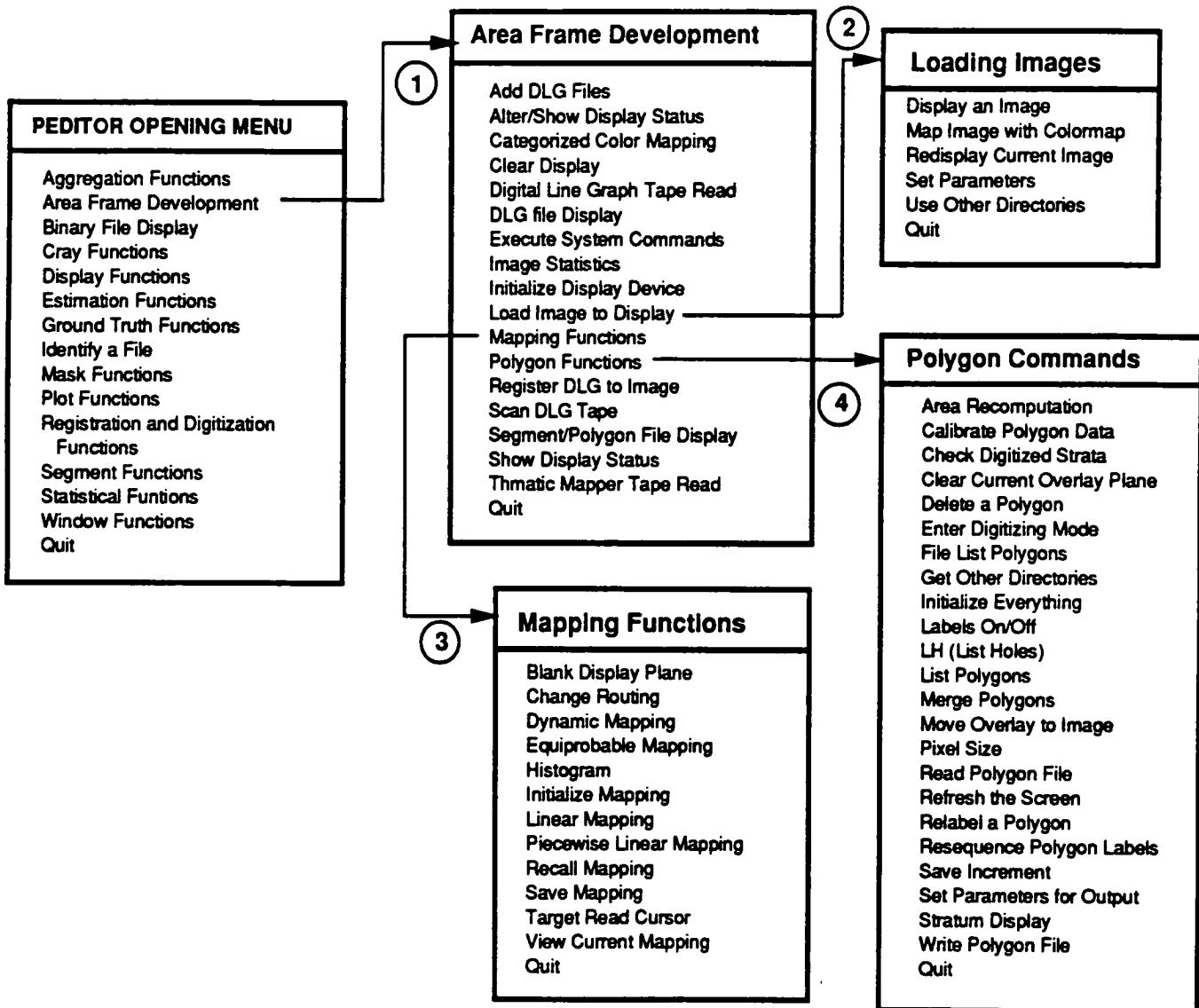Target Read Cursor
View Current Mapping
Quit

Figure 1.– Menu selections for the computer-aided stratification procedure (Cotter 1989).
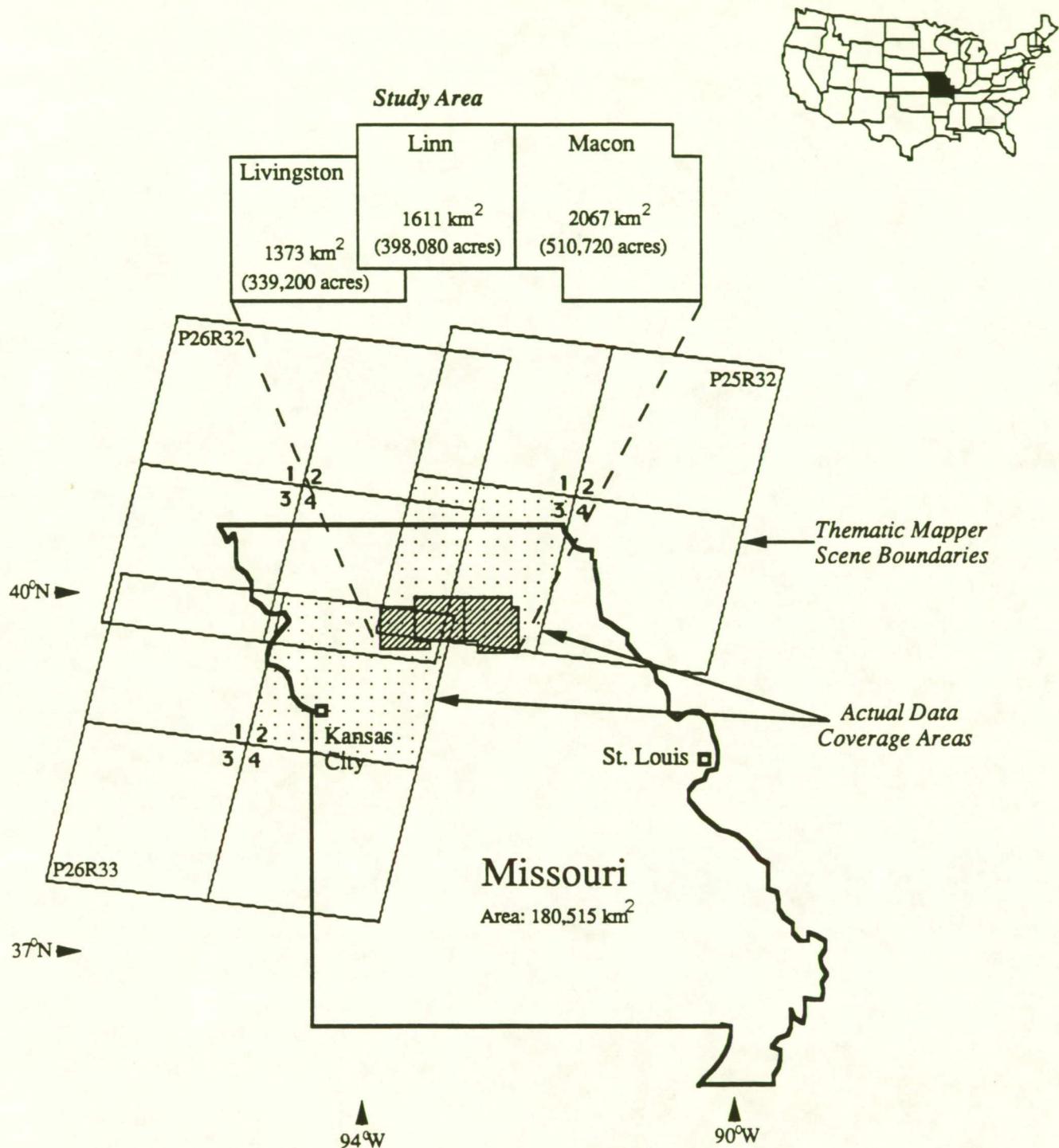
20

Figure 2.– Study area.

Figure 3.– DLG and PSU boundaries. A 512- by 512-pixel window of TM image of Linn County was displayed using channels 1, 3, and 5 in blue, green, and red, respectively. Therefore, the highly vegetated areas appear to be dark red. Using the two graphic overlay planes, DLG transportation boundaries were overlaid in blue, and the boundaries of the PSUs were overlaid with strata labels in white.

23

NASA
National Aeronautics and
Space Administration

Report Documentation Page

1. Report No. NASA TM-102243	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Computer-Aided Boundary Delineation of Agricultural Lands		5. Report Date November 1989
		6. Performing Organization Code
7. Author(s) Thomas D. Cheng (TGS Technology, Inc., Moffett Field, CA), Gary L. Angelici (Sterling Federal Systems, Inc., Palo Alto, CA), Robert E. Slye, and Matt Ma (TGS Technology, Inc., Moffett Field, CA)		8. Performing Organization Report No. A-89264
		10. Work Unit No. 672-90-20-03
9. Performing Organization Name and Address Ames Research Center Moffett Field, CA 94035-1000		11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546-0001		13. Type of Report and Period Covered Technical Memorandum
		14. Sponsoring Agency Code

15. Supplementary Notes

Point of Contact: Robert E. Slye, Ames Research Center, MS 242-4, Moffett Field, CA 94035-1000
(415) 604-6031 or FTS 464-6031

16. Abstract

The National Agricultural Statistics Service of the United States Department of Agriculture (USDA) presently uses labor-intensive aerial photographic interpretation techniques to divide large geographical areas into manageable-sized units for estimating domestic crop and livestock production. Prototype software, the computer-aided stratification (CAS) system, was developed to automate the procedure, and currently runs on a Sun-based image processing system. With a background display of Landsat Thematic Mapper and United States Geological Survey Digital Line Graph data, the operator uses a cursor to delineate agricultural areas, called sampling units, which are assigned to strata of land-use and land-cover types. The resultant stratified sampling units are used as input into subsequent USDA sampling procedures. As a test, three counties in Missouri were chosen for application of the CAS procedures. Subsequent analysis indicates that CAS was five times faster in creating sampling units than the manual techniques were.

17. Key Words (Suggested by Author(s)) Boundary delineation Image processing Agricultural stratification	18. Distribution Statement Unclassified-Unlimited Subject Category - 43		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 27	22. Price A03

NASA FORM 1626 OCT 86

For sale by the National Technical Information Service, Springfield, Virginia 22161