

Selected Papers in the Applied Computer Sciences 1988

U.S. GEOLOGICAL SURVEY BULLETIN 1841



AVAILABILITY OF BOOKS AND MAPS OF THE U.S. GEOLOGICAL SURVEY

Instructions on ordering publications of the U.S. Geological Survey, along with prices of the last offerings, are given in the current-year issues of the monthly catalog "New Publications of the U.S. Geological Survey." Prices of available U.S. Geological Survey publications released prior to the current year are listed in the most recent annual "Price and Availability List." Publications that are listed in various U.S. Geological Survey catalogs (see back inside cover) but not listed in the most recent annual "Price and Availability List" are no longer available.

Prices of reports released to the open files are given in the listing "U.S. Geological Survey Open-File Reports," updated monthly, which is for sale in microfiche from the U.S. Geological Survey, Books and Open-File Reports Section, Federal Center, Box 25425, Denver, CO 80225. Reports released through the NTIS may be obtained by writing to the National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161; please include NTIS report number with inquiry.

Order U.S. Geological Survey publications **by mail** or **over the counter** from the offices given below.

BY MAIL

Books

Professional Papers, Bulletins, Water-Supply Papers, Techniques of Water-Resources Investigations, Circulars, publications of general interest (such as leaflets, pamphlets, booklets), single copies of Earthquakes & Volcanoes, Preliminary Determination of Epicenters, and some miscellaneous reports, including some of the foregoing series that have gone out of print at the Superintendent of Documents, are obtainable by mail from

U.S. Geological Survey, Books and Open-File Reports
Federal Center, Box 25425
Denver, CO 80225

Subscriptions to periodicals (Earthquakes & Volcanoes and Preliminary Determination of Epicenters) can be obtained **ONLY** from the

Superintendent of Documents
Government Printing Office
Washington, D.C. 20402

(Check or money order must be payable to Superintendent of Documents.)

Maps

For maps, address mail orders to

U.S. Geological Survey, Map Distribution
Federal Center, Box 25286
Denver, CO 80225

Residents of Alaska may order maps from

Alaska Distribution Section, U.S. Geological Survey,
New Federal Building - Box 12
101 Twelfth Ave., Fairbanks, AK 99701

OVER THE COUNTER

Books

Books of the U.S. Geological Survey are available over the counter at the following Geological Survey Public Inquiries Offices, all of which are authorized agents of the Superintendent of Documents:

- WASHINGTON, D.C.--Main Interior Bldg., 2600 corridor, 18th and C Sts., NW.
- DENVER, Colorado--Federal Bldg., Rm. 169, 1961 Stout St.
- LOS ANGELES, California--Federal Bldg., Rm. 7638, 300 N. Los Angeles St.
- MENLO PARK, California--Bldg. 3 (Stop 533), Rm. 3128, 345 Middlefield Rd.
- RESTON, Virginia--503 National Center, Rm. 1C402, 12201 Sunrise Valley Dr.
- SALT LAKE CITY, Utah--Federal Bldg., Rm. 8105, 125 South State St.
- SAN FRANCISCO, California--Customhouse, Rm. 504, 555 Battery St.
- SPOKANE, Washington--U.S. Courthouse, Rm. 678, West 920 Riverside Ave..
- ANCHORAGE, Alaska--Rm. 101, 4230 University Dr.
- ANCHORAGE, Alaska--Federal Bldg, Rm. E-146, 701 C St.

Maps

Maps may be purchased over the counter at the U.S. Geological Survey offices where books are sold (all addresses in above list) and at the following Geological Survey offices:

- ROLLA, Missouri--1400 Independence Rd.
- DENVER, Colorado--Map Distribution, Bldg. 810, Federal Center
- FAIRBANKS, Alaska--New Federal Bldg., 101 Twelfth Ave.

Selected Papers in the Applied Computer Sciences 1988

Edited by D.A. WILTSHIRE

U.S. GEOLOGICAL SURVEY BULLETIN 1841

DEPARTMENT OF THE INTERIOR
DONALD PAUL HODEL, Secretary

U.S. GEOLOGICAL SURVEY
Dallas L. Peck, Director



UNITED STATES GOVERNMENT PRINTING OFFICE: 1988

For sale by the Books and Open-File Reports Section,
U.S. Geological Survey, Federal Center,
Box 25425, Denver, CO 80225

Library of Congress Cataloging in Publication Data

Selected papers in the applied computer sciences, 1988.
(U.S. Geological Survey bulletin ; 1841)

Bibliography: p.

Supt. of Docs. no.: I 19.3:1841

1. Remote sensing—Data processing. I. Wiltshire, Denise A. II. Series.
QE75.B9 no. 1841 [G70.4] 557.3s 88-600051 [621.36'78]

FOREWORD

This compilation of short papers is intended to describe technical advances in the applied computer sciences. The aim of the series is to provide a forum for discussions on innovative uses of computing techniques in the earth sciences. Papers are presented on (1) an expert system to facilitate land analysis, (2) data-generalization processes for use with geographic information systems, (3) digital image processing for map production, (4) connecting local- and wide-area networks, and (5) generic file structures for hydrologic data.

CONTENTS

FOREWORD

CHAPTER A

An expert-system front end for the Land Analysis System

By Stuart W. Doescher, Donna K. Scholz, and Bruce K. Quirk

CHAPTER B

A study of four raster-based data generation procedures

By Margaret G. Mayers, Eugene A. Fosnight, and James A. Sturdevant

CHAPTER C

Digital image processing of Landsat image map products at the U.S. Geological Survey

By Richard D. Kidwell, Douglas R. Binnie, and Sheila E. Martin

CHAPTER D

Connecting local- and wide-area networks for distributed access to a microcomputer application

By Richard A. Hollway and Denise A. Wiltshire

CHAPTER E

Generic file structures for selected applications of water-resources data

By Mark L. Farmer

CHAPTER A

An Expert-System Front End for the Land Analysis System

By STUART W. DOESCHER, DONNA K. SCHOLZ, and
BRUCE K. QUIRK

U.S. GEOLOGICAL SURVEY BULLETIN 1841

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES

CONTENTS

Abstract	A1
Introduction	A1
Expert systems	A2
Expert-system characteristics	A2
Environments	A3
Land Analysis System front-end concepts	A3
Expert-system shell	A4
Sample knowledge bases	A4
Elements of a knowledge base	A5
Product clauses	A5
Control clauses	A5
Set parm clauses	A6
Procedure clauses	A6
Parameter clauses	A6
Check parm clauses	A8
Translate clauses	A8
Help clauses	A8
Conclusion	A9
Plans and recommendations	A9
References	A11

FIGURES

1. Relation among the expert-system shell, the knowledge bases, and the Land Analysis System A4
2. Expert-system interaction with the user and the “shade” knowledge base A10

An Expert-System Front End for the Land Analysis System

By Stuart W. Doescher, Donna K. Scholz,¹ and Bruce K. Quirk¹

Abstract

A Land Analysis System front-end expert system was built to demonstrate and assess the utility of expert systems in assisting users that have image-processing tasks and to gain experience. This expert system implements the concept that the users of computer systems must be provided with increasing expertise as the software system that they use becomes more complex. Expertise and knowledge of the software system is needed to perform efficiently the proper sequencing of modules that will produce the desired results. The users of these software systems often know "what" they want to do, but the problem they encounter is "how" to get it accomplished. A prototype expert system incorporating the knowledge associated with "how" to generate shaded-relief images and to produce raster and vector combination Applicon plots was implemented by using Prolog.

INTRODUCTION

The U.S. Geological Survey uses a variety of geographic information systems (GIS's) in accomplishing its mission. Some of these GIS's, which are raster or grid-cell based and support remote sensing activities, are known as image-processing systems. In cooperation with the National Aeronautics and Space Administration Goddard Space Flight Center, the National Mapping Division at the Earth Resources Observation Systems (EROS) Data Center has been developing an image-processing system called the Land Analysis System (LAS). This system is used in research and in production activities at the EROS Data Center.

LAS consists of over 200 modules that support the functional areas of data ingest, radiometric and geometric correction, signal processing, logical operations, classification, display management, tabular analysis, and output product generation (Quirk and Oleson, 1987). Through a LAS application executive, the user is provided with an interface to control the functionality and the execution of these modules. The functionality of these modules is controlled through parameters that are set by the user.

When these LAS modules and the user interface are in place, the research scientist and the production personnel have the functionality available to perform their particular tasks. However, the user still may be perplexed by such concerns as "What is the best combination and proper ordering of modules to accomplish my specific task or to generate my desired product?" or "What is the best parameter setting for these modules based upon what I know about the characteristics of the data and the type of results I expect?" In general, the user

¹ TGS Technology, Inc., Earth Resources Observation Systems Data Center, Sioux Falls, SD 57189. Work performed under U.S. Geological Survey contract 14-08-0001-22521.

Any use of trade names and trademarks in this publication is for descriptive purposes only and does not constitute endorsement by the U.S. Geological Survey.

is saying, "I know 'what' I want to do, but I am not sure 'how' to get it accomplished." The answers to these "how" questions traditionally have come from people who are the "experts" in using image-processing systems.

This paper first discusses characteristics of expert systems and then describes the LAS knowledge bases using examples of the Prolog implementation of a shaded-relief expert-system prototype.

EXPERT SYSTEMS

By using definitions of expert systems such as "a system that utilizes a significant amount of expert information about a particular domain to solve problems in that domain" (Fischler and Firschein, 1986) and "a system that processes data and delivers or administers the advice or expertise that a human expert would give in the same circumstances" (Rolandi, 1986), one could expect that expert systems might assist in answering the "how" part of the user's concern. However, it must be recognized that expert systems, also known as knowledge-based systems and rule-based representation systems, have limitations. These limitations are the lack of automated learning, the inability to generalize, and the loss of competency when applied outside the intended or defined domain of the expert system.

EXPERT-SYSTEM CHARACTERISTICS

A closer review of the expert-system characteristics as they apply to our situation was performed to determine if an expert system was feasible to develop and if it had a potential usefulness. The following characteristics of expert systems (Feigenbaum and McCorduck, 1983) were evaluated:

1. Usually, expert systems are applied to limited or specific problem areas that are referred to as domains.

In our case, the domain is limited to image processing that has a special emphasis on LAS-processing capabilities. Further refinement of this domain can be made by concentrating on specific image-processing problem areas and by allowing growth into other areas.

2. The expert system contains knowledge or expertise supplied by a specialist.

By knowing the specific problem of "what" is needed, the image-processing experts at the EROS Data Center can define the "how" methodologies. They can identify an appropriate module sequence and, on the basis of the specific problem, can provide appropriate parameters. The needed expertise may not reside totally with one expert, but may require several experts to accomplish a particular task. The development of an expert system's set of rules may consolidate the experts' views.

3. A knowledge engineer working with the experts develops the rules for the knowledge base.

The definition of the rules that make up the knowledge base is the critical step in building an expert system. The knowledge engineer is needed to give structure or organization to the rule-building process and, therefore, must have an understanding of the mechanisms by which the rules are to be used by the expert system and must be able to capture the knowledge supplied by the experts in the form of these rules. In our situation, the rules for our prototype were built through iterative conversations with the users who would ask "How do I do this?" and consultation with the expert, who would answer, "In this circumstance, I would do it this way."

4. The expert system will provide consultation, advisory, or training functions for users.

Once the knowledge has been encoded in the expert system, does the system provide a useful tool for the user community? Will members of the user community accept the advice or the assistance if they consider themselves experts? These questions remain to be answered, but it is our hope that a proper selection of a domain was made to demonstrate more than just a concept. The selection of the domain in the area of image processing must be balanced between the “too trivial,” where the user would not bother to use the expert system, and the “too complicated,” where the expert-system builders could not complete the knowledge base.

5. The knowledge base should be easy to modify or augment.

New image-processing methodologies are being developed and enhanced continuously. The knowledge base should represent “how” to process the data by using current techniques. It should allow for the evolution and the growth of the knowledge base. Because the user should have available the best current technology, the knowledge base must be dynamic and easily modified.

ENVIRONMENTS

In the determination of an environment for an expert system, many hardware and software choices and trade-offs must be made. The expert system should be easily accessible and comfortable to operate. The developer and the knowledge engineer should be provided with sufficient tools to allow easy development and maintenance of the system. The computer environment should be structured to accommodate growth beyond the demonstration phase. For the initial prototype, a microcomputer that had an expert-system shell written in Prolog (Clocksin and Mellish, 1984) seemed most suitable. Prolog was chosen because it gave the developers a fundamental understanding and control of the user interface and experience in the modeling aspects of expert-systems technology. Prolog also permits the customized creation of command language statements that are used as inputs to LAS and provides a pathway for evolution to larger computers.

LAND ANALYSIS SYSTEM FRONT-END CONCEPTS

The demonstration expert system has been designed as a LAS front end, which provides a user interface having a question-and-answer-type format. The system asks only pertinent questions of the user and then determines the next question based on the user's answers. The knowledge or rule base has been structured to support this type of question-and-answer interface by using a deterministic decision tree where the answer to one question determines the next question.

The expert system and its knowledge base should provide the “how” side of running LAS and should eliminate the user's need to know LAS module names and their functions, parameter names and appropriate values, and the proper module sequencing. Thus far, the knowledge base contains rules that will produce standard as well as customized products. The user's interaction with the system can be set either at a novice level, where the knowledge base is used to set specified parameter settings by default, or at an expert level, where the user is able to modify or refine the initial parameter settings that have been provided from the knowledge base.

The result of the question-and-answer session is the generation of a set of LAS commands that can be studied by the user as a learning aid or can be sent as a command file to a system that supports the LAS image-processing system. These LAS commands contain the instructions that reflect the knowledge of “how” an expert would have performed the task.

EXPERT-SYSTEM SHELL

The expert-system shell has three main sections. The introductory section allows the user to specify the knowledge base of interest, to select a novice or expert consultation level, to view the products supported in the knowledge base, and to initiate a consultation. The body of the expert-system shell allows the user to specify the product that is desired and initiates the root level of questioning and control within the deterministic decision tree of the knowledge base. On the basis of the user-supplied answers, elements of the knowledge base are cycled until the branches of the tree are exhausted. The closing section prompts the user for the type of output that is desired. The user can review the LAS commands that have been prepared or can save them in a file to be submitted to the appropriate system for LAS processing. The user also can restart the consultation or can exit the expert system. It is intended that the expert-system shell remain unchanged when additional knowledge bases are built or modified.

SAMPLE KNOWLEDGE BASES

Two sample knowledge bases were prepared to demonstrate the possibilities of an expert system as a front end to LAS (fig. 1). As instructed by the user, a knowledge base is consulted by the expert-system shell. Upon completion of the question-and-answer dialog with the user, LAS commands that are built by the expert-system shell may be submitted for LAS processing.

The “shade” knowledge base was built to assist in the preparation of shaded-relief products. A shaded-relief product is generated by “illuminating” Digital Elevation Model (DEM) data, which represent the uses and the types of terrain, from a user-specified illumination source. The position of the source is identified by an elevation angle above the horizon and a direction angle measured clockwise from north. The algorithm determines the positional vector

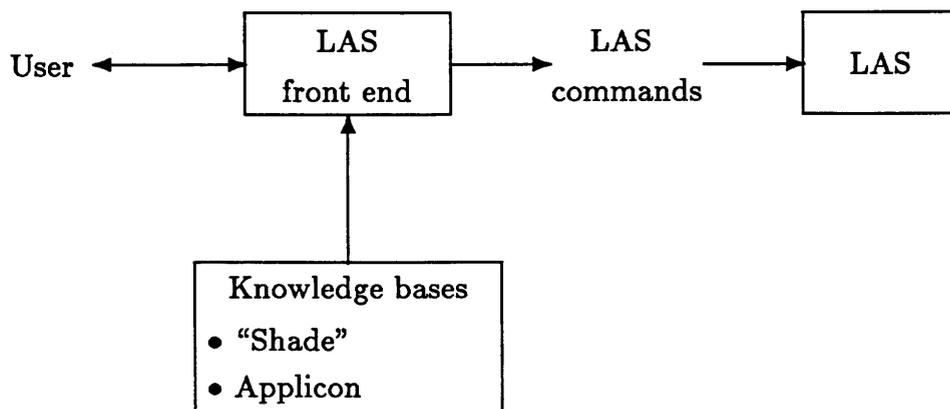


Figure 1. Relation among the expert-system shell, the knowledge bases, and the Land Analysis System (LAS).

of the illumination source and a vector normal to the terrain's surface at each point within the DEM to calculate the level of reflectance that would be observed by the viewer. The algorithm uses a variety of parameters to produce any of the potential products. This knowledge base contains the rules that control the different parameter settings for each base-map-type product; these settings depend on whether the result will be combined with other thematic data layers or whether the main purpose is to depict the terrain relief in detail. Further parameter settings must be specified on the basis of the type of terrain. Mountainous, hilly, or flat landscapes each require different types of illumination to optimize shading of their features. This knowledge base also walks the user through the steps that are necessary for loading digital data into the computer system for LAS processing.

The Applicon base of knowledge was constructed to demonstrate the expert system's ability to ensure proper data flow through a series of LAS modules to produce color inkjet plots. The Applicon inkjet plotter is used at the EROS Data Center to produce a variety of map products. It supports the plotting of vector data (points, lines, polygons) and raster data (DEM's, Landsat images), as well as combinations of vector and raster data types. Users who are not expert in the use of the five or six modules that are needed to construct Applicon products spend several hours reading and interpreting the documentation to generate a single product. The documentation is not inadequate, but it does present an abundance of alternatives. This knowledge base assists in preparing a combination of LAS commands for generating an Applicon product.

ELEMENTS OF A KNOWLEDGE BASE

A knowledge base consists of Prolog clauses that contain the information and the control for questioning the user and that create the LAS commands. A knowledge base is structured so that it can be built and maintained by the domain expert. Elements include the product clause for describing the available end products, control clauses for controlling the session flow, set parm clauses for setting parameter values, procedure clauses for building the LAS command line, and parameter, check parm, translate, and help clauses for controlling parameter settings. The knowledge base elements as used in the "shade" example and their functions are described in detail below.

Product Clauses

The product clauses contain a description of each of the end products that are contained within the knowledge base. The user can request this information as a starting point for use of the control clauses.

Control Clauses

The control clauses provide for the overall flow of the questions and answers and are of three types. The first type provides information to the user. The second responds to a yes or no answer from the user. The third presents the user with a multiple choice. Each control clause specifies text that is presented to the user as information or as a question. Each clause also contains lists of further clauses that will be used to govern the overall flow of the expert system. The type of question (informational, yes or no, multiple choice) and the response from the user determine which list is to be processed next.

The example control clause shown below questions the user about the type of terrain in the area of the DEM. The expected answer is “m, n, or f.” This answer will cause control to pass to the set parm clauses, which will specify the parameter setting of Sun angle appropriate for the type of relief in the image.

```
control(display,          var,
        ['Is the area you want to create a shaded-relief for, one that is',
         '(m)ountainous, (h)illy, or relatively (f)lat '],
        [
          m,[millum],
          h,[hillum],
          f,[fillum]  ],          []).
```

Set Parm Clauses

A set parm clause is used to set a parameter or a group of LAS parameter values. During the questions and answers, the user responses set parameter values required when procedure clauses are processed.

An example of set parm that would set the Sun angle according to terrain characteristics, as discussed above, is as follows:

```
set_parm(fillum,  [angle,'10']).
set_parm(hillum,  [angle,'30']).
set_parm(millum,  [angle,'50']).
```

Procedure Clauses

A procedure clause is used to build the LAS command line. This clause consists of the LAS module name along with a list of parameters that are needed for this LAS function. The expert-system shell determines the value for each parameter name in the list on the basis of either the user's answers or the information in the parameter clauses.

For the shaded-relief example, the procedure clause specifies the parameters that are required for the shading algorithm as follows:

```
proc(pic_c,          picshade-create,
     [in,cout,method,pixsize,mask,isf,angle,azimuth,
     expnt,osf,offset,odtype]).
```

Parameter Clauses

The parameter clauses are unique for each LAS parameter. These clauses contain an internal parameter name, the LAS parameter name, the default value to be used, the question that may be posed to the user, and a list of other parameters' names whose values will be set identical to this one. The user will be prompted by the question in the parameter clause only if the parameter value has not been set previously as a result of an earlier set parm clause. If a default value of “none” is specified, then the question will be presented to the user to

determine the appropriate parameter value. However, if a default value is present within the clause and the consultation level is set to "novice," then the default value is used. If the expert system consultation level is set at "expert," then the question is presented to the user, and the default value is shown.

In the following example, the parameter clauses shown are called by the picshade-c procedure clause. Because the parameter clauses for in and cout have "none" specified for the default parameter setting, these two parameters must always be supplied by the user. The remaining parameters have defaults (method = 1, pixsize = 30,30, and so forth); therefore, novice users are not required to supply them.

```
parm(in,      in,      none,      [],
     ['What is the name of the input DEM file']).

parm(cout,    out,    none,    [],
     ['What name do you wish to give to your resulting shaded',
      'relief image']).

parm(method,  method,  '1',      [],
     ['What is the method of calculating the directional',
      'derivative (1) 1x3 or (2) 3x3 convolution']).
check_parm(method,  specific,      ['1','2']).

parm(pixsize, pixsize,  '(30,30)', [],
     ['What is the pixel size of the data in line and sample',
      'directions measured in the same units as the input',
      'elevation values']).

parm(mask,    mask,    '0',      [],
     ['If there is a value used to mask your study area',
      'or areas of bad data values, what is this value']).

parm(isf,     isf,     '1',      [],
     ['If you would like to accentuate the scale of the relief ',
      'please provide a scaling factor to multiply elevations by']).
check_parm(isf,  range,      ['1','100']).

parm(angle,   angle,   '30',     [],
     ['The angle of the Sun above the horizon controls the amount ',
      'of shadowing in the resultant output image.',
      'Would you like HEAVY, MODERATE, or LIGHT shadowing']).
check_parm(angle,  range,      ['0','90']).
translate(angle,
           [heavy,moderate,light],
           ['10', '30',  '50']).

parm(azimuth, azimuth,  'nw',     [],
     ['From what direction would you like the sun to shine ']).
check_parm(azimuth,  range,      ['0','360']).
translate(azimuth,
           [ north,  east,  south,  west  ],
           ['0',    '90',   '180',   '270' ]).
```

```

parm(expnt,    expnt,    '0',    [],
     ['To what degree would you like to subdue low relief features',
      'and accentuate high relief detail of your output image',
      'on a scale of 0(none) to 10(much)']).
translate(expnt,
          ['0','1', '2', '3', '4', '5', '6', '7', '8', '9', '10'],
          ['1','.9','.8','.7','.6','.5','.4','.3','.2','.1','0'] ).

parm(osf,      osf,      '0',      [],
     ['To what degree would you like to accentuate low relief detail',
      'at the expense of saturating high relief detail in your output',
      'image on a scale of 0(none) to 2(much)']).
translate(osf,
          ['0','1','2'],
          ['1','2','3']).

parm(offset,   offset,   '0',      [],
     ['To what degree would you like to subdue the overall contrast',
      'of your output image on a scale of 0(none) to 10(much)']).
translate(offset,
          ['0','1', '2', '3', '4', '5', '6', '7', '8', '9', '10'],
          ['0','.1','.2','.3','.4','.5','.6','.7','.8','.9','1']).

parm(odtype,   odtype,   '1',      [],
     ['What data type would you like your output image to be',
      ' (1) = byte and (4) = real*4']).
check_parm(odtype,    specific,    ['1','4']).

```

Check Parm Clauses

The check parm clauses contain information that is used to check the user's response to a setting of a parameter. If the user's response is not proper, then the user is notified and informed what responses might be acceptable. The example above contains some sample check parm clauses; for instance, azimuth will accept any number within the range of values between 0 and 360.

Translate Clauses

In some situations, the translate clauses are used to assist the user by translating English-language-type answers to the appropriate LAS parameter values. A user may feel more comfortable answering a question on compass direction by providing an answer like "east." In this situation, the translate clause would convert "east" to the value of 90 degrees, which is an acceptable numerical value for the LAS parameter azimuth (see the example in the section "Parameter Clauses").

Help Clauses

The help clauses teach the user the "what's" and the "why's" of questions being asked. He can request help at any time and be presented with a rationale

that an expert would provide at that point in the decision process. The user can request this information to enhance his own level of expertise.

The example below is a help clause explaining the importance of the display control clause. This information will assist the user in understanding how Sun elevation and terrain type will impact the final results of the shading process.

```
help(display,e,  
  ['The amount of topographic relief in an area is a factor ',  
   'when deciding on an optimum sun elevation for creating an ',  
   'attractive shaded-relief product. In relatively flat areas ',  
   'you will need a lower Sun angle to bring out shadows for ',  
   'subtle terrain features. In hilly areas you still want a ',  
   'moderately low Sun angle to ensure enough shadowing to depict ',  
   'rolling terrain. In mountainous areas a high Sun angle will ',  
   'illuminate the numerous slope faces without sacrificing the ',  
   'more rugged appearance of the area.']).
```

Figure 2 shows an interactive session of the LAS front-end expert system. The example illustrates the questions that are asked of the novice user on the computer terminal screen for generation of a shaded-relief display product. Note that the example is for a DEM of a mountainous area. That information and the intended use of the image for display purposes sets the LAS modules' parameters to create an optimal product. These are set automatically by Prolog control and set parm clauses that use predetermined values provided by an expert.

CONCLUSION

The evaluation of the expert system as a front end to the LAS is not complete, but some insights have been obtained. The expert-system shell and a knowledge base are not difficult to build, perhaps because the domain of the expert system is intended to support only a subset of the LAS functionality. The expert-system shell could have been implemented in Fortran or C, but Prolog, which is very easy to learn, seems better suited for nonalgorithmic applications, such as the LAS front end. Each knowledge base can be built by a domain expert who has a minimum understanding of Prolog or assistance from the knowledge engineer.

PLANS AND RECOMMENDATIONS

Thus far, the expert system has been used only for demonstration. An implementation of a production-level expert system for determining satellite-data processing based on the customer's description of intended data use is being planned. It is expected that this knowledge base will fit very easily within the LAS front-end expert-system shell. The success we have experienced with the moderate use of the prototype expert system has lead us to expect operational success when applied in a production environment.

LAS Front End .. Version 1.0

Knowledge base (Type filename or
'c' to use current): shade

Current level is set to Novice

PRE-CONSULTATION OPTIONS

Set consultation level (s)
See possible products (p)
Start a consultation (c)
Exit from the program (e)

Option? c

State what product or procedure is desired: shade

Help may be requested with (?) or (help)

Do you desire a shaded-relief image product
that is used primarily for display purposes.--> (d)isplay
or a shaded-relief image product that you
want to use as a base map upon which other data
can be later overlain. -----> (b)asemap
or an image product in which you have control
over the amount of shadowing, Sun location and
accentuation of relief. -----> (s)pecial? d

Is the area you want to create a shaded-relief for, one that is
(m)ountainous, (h)illy, or relatively (f)lat ? m

Is your digital topo data online? n

Do you desire to retain your raw digital topo data online? n

Is the digital topo data tape on the computer floor? y

What is the tape description or tape number? AA123

Is your digital topo data on a dal transfer tape-> (t)

or on a dem formatted tape-----> (d)? d

What name do you wish to give to your resulting shaded
relief image? Shade

END OF CONSULTATION OPTIONS

Figure 2. Expert-system interaction with the user and the "shade" knowledge base.

```
Display commands      (d)
Save commands to file (s)
Restart consultation  (r)
End this program      (e)
```

```
Option? d
```

```
fmount +
  drive=MF: density=6250 tapeid=AA123
```

```
dement +
  out=temp.dem skip=0 size=33000 dtype=dem +
  drive=MF: print=term
```

```
picshade-create +
  in=temp.dem out=Shade method=1 pixsize=(30,30) +
  mask=0 isf=1 angle=50 azimuth=315 +
  expnt=1 osf=1 offset=0 odtype=1
```

```
lasdel +
  name=temp.* verify=no
```

Figure 2. Continued.

REFERENCES

- Clocksinn, W. F., and Mellish, C. S., 1984, *Programming in Prolog*: New York, Springer Verlag, 293 p.
- Feigenbaum, E. A., and McCorduck, P., 1983, *The fifth generation*: Reading, Mass., Addison-Wesley Publishing Company, p. 76-80.
- Fischler, M. A., and Firschein O., 1986, *Intelligence and the computer: AI Expert*, v. 1, no. 3, p. 47.
- Quirk, B. K., and Oleson, L. R., Overview of the Land Analysis System (LAS), *in Pecora Symposium*, 11th, Sioux Falls, S. Dak., 1987, *Proceedings: American Society of Photogrammetry and Remote Sensing*, p. 133-148.
- Rolandi, W. G., 1986, *Knowledge engineering in practice: AI Expert*, v. 1, no. 3, p. 47.

CHAPTER B

A Study of Four Raster-Based Data Generalization Procedures

By MARGARET G. MAYERS, EUGENE A. FOSNIGHT, and
JAMES A. STURDEVANT

U.S. GEOLOGICAL SURVEY BULLETIN 1841

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES

CONTENTS

Abstract	B1
Introduction	B1
Generalization strategies	B1
Area filtering	B2
Category aggregation and area filtering	B3
Systematic resampling	B3
Modal search and replace	B3
Results	B5
Evaluation of generalization techniques	B5
Area filtering	B5
Category aggregation and area filtering	B6
Modal search and replace	B6
Systematic resampling	B7
Conclusion	B7
References	B10

FIGURES

1–5. Maps showing:

1. Composition of ungeneralized U.S. Geological Survey interim land cover test data **B3**
2. Area filtering procedure **B4**
3. Modal search and replace and systematic resampling generalization strategies **B6**
4. Effects of three generalization strategies on the original land cover data set **B10**
5. Effects of area filtering on the original land cover data set **B10**

TABLES

1. U.S. Geological Survey land cover classification for Alaska **B2**
2. Statistical measurements **B5**
3. Area filtering statistical results **B6**
4. Probability for nonomission and noncommission for area filtered categories **B7**
5. Category aggregation and area filtering statistical results **B7**
6. Probability of nonomission and noncommission for aggregated area filtered categories **B8**
7. Modal search and replace statistical results **B9**
8. Probability of nonomission and noncommission for replaced categories **B9**
9. Systematic resampling statistical results **B9**
10. Probability of nonomission and noncommission for resampled categories **B9**

A Study of Four Raster-Based Data Generalization Procedures

By Margaret G. Mayers,¹ Eugene A. Fosnight,¹ and James A. Sturdevant

Abstract

The transfer of complex land cover data generated from digitally processed Landsat grid cell data to a vector-based geographic information system can require a drastic reduction in data complexity. In developing effective procedures for reduction of grid cell data sets, the following generalization strategies were investigated: (1) area filtering, (2) category aggregation, (3) systematic resampling, and (4) modal search and replace.

Generalization of the land cover data is justified for two reasons—reduction of the processing time required for analysis and simplification of the land cover map. The ultimate goal is to reduce the volume of land cover data while maintaining the integrity of the land cover information. The quality of the land cover information after the application of a generalization strategy was measured by omission and commission errors, percent unchanged, Cohen's statistic of agreement, and the reduction of land cover regions.

Although the four generalization strategies significantly reduced the number of gridded land cover regions, we found that aggregating the categories gave the most region reduction. However, the gain in region reduction was offset by a loss in information about the subcategories. Area filtering alone and in combination with category aggregation resulted in the fewest category-specific errors relative to the amount of overall region reduction. Consideration of pixel contiguity for each land cover category by the area filtering algorithm produced the greatest reduction in land cover regions with the least amount of overall image and category-specific error.

INTRODUCTION

Land cover data produced from digitally classified Landsat multispectral scanner data can produce an overwhelming volume of gridded homogeneous and spatially contiguous areas that will be referred to

¹TGS Technology, Inc., Earth Resources Observation Systems Data Center, Sioux Falls, SD 57189. Work performed under U.S. Geological Survey contract 14-08-001-22521.

Any use of trade names and trademarks in this publication is for descriptive purposes only and does not constitute endorsement by the U.S. Geological Survey.

as regions. Each region is converted to a polygon when incorporated into a vector-based geographic information system (GIS); for example, the Alaska interim land cover data set of the U.S. Geological Survey (USGS) Valdez 1:250,000-scale quadrangle contains approximately 98,000 land cover regions. Although this and similar data sets provide a viable source of land cover information for land managers, their conversion to vector data structures is needed for use in some systems and applications. Because of the size and the complexity of these data, a method is needed for generalizing them before conversion to vector form.

The USGS Alaska interim land cover categories are a combination of the Anderson (Anderson and others, 1976) Level I and II map classification system and the Viereck and Dryness (1980) vegetation classification system (Shasby and Carneggie, 1986). The USGS Alaska interim land cover categories and their definitions are listed in table 1.

The Valdez quadrangle contains 98,000 land cover regions, which, by itself, challenges the limitations of most vector GIS software and requires excessive file space and processing time. Also, regional applications of interim land cover data require that the data be generalized to simplify the land cover maps for broad land management applications.

GENERALIZATION STRATEGIES

After much consideration, the following generalization strategies were chosen: (1) area filtering, (2) category aggregation, (3) systematic resampling, and (4) modal search and replace. These strategies were selected on the basis of the range of techniques typically used for simplification. A 1,185- by 875-pixel subscene extracted from the Valdez interim land cover data was selected for testing purposes (fig. 1). This study was conducted by the Earth Resources Observation Systems Data Center by using the raster-

Table 1. U.S. Geological Survey land cover classification for Alaska

[Adapted from Shasby and Carneggie (1986)]

Level I	Level II
I. Forest (forest canopy cover of one-third or more)	A. Needleleaf Forest (over two-thirds of tree cover contributed by needleleaf species) B. Broadleaf Forest (over two-thirds of tree cover contributed by broadleaf species) C. Mixed Forest (broadleaf or needleleaf species contribute one-third to two-thirds of the tree cover)
II. Shrub (forest canopy cover of less than one-third and shrub canopy cover of one-third or more)	A. Shrublands B. Dwarf Shrublands and Related Communities (rarely exceeding 50 centimeters in height)
III. Herbaceous (vegetation with 5 percent or more of vascular and nonvascular (mosses and lichens) cover and less than one-third cover of woody plants)	A. Dry or Moist (seasonally or systematically submerged) B. Wet (sites are typically large ponds in the lower elevations) C. Very Wet Aquatic (sites are typically large ponds in the lower elevations) D. Mosses E. Lichens
IV. Agriculture	
V. Urban	
VI. Barren	A. Sparse Vegetation B. Nonvegetated (rock, soil)
VII. Water	A. Deep Clear Water (water bodies typically deeper than 1 meter that have low turbidity) B. Turbid Shallow Water (water bodies typically less than 1 meter deep or that have significant amounts of suspended sediments)
VIII. Ice and Snow (includes snowfields and glaciers)	
IX. Shadow (areas in rugged terrain in which incident solar illumination is too low to permit the discrimination of vegetation types)	
X. Clouds	

based Land Analysis System, which runs on a Digital Equipment Corporation VAX 11/780 computer.

Area Filtering

Area filtering may be defined as the reassignment of homogeneous regions that do not contribute to the information needed to solve problems as defined by the user. Area filtering is done for several very different purposes. First, a classified image may be filtered to reduce the number of pixels thought to

contain error. Second, a classified image may be assumed to be correct, and filtering may be used for data reduction. Third, area filtering may be used to create polygons of a desired minimum map size. The research described in this chapter was directed by the second purpose, and, because of its exploratory nature, several different minimum threshold areas were selected for filtering. Impacts on the land cover categories and on the total number of regions were studied.

The area filtering functions used in this project were based on the segmentation procedure developed by Rosenfeld (1978) and implemented by Nichols (1981). First, regions were identified and labeled with unique numbers in a process called region labeling. Once identified as a unique region, the area of each region was calculated. Areas smaller than a given minimum threshold value (sub-minimum regions) were replaced by the most frequently occurring nonminimum land use category that abutted each of these subminimum regions (fig. 2).

The minimum threshold area value may be altered for the entire classification system or for individual land cover categories. Also, a pairwise weight table may be applied to the process. It consists of a list of every possible category pair and an associated numeric weight that is used to control reclassification of a flagged pixel. The table also may cause filtering to be unidirectional or bidirectional. Unidirectional filtering allows a specific category to be replaced in the smoothing process by another category, but the reverse replacement is not allowed. Unidirectional filtering may be used to protect small, but significant, classes. Bidirectional filtering gives the same weighting for a pair of categories in either direction. Seven different minimum threshold sizes, ranging from 0.5 to 128.0 hectares, were selected for use in area filtering of the land cover data.

Unidirectional area filtering tests were conducted to reduce the number of regions while restricting the increase of the shadow category. The same seven minimum threshold sizes listed above were used; however, the pairwise weight table was manipulated to create the unidirectional filter. Shadow was allowed to be reclassified to any other land cover category, but other land cover categories were not allowed to be changed to shadow. All shadow-to-category pairs were assigned a value of 1 in the pairwise weight table, and all category-to-shadow pairs were assigned a value of 0 in the pairwise weight table.

As each area filtering test was completed, each filtered image was compared to the original unfiltered Valdez land cover test area data set. A pixel

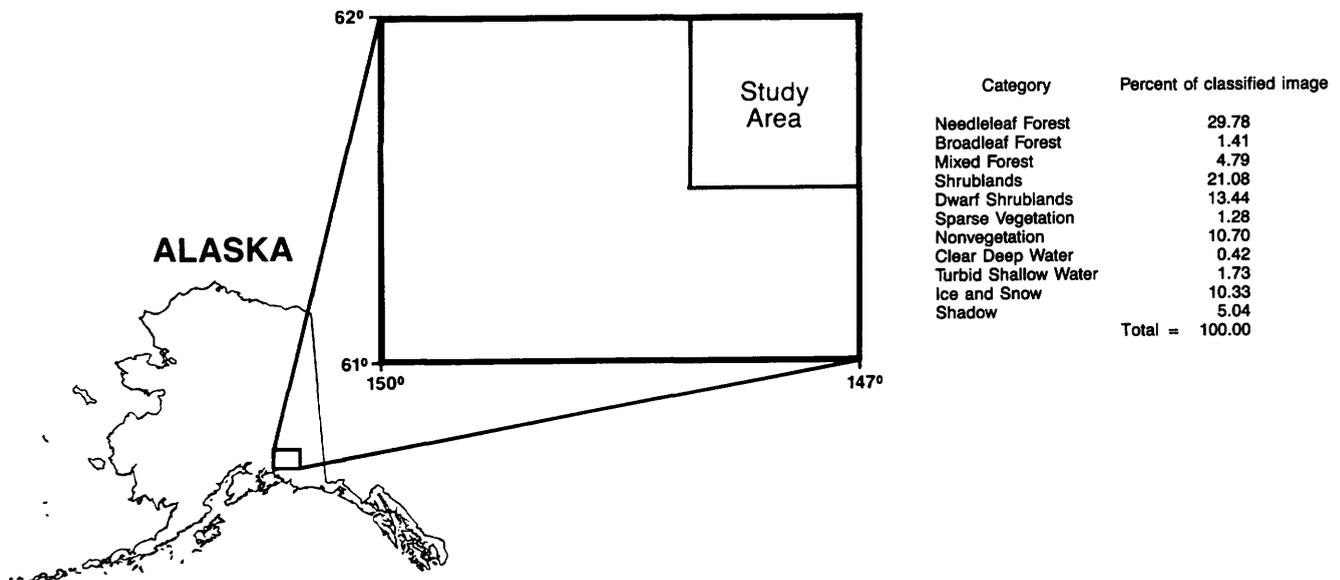


Figure 1. Composition of ungeneralized U.S. Geological Survey interim land cover test data.

count and a contingency table were generated to study the overall image and individual category changes. Finally, the new reduced number of regions was tallied and compared with the original number of land cover regions. The following comparisons were made between each area filtered image and the original unfiltered image (table 2): (1) percent unchanged, (2) errors of omission, (3) errors of commission, (4) KHAT statistic of agreement (Cohen, 1960), and (5) a fragmentation index.

The statistical measurements chosen for data analysis were used to study the overall image and individual category change. Percent unchanged represents the overall percentage of similarity between the generalized and the ungeneralized images based totally upon the diagonal elements in a contingency table. Errors of omission indicate the probability that a pixel in the original image was omitted in the generalized image. Errors of commission indicate the probability of a pixel incorrectly added to the image (Bishop and others, 1975). Cohen's KHAT statistic of agreement represents a percentage agreement with the original image and compensates for the probability of chance agreement (Cohen, 1960). Finally, a fragmentation index was calculated to determine the complexity present within a generalized land use map (Monmonier, 1974).

Category Aggregation and Area Filtering

In the second generalization strategy, category aggregation, all Level II categories present in the Valdez test area were aggregated to the more general

Level I category, and the area filtering procedure was repeated. The aggregated, area filtered image results were compared to the unfiltered Level I image, rather than to the original land cover data. The five statistical measurements (table 2) were calculated on the resultant images.

Systematic Resampling

The third generalization strategy, systematic resampling, was accomplished by imposing larger area grids over the land cover cells, then selecting the central land cover cell within each imposed grid, and assigning that cell value to the larger grid cell (fig. 3). The imposed grid sizes used in the resampling tests were 150, 250, and 350 meters. The resultant images were compared to the original land cover image by using the tests in table 2.

Modal Search and Replace

The final generalization strategy was a neighborhood modal search and replace. For a given input pixel radius, each pixel is used as the center of a modal tabulation. The most frequently occurring land cover category within the specified radius becomes the grid cell value (fig. 3). The radii used for the cartographic neighborhood function in this study were 1, 2, and 3 pixels. The resultant images were compared to the original land cover image.

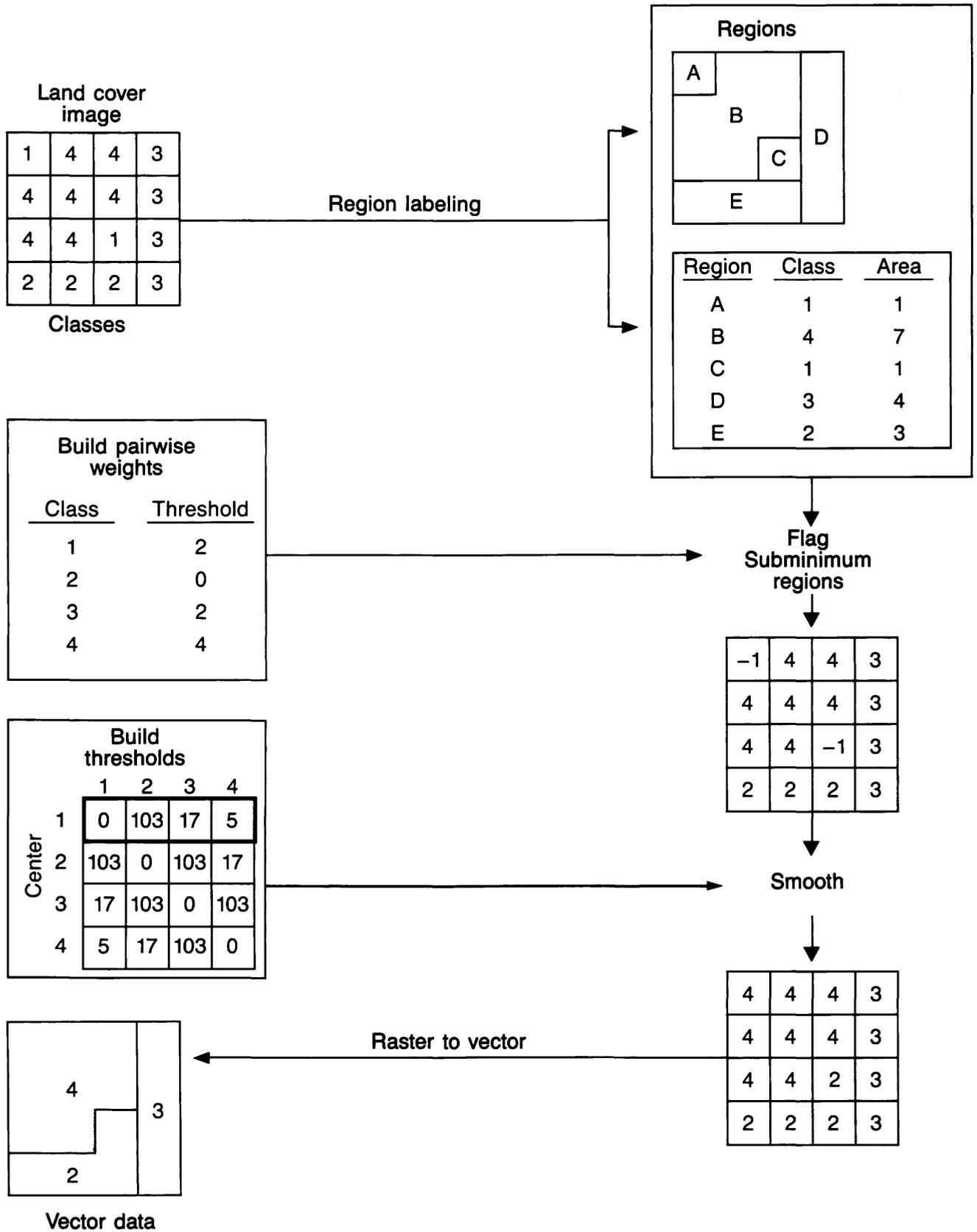


Figure 2. Area filtering procedure.

Table 2. Statistical measurements

-
1. Percent unchanged = $\Sigma X_{ij}/N$,
where X_{ij} = the number of elements in a particular generalized category that correctly represent the original data.

 N = the total number of elements.
 2. Probability of nonomission = $(1 - (X_{ij}/X_i)) \times 100$,
where X_{ij} = the number of elements in a particular generalized category that correctly represent the original data.

 X_i = the total number of elements in a particular category in the original image.
 3. Probability of noncommission = $(1 - (X_{ij}/X_j)) \times 100$,
where X_{ij} = the number of elements in a particular generalized category that correctly represent the original data.

 X_j = the total number of elements in a particular category in the generalized image.
 4. Cohen's KHAT statistic = $((N \cdot \Sigma X_{ij}) - X_i + X + i) / (N - \Sigma(X_i + X + i))$,
where X_{ij} = the total number of elements in a contingency table.

 X_{ij} = the number of elements in a particular generalized category that correctly represent the original data.

 $(X_i + X + i)$ = the product of each category row and column.
 5. Fragmentation index (F) = $((m-1)/(n-1))/F_s$,
where m = the number of regions or parcels.

 X_{ij} = the number of regions or parcels.

 F_s = $(m-1)/(n-1)$ Monmonier's fragmentation index for the initial, ungeneralized image.
-

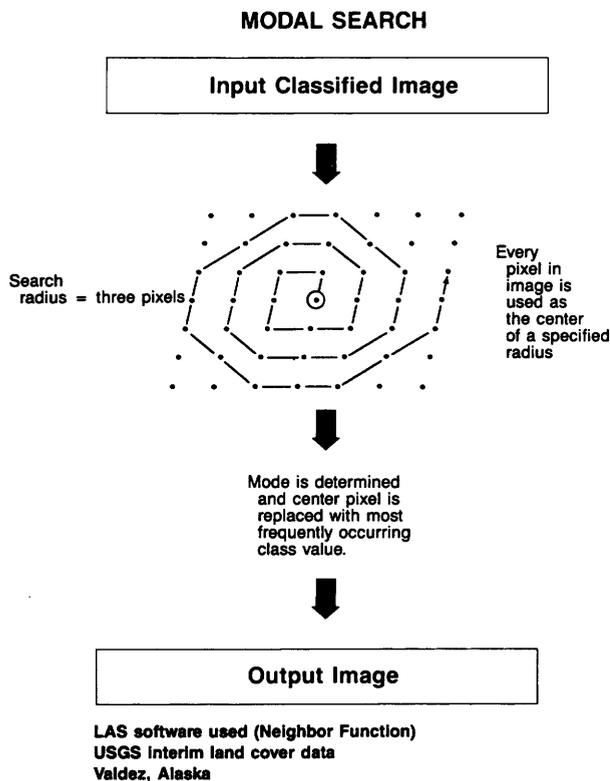
RESULTS

Area Filtering

Evaluation of Generalization Techniques

Each technique required different inputs, such as minimum threshold size, imposed grid size, and search radius; thus, the results were not directly comparable. However, the strategies were evaluated by comparing the relative reduction in region numbers of each technique and the category-specific errors associated with the region reduction.

Overall agreement between the original data and area filtered data was high, and category-specific errors were low relative to the number of regions reduced (table 3). The implementation of the unidirectional shadow filter during area filtering eliminated commission errors for the shadow category and decreased the presence of shadow by as much as 3 percent but resulted in a slight increase in the number of regions. Generally, omission and commission errors increased as minimum threshold size increased



Algorithm for 150-meter grid (Each pixel = 50 meters on one side)

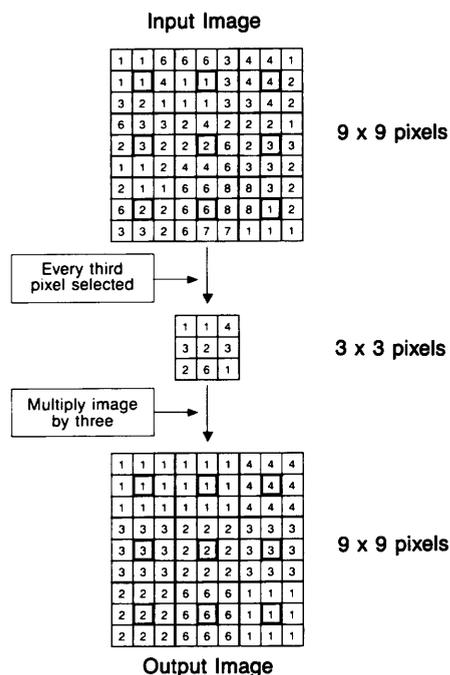


Figure 3. Modal search and replace and systematic resampling generalization strategies.

Table 3. Area filtering statistical results

	Minimum threshold size (pixels)						
	2	16	68	100	128	256	512
Number of regions ¹	7138	1916	466	310	249	121	59
Percent reduction ²	23.1	79.3	95.0	96.7	97.3	98.7	99.4
Percent unchanged	99.7	95.3	88.5	86.7	85.5	81.6	77.5
KHAT (percent)	99.6	94.3	85.9	83.6	82.1	77.3	71.3
Fragmentation	0.77	0.21	0.05	0.03	0.03	0.01	0.01

¹ The number of land cover regions before generalization = 9278.

² Percent reduction in number of regions is based on comparisons with the original, ungeneralized image.

(table 4). Filtering out lone pixels (minimum threshold size of 2) resulted in approximately 2,000 regions (23-percent reduction), which left all category probabilities in the 99 percentile.

Increasing minimum threshold size had varying effects on the individual categories. “Needleleaf Forest” and “Ice and Snow” categories remained in the 85 percentile of nonomission and noncommission up to the 128-pixel range (table 4). Conversely, “Broadleaf Forest” and “Sparse Vegetation” categories had small nonomission and noncommission probabilities relative to the minimum threshold size. These results indicate the need for further work in altering the minimum thresholds by category.

Category Aggregation and Area Filtering

Reduction in the number of regions from simple category aggregation alone was approximately 42 percent. This 42-percent reduction was achieved by having no category omission or commission error; however, a loss in category precision, or level of detail, was experienced. Simple category aggregation and area filtering with a 68-pixel-area threshold were equally effective in reducing the number of regions (table 5). Thus, this technique provided better results than area filtering only if the minimum threshold size was smaller than 68 pixels and Level II categories were not needed. As with simple area filtering, aggregated area filtered categories experienced an increase in omission and commission errors as threshold size increased (table 6). Further region reduction accompanied by little increase in error might be accomplished by increasing the minimum threshold size for the low-error categories and by decreasing the minimum threshold size for the high-error categories.

Modal Search and Replace

Region reduction resulting from modal search and replace using a 3-by-3 search area (radius = 1)

Table 4. Probability of nonomission and noncommission for area filtered categories
[See table 1 for explanation of abbreviations]

Nonomission							
CATEGORY	Minimum threshold area (pixels)						
	2	16	68	100	128	256	512
NF	99.81	97.41	94.03	92.97	92.39	90.96	90.10
BF	98.58	80.80	58.44	48.02	46.76	24.26	0.00
MF	98.66	79.58	51.39	44.13	39.54	27.85	15.36
SL	99.84	97.63	94.12	93.40	93.03	90.30	87.80
DS	99.78	96.63	91.71	90.75	90.35	87.08	80.85
SV	97.92	65.48	24.03	15.97	8.01	0.00	0.00
NV	99.62	93.68	82.69	80.04	79.02	73.44	68.82
DW	99.43	88.69	69.62	60.12	52.04	45.57	19.82
SW	99.88	97.89	94.78	94.43	93.44	92.05	91.21
IS	99.88	97.89	94.78	94.43	93.44	92.05	91.21
SD	99.66	94.28	82.77	76.51	70.64	57.59	40.80

Noncommission							
CATEGORY	Minimum threshold area (pixels)						
	2	16	68	100	128	256	512
NF	99.71	95.85	89.99	88.31	87.56	84.44	79.55
BF	99.70	94.00	82.46	78.63	77.63	74.35	0.00
MF	99.43	91.74	79.24	78.40	77.92	77.74	78.77
SL	99.74	95.95	90.00	88.36	86.95	82.91	76.91
DS	99.64	94.43	84.47	81.19	78.95	73.29	69.83
SV	99.40	91.03	79.55	75.57	72.82	0.00	0.00
NV	99.63	94.05	86.23	86.23	80.74	74.70	69.68
DW	99.57	97.46	96.55	97.24	99.86	99.84	99.64
SW	99.66	94.64	85.24	81.68	80.69	76.37	75.01
IS	99.92	98.50	96.59	96.04	96.60	95.56	93.27
SD	99.56	92.47	81.91	80.61	79.55	73.07	73.47

Table 5. Category aggregation and area filtering statistical results

	Minimum threshold size (pixels)						
	2	16	68	100	128	256	512
Number of regions ¹	4050	1222	322	219	177	84	41
Percent reduction ²	21.2	76.3	93.5	95.7	96.6	98.4	99.2
Percent unchanged	99.8	97.5	93.2	91.9	91.1	88.4	85.6
KHAT (percent)	99.8	96.5	90.6	88.7	87.6	83.9	79.8
Fragmentation	0.79	0.24	0.06	0.04	0.03	0.02	0.01

¹The number of land cover regions before generalization = 5141.

² Percent reduction in number of regions is based on comparisons with the aggregated, ungeneralized image.

was similar to that achieved by simple category aggregation. However, KHAT and percent unchanged statistics were lower, and individual category error was higher for modal search and replace results (table 7). Errors of commission for modal search had a different trend among the categories compared to area filtering. The "Mixed Forest" and, to some extent, the "Broadleaf Forest" categories had high commission errors in the modal search and replace technique but low commission errors in the area filtering results (table 8). This difference was due to the large circumference relative to the area of "Mixed Forest" region

boundary that was smoothed out in the modal search strategy.

Systematic Resampling

Of all the generalization strategies examined in this study, resampling produced the least desirable results. Although the simplicity of this technique allowed the algorithm to be implemented quickly, overall agreement between the original data and the low category-specific probabilities was poor. As expected, the number of regions decreased as grid cell sizes increased (table 9). However, relatively low KHAT and percent unchanged values, as well as low probabilities of nonomission and noncommission, were obtained (table 10).

CONCLUSIONS

Of the strategies tested, simple category aggregation to a lower category precision resulted in the highest reduction in the number of homogeneous regions. Area filtering that was applied to the results of category aggregation resulted in an image of very

Table 6. Probability of nonomission and noncommission for aggregated area filtered categories

[See table 1 for explanation of abbreviations]

Nonomission							
CATEGORY	Minimum threshold area (pixels)						
	2	16	68	100	128	256	512
FT	99.86	98.03	94.67	93.44	93.19	91.35	89.60
SH	99.91	98.46	95.76	95.10	94.92	93.44	91.14
BN	99.66	94.14	85.05	82.62	81.22	75.64	72.69
WT	99.82	96.89	91.60	89.31	87.76	86.66	79.28
IS	99.88	97.89	94.78	94.42	93.42	92.03	91.14
SD	99.66	94.27	82.49	76.10	70.29	57.44	40.58

Noncommission							
CATEGORY	Minimum threshold area (pixels)						
	2	16	68	100	128	256	512
FT	99.86	97.73	93.78	92.87	92.16	91.07	88.74
SH	99.87	97.83	93.64	91.85	91.02	87.32	84.05
BN	99.74	96.06	89.83	87.77	86.12	81.24	76.00
WT	99.66	96.16	89.49	86.43	85.38	80.67	78.20
IS	99.94	98.68	96.98	98.57	96.99	96.25	94.22
SD	99.66	94.69	87.64	85.94	85.05	79.56	78.29

high overall agreement with the original data and in little individual category error; however, the data did experience a loss in category precision (fig. 4).

Of the three strategies conducted by using unaggregated categories, area filtering resulted in very high overall agreement with the original data and in little individual category error. This was derived from the algorithm's consideration of contiguous pixels for reassignment based on a minimum threshold area. The data generalized by using the modal search and replace strategy had lower KHAT and category-specific agreement with the unfiltered data. Relatively poor results were caused by the small search radius and the loss of contiguously occurring pixels because of nearby domination of a larger homogeneous land cover category (fig. 5).

The strategy that produced the most undesirable results was the systematic resampling technique. It produced data that were based largely upon the individual category's total percent of an image. Overall agreement values were low, and individual category errors were high. The most significant problem with this algorithm is the loss of natural category boundaries when the multiplication factor is applied during this technique (fig. 4).

The results of area filtering tests point to the need for additional research in the 2- to 68-pixel

minimum threshold sizes and for alteration of these threshold sizes by category. The optimum generalization strategy can be selected only on the basis of a specific purpose for the particular land use-land cover map. Although this research was exploratory in nature and was not directed toward a particular use of land cover maps, the results have indicated a general superiority of area filtering over the other techniques tested. Also, the resulting category error patterns from the generalization process have provided a basis for continuing research in filtering techniques (fig. 5).

REFERENCES

- Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E., 1976, A land use and land cover classification system for use with remote sensor data: U.S. Geological Survey Professional Paper 964, 28 p.
- Bishop, Y. M., Fienberg, S. E., and Holland, P. W., 1975, Discrete multivariate analysis: theory and practice: Cambridge, Mass., The MIT Press, p. 395-398.
- Cohen, J., 1960, A coefficient of agreement of nominal scales: Educational and Psychological Measurement, v. 20, no. 1, p. 37-46.
- Earth Resources Observation Systems Data Center, 1986, LAS user's manual: Sioux Falls, S. Dak., 53 p.

Table 7. Modal search and replace statistical results

	Search radius (pixels)		
	1	2	3
Number of regions ¹	5414	3333	2340
Percent reduction	47.1	64.1	74.8
Percent unchanged	96.4	92.5	89.3
KHAT (percent)	95.6	90.9	86.9
Fragmentation	0.58	0.36	0.25

¹The number of land cover regions before generalization = 9278.

Table 8. Probability of nonomission and noncommission for replaced categories
[See table 1 for explanation of abbreviations]

CATEGORY	Nonomission		
	Search radius (pixels)		
	1	2	3
NF	98.14	95.84	94.00
BF	86.41	72.00	60.97
MF	87.04	72.84	61.65
SL	97.69	95.41	93.44
DS	96.87	93.86	91.15
SV	80.62	60.85	44.90
NV	94.63	88.48	83.07
DW	93.17	84.53	77.13
SW	98.32	96.47	95.30
IS	98.22	96.35	94.77
SD	95.03	89.99	85.35

CATEGORY	Noncommission		
	Search radius (pixels)		
	1	2	3
NF	96.41	93.23	90.58
BF	93.25	85.83	78.88
MF	92.19	84.32	76.93
SL	97.27	93.89	91.04
DS	96.08	91.66	87.89
SV	92.92	84.19	75.64
NV	95.53	90.31	86.00
DW	99.13	96.27	95.05
SW	97.38	92.77	89.11
IS	98.47	96.83	95.43
SD	96.41	90.96	86.18

Table 9. Systematic resampling statistical results

	Imposed grid size (pixels)		
	3	5	7
Number of regions ¹	6359	3425	2093
Percent reduction	31.5	63.1	77.4
Percent unchanged	84.0	75.3	70.4
KHAT (percent)	80.5	70.0	63.9
Fragmentation	0.69	0.37	0.23

¹The number of land cover regions before generalization = 9278.

Table 10. Probability of nonomission and noncommission for resampled categories
[See table 1 for explanation of abbreviations]

CATEGORY	Nonomission		
	Imposed grid size (pixels)		
	3	5	7
NF	88.56	82.69	79.55
BF	60.45	42.56	34.88
MF	61.59	44.64	36.52
SL	88.04	80.96	76.30
DS	83.65	73.83	68.60
SV	51.28	31.19	23.59
NV	77.06	64.78	57.10
DW	76.38	64.14	51.90
SW	87.60	78.92	73.10
IS	92.00	87.60	84.89
SD	74.96	59.02	49.14

CATEGORY	Noncommission		
	Imposed grid size (pixels)		
	3	5	7
NF	88.61	82.64	79.25
BF	60.95	42.01	34.04
MF	61.37	43.98	36.33
SL	88.08	81.18	76.40
DS	83.31	74.15	67.35
SV	51.08	30.26	24.02
NV	76.54	64.14	57.85
DW	76.71	61.64	51.59
SW	87.43	79.95	74.15
IS	92.16	87.10	85.01
SD	75.17	59.18	48.75

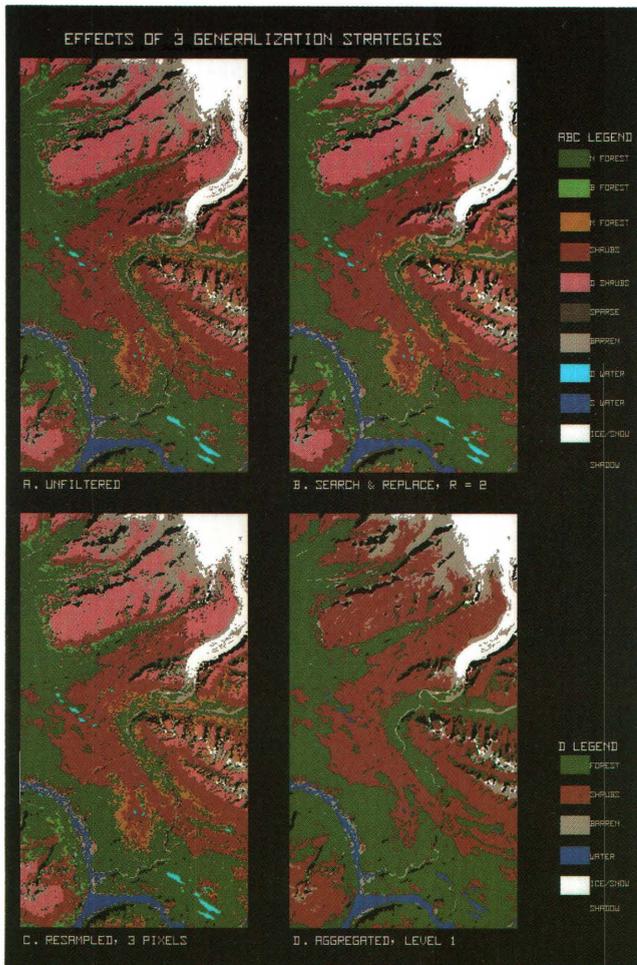


Figure 4. Effects of three generalization strategies on the original land cover data set.

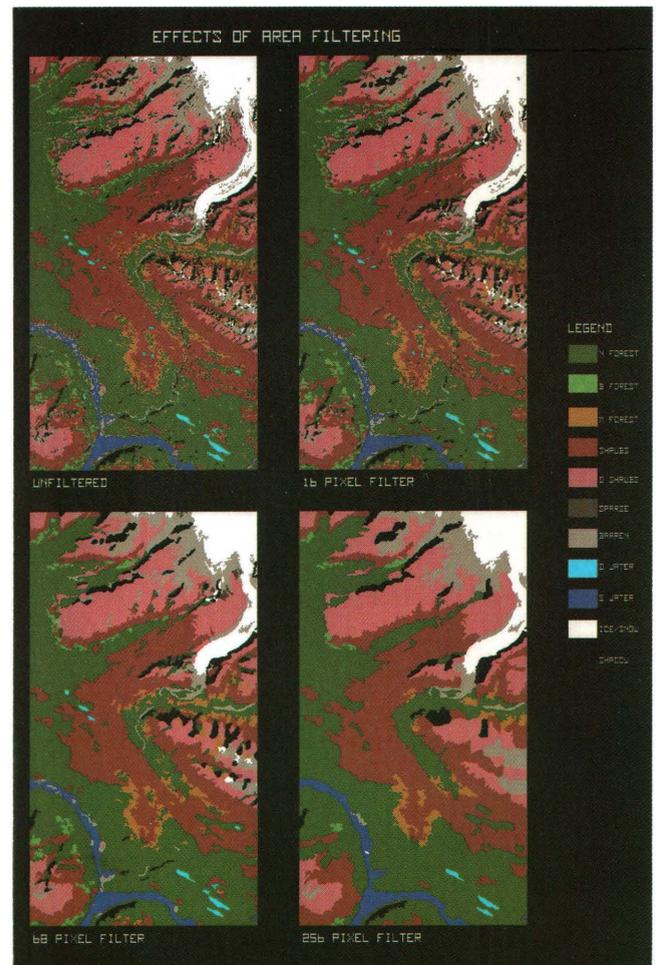


Figure 5. Effects of area filtering on the original land cover data set.

- Monmonier, M. S., 1974, Measures of pattern complexity for choropleth maps: *American Cartographer*, v. 1, no. 2, p. 159–169.
- Nichols, D. A., 1981, Conversion of raster coded images to polygonal data structures, in William T. Pecora Memorial Symposium, 7th, Sioux Falls, S. Dak., 1981, Proceedings: Falls Church, Virginia, American Society for Photogrammetry, p. 508–515.
- Rosenfeld, A., 1978, Extraction of topological information from digital images, in International Advanced Symposium on Topological Data Structures for Geographic

- Information Systems, 1st, Cambridge, Mass., 1977, Proceedings: Cambridge, Harvard University Press, v. 6, p. 790–805.
- Shasby, M. B., and Carneggie, D. M., 1986, Vegetation and terrain mapping using Landsat MSS and digital terrain data: *Photogrammetric Engineering and Remote Sensing*, v. 52, no. 6, p. 779–786.
- Viereck, L. A., and Dryness, C. T., 1980, A preliminary classification system for vegetation of Alaska: Forest Service General Technical Report PNW-106, U.S. Department of Agriculture, 36 p.

CHAPTER C

Digital Image Processing of Landsat Image Map Products at the U.S. Geological Survey

By RICHARD D. KIDWELL, DOUGLAS R. BINNIE, and
SHEILA E. MARTIN

U.S. GEOLOGICAL SURVEY BULLETIN 1841

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES

CONTENTS

Abstract	C1
Introduction	C1
Processing procedures	C2
Data acquisition and evaluation	C2
Digital mosaicking procedure	C2
Halftone aimpoints	C3
Photomechanical film mosaicking procedure	C4
Graphic arts electronic scanning procedures	C6
Conclusion	C8
References	C9

FIGURES

1. Predigital and postdigital mosaic procedure diagram C3
2. Density curve transformation of Landsat digital brightness values to final photographic color proof C4
3. Halftone reproduction curves of the Washington, D.C., and vicinity Thematic Mapper Landsat image map C5
4. Commission International d'Eclairage diagram showing printing color gamut limitation compared to electronic and visual color displays C5
5. Photomechanical mosaic procedure C6
6. Procedure used to calibrate the enlargement scale factor for x and y dimensions and for the halftone cyan density scale aimpoints C7
7. Graphic arts electronic scanner procedure C8

Digital Image Processing of Landsat Image Map Products at the U.S. Geological Survey

By Richard D. Kidwell, Douglas R. Binnie,¹ and Sheila E. Martin

Abstract

For over a decade, the U.S. Geological Survey has produced multicolor satellite image maps from Landsat imagery. Image map research has progressed from the development of photographic techniques for mosaicking of Landsat scenes to the refinement of computer techniques for large-area mosaicking of Landsat digital data and subsequent extraction of up to twelve 1:250,000-scale quadrangle areas for lithographic map production.

To maintain consistently high quality satellite image map production, the U.S. Geological Survey has developed standard technical instructions for all procedures, including digital processing of Landsat image map data, photographic and electronic production of publication-scale halftone image separates, and selection of density aimpoints for lithographic platemaking and printing.

INTRODUCTION

Image maps from Landsat multispectral scanner (MSS) and Thematic Mapper (TM) data are produced by the U.S. Geological Survey (USGS) using a combination of digital, photographic, and lithographic procedures. Since 1972, Landsat image mapping has progressed from simple procedures that create a single map from a single Landsat scene to sophisticated computer techniques for digital mosaicking of large blocks of Landsat data and automated extraction of quadrangle-formatted areas. Typically, the imagery is recorded on color film directly from the digital data. A graphic arts color scanner is used to produce enlarged, color-corrected halftone separations for lithographic four-color process printing. A photomechanical film mosaicking procedure, developed in 1975, continues to be used for selected projects (Dahlberg, 1987).

¹ Earth Resources Observation Systems Data Center, Sioux Falls, SD 57189.

Any use of trade names and trademarks in this publication is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

Landsat MSS and TM data used for image mapping are evaluated and processed at the Earth Resources Observation Systems (EROS) Data Center in Sioux Falls, South Dakota. Graphic arts color scanner and plotting work is contracted out to commercial companies. Photomechanical processing, map design and layout, and printing are done at the USGS Headquarters in Reston, Virginia.

The digital mosaicking process uses a 9.5- by 9.5-inch color film transparency produced on a MacDonald Dettwiler Color Fire 240 recorder. The color film is electronically scanned, enlarged, and halftoned using a graphic arts scanner-plotter for four-color process printing (cyan, magenta, yellow, and black). Recent production samples from the Color Fire 240 and laser beam recorder (LBR) imagery are Denali National Park and Preserve, Alaska (Color Fire 240), and Anchorage, Alaska (LBR) (U.S. Geological Survey, 1986). Future plans include production of four-color halftone separations at publication scale using a Scitex plotter; these would be made directly from computer-compatible tapes that were digitally processed at the EROS Data Center. Additional plans include the digital merging of MSS (approximately 75 meters) and (or) TM data (30 meters) with System Probatoire d'Observation de la Terre (SPOT) panchromatic data (10 meters) to produce an image map that has enhanced image detail (Welch and Ehlers, 1987). Graphic displays in various forms and scales of the Phoenix, Arizona, area were published to demonstrate this procedure (U.S. Geological Survey, 1987).

The photomechanical process uses spectral bands 3, 4, and 5 from MSS data for preparation of the yellow, magenta, and cyan film separations. Comparable bands from TM and SPOT data are available for similar film separations. The digital image data for each spectral band are recorded in negative and positive form on 9.5- by 9.5-inch black-and-white film by an LBR. The LBR film transparencies of single scenes are mosaicked photomechanically and

exposed through a halftone screen for three-color process printing. The most recent example is the Illinois State Geological Survey 1:500,000-scale map of Illinois (Illinois Department of Energy and Natural Resources, 1985).

Halftone separates produced by either digital or photomechanical mosaicking method are registered to lettering separation, and a DuPont Cromalin color prepress proof is prepared. The color proof, which is controlled to match the press printing, is used to evaluate the initial press sheets for registration, completeness of linework, and color balance.

PROCESSING PROCEDURES

Data Acquisition and Evaluation

The first step in an image map project is a thorough search of acquired satellite data at the EROS Data Center to identify the best possible digital source. The digital data are purchased from the National Oceanic and Atmospheric Administration (NOAA) archive, and working copies of the digital data are made. If the data are not in the NOAA archive, then an order is placed by NOAA to acquire the data.

A project plan is prepared, outlining the steps necessary to complete the image map. Specific variable output parameters include

- Output projection and projection parameters,
- Output pixel size,
- Resampling algorithm,
- Digital enhancements,
- Output scale, and
- Working and final product map coordinates.

After photographic film transparencies are made from the acquired digital tapes for all bands of each scene that will be used in the project, the imagery is assessed. Any bad data lines or missing pixels are compensated for by substituting data from an adjacent spectral band. This is a cosmetic correction that does not affect map accuracy as pixels are not displaced.

Digital Mosaicking Procedure

A small-scale paper mosaic is prepared from black-and-white prints from the film transparencies. This photomosaic is used to check for adequate coverage and to identify the placement of the mosaic cut lines.

The digital image data are registered to a geodetic reference system from USGS topographic maps

derived from control points. A 1:24,000-scale map is registered on a digitizer in conjunction with a full-resolution subscene of the digital data to identify map coordinates and corresponding line-sample locations for a series of control points. Transformation coefficients are calculated from these control points to rectify the image data and to generate the final registered image. Control points may be selected from maps at other scales if 1:24,000-scale maps are not available.

The radiometric aimpoint for the midtone should be the Landsat data median brightness value. The endpoints used in determining the median should be values containing useful data. Large bodies of water, clouds, cloud shadow, or other shadow should not be used when determining the median value. Image tonal match is accomplished between scenes most often by multipoint linear mapping, which is based on matching cumulative percentages from the overlap area in the two adjacent scene histograms.

Scenes for the quadrangle area are digitally mosaicked by using one of the following two methods. The scenes may be mosaicked with the image 100 geometric registration system connected to a Digital PDP 11/60 computer when fewer than five scenes are involved in the quadrangle area. The Large Area Mosaicking Software on the VAX 11/780 system will handle more than five scenes. However, the maximum number depends on available disk storage space (fig. 1).

Radiometric differences are balanced and images abutted as a part of the mosaicking step. Carefully chosen cuts are used to mask the mosaic join and to eliminate undesirable conditions, such as cloud or snow. Radiometric matching is accomplished by using a multipoint linear mapping that standardizes the brightness value distribution of the component images. If multiple quadrangles are represented in the digital data, then each quadrangle is extracted after the mosaicking and enhancement process.

The formatted digital mosaic data are used to produce final small-scale (9.5- by 9.5-inch) black-and-white film transparencies on the LBR. The transparencies are used to produce a color photographic proof (fig. 2) and to prepare the final-scale color separates for lithography. The photographic proof is used by the EROS Data Center to evaluate the color balance, density, and contrast of the LBR output product. A black-and-white continuous-tone contact print is prepared from the small-scale film separate planned for final printing in cyan. The cyan separation controls the image map contrast. Halftone aimpoints are selected based on pleasing appearance of contrast in the black-and-white cyan separation. Adja-

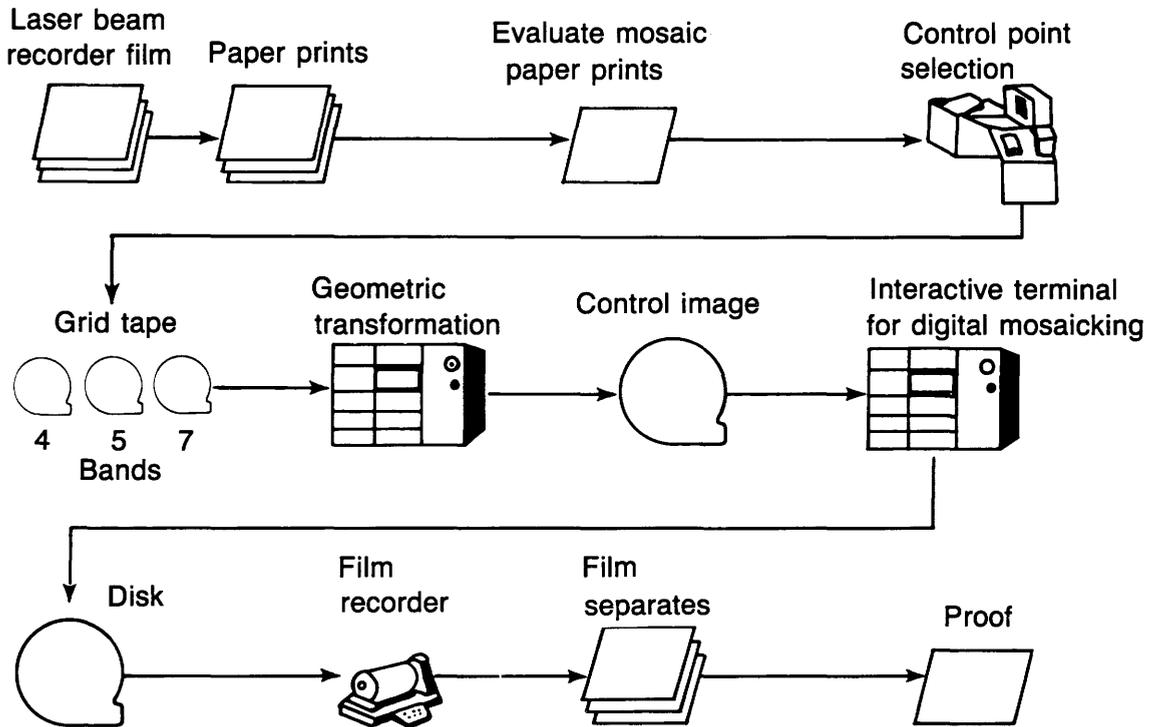


Figure 1. Predigital and postdigital mosaic procedure diagram.

cent maps can be made to match radiometrically by controlling the midtone halftone value between scenes (fig. 2).

Halftone Aimpoints

The EROS Data Center exposes a 16-step density control scale on each LBR separation that has a 0.10 density increase between steps (0–1.60 density range for scale). In graphic arts, five aimpoints are selected for stretching the continuous-tone imagery to fit the appropriate representation for the imagery. One aimpoint is assigned a 3- to 5-percent dot value in an area of interest in the highlight imagery. This is step 2 or 3 (about 0.20–0.30 transmission density) on the 16-step linear control scale exposed on each LBR transparency. The selection of step 2 or 3 depends upon the desired contrast in the imagery. Reduced steps between highlight and quarter-tone steps increase contrast. The shadow aimpoint, step 14 (density about 1.35), is selected for a 90- to 97-percent dot, and the midpoint aimpoint, step 8 (density about 0.80), is assigned a 45- to 65-percent dot.

Quarter tones at 25 and 75 percent vary with the images, but, in general, they are assigned closer to endpoints than to the midpoint to improve contrast in the highlight and shadow areas. The cyan separation

is evaluated first because it carries the shape and contrast of the features in the imagery. If the cyan halftone of the black-and-white film separation looks good as viewed on white paper (not backlighted), then the final printed map is likely to look good also. Yellow and magenta separations, which generally follow the cyan tone curve, drop 10 percent at the midtone and 5 percent in the quarter tones. Reductions in the yellow and magenta tone curves are made to correct for lithographic ink deficiencies. If the ink pigments were pure, then equal densities of ink in each of the separations would combine to make tints of gray (fig. 3; Kidwell and McSweeney, 1985).

Adjoining map sheets generally will be visually acceptable if the adjoining density values for the cyan separation can be made to fit between the quarter tone to the midtone points and the midtone to the three-quarter tone points. Endpoints for highlight and shadow will be forced to match in reproduction by the offset printing halftone process as highlight is limited by color of paper and darkest shadow by maximum ink density.

Electronic monitor interactive color gamut displays and photographic color dyes contain a wider range of colors than are possible by using process inks of yellow, magenta, and cyan. The approximate differences are shown in the Commission International d'Eclairage diagram (fig. 4). The color hue and

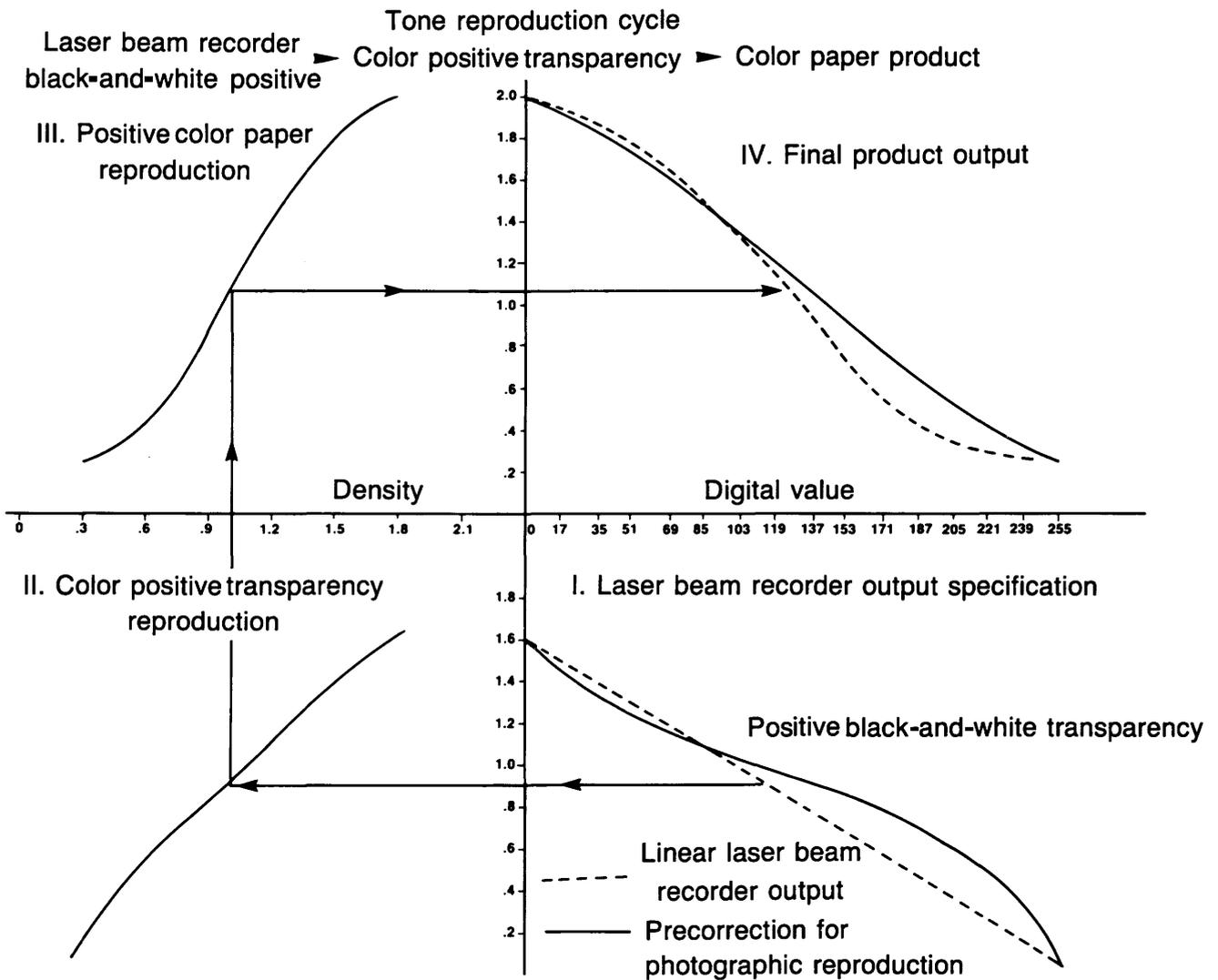


Figure 2. Density curve transformation of Landsat digital brightness values to final photographic color proof that approximates the color rendition desired in the lithographic product.

saturation attained on an electronic color display or a photographic color proof can be made to closely approximate the final lithographic print by linearizing the output of the color photographic proof described in the digital mosaic procedure (fig. 1).

Photomechanical Film Mosaicking Procedure

Laser beam recorder film negative images and paper positive prints of MSS bands 4, 5, and 7 or TM bands 1, 2, or 3 and 4, 5, or 7 of each Landsat scene to be mosaicked are evaluated for adequate coverage, uniformity in image detail, and range of contrast before geometric correction and film mosaicking. These three-band combinations from either MSS or TM data are printed in yellow, magenta, and cyan in

the band order listed for the typical false-color Landsat image map product. Geometric corrections are determined from large-scale (usually 1:24,000) maps as described for digital mosaicking.

In the photomechanical film mosaicking procedure, negative film scenes are assembled on clear stable plastic carrier sheets. A maximum of four sheets are needed. To control a film mosaic, two types of control bases generally are required. A tie-point control base is plotted from selected points on line maps at 1:63,000 scale or larger, and a line map drainage separate is used as a supplemental control base. By using the control bases, a scene for the cyan film separation is oriented to fit the control points and the drainage pattern and is taped on all corners to a carrier sheet punch-registered to the control.

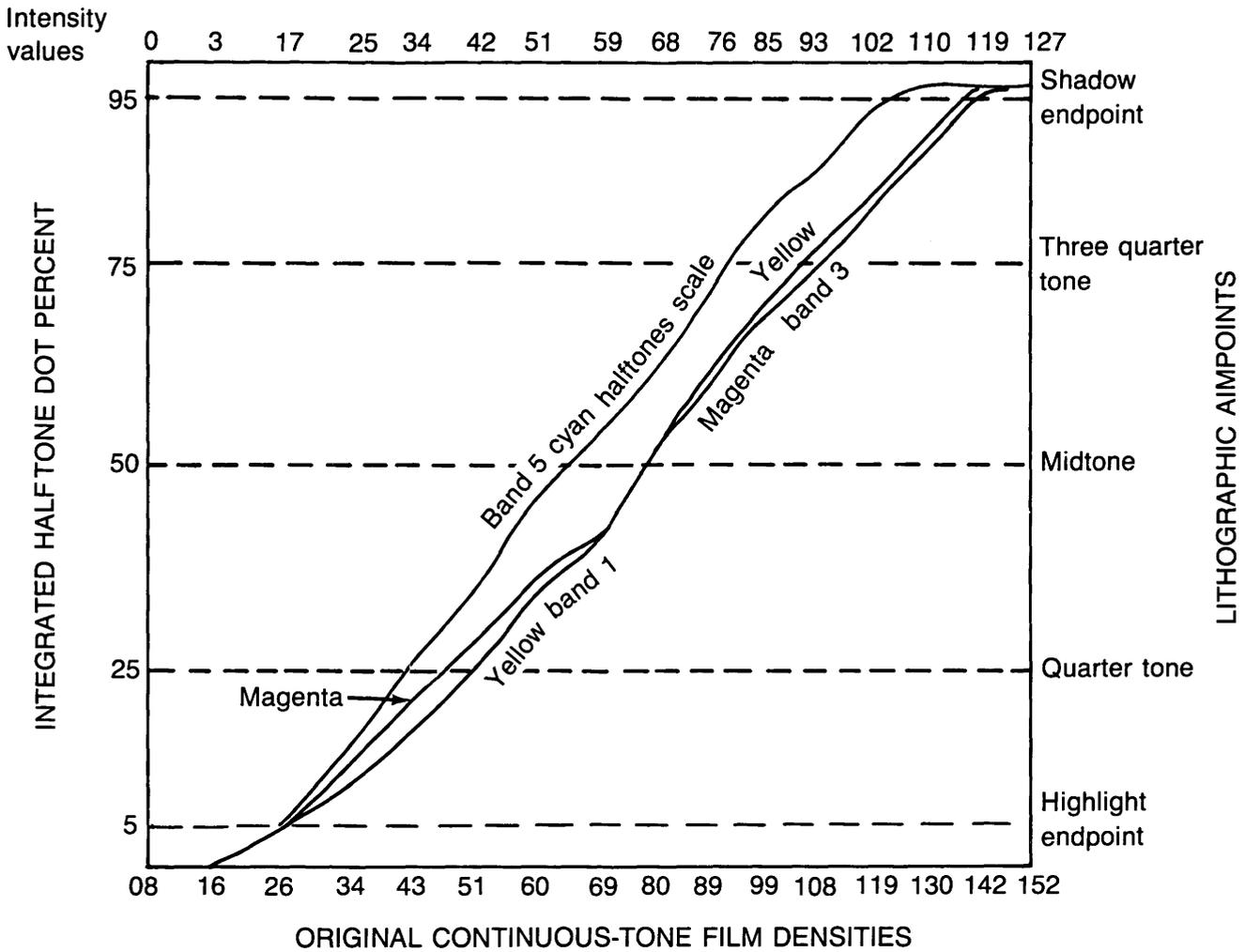
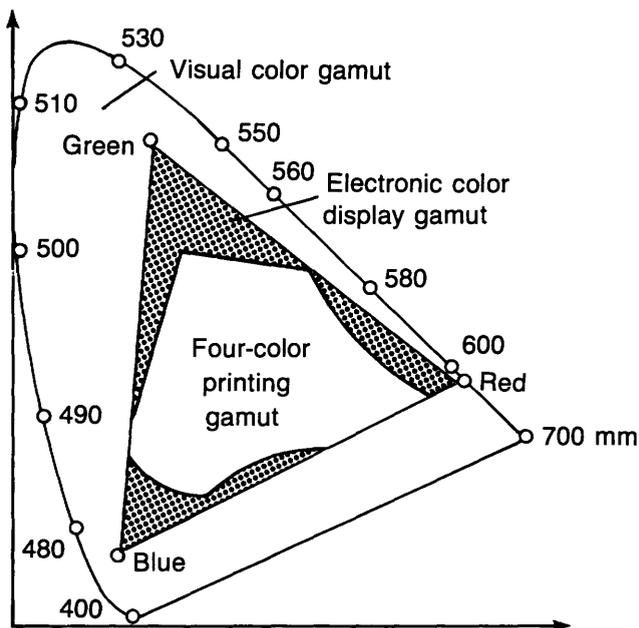


Figure 3. Halftone reproduction curves of the Washington, D.C., and vicinity TM Landsat image map (U.S. Geological Survey, 1984).



Other scenes for the same band are positioned in a checkerboard pattern on which no two scenes overlap on the same carrier. This procedure is continued until all scenes are registered to the control on all four carrier sheets.

A single-image mosaic of the cyan separation film positive becomes the control base for assembling the other bands by using the same technique. Masks are prepared to expose selected areas of each scene mosaicked. The masks are prepared by using peel-coat material photographically exposed and chemically etched from scribed join lines. A sheet of scribecoat material is overlaid on the stack of scene carrier sheets with the nonemulsion side up.

Figure 4. Commission International d'Eclairage diagram showing printing color gamut limitations compared to electronic and visual color displays (adapted from Williamson, 1983).

Join lines are drawn in the overlap areas where the images will blend together in a photographic step. Usually, 0.003-inch-wide lines are scribed on the emulsion side of the scribecoat. Four peelcoats are exposed from the scribecoat, and windows are opened in the peelcoat to expose each scene on each corresponding carrier sheet. All remaining lines on the peelcoat masks are opaqued. A diffuser sheet is placed between the light source and the vacuum frame during exposure of a sheet of unexposed continuous-tone film so as to blend the images in the 0.003-inch overlap. The same masks used to make the film positive of the cyan mosaic are used for compositing the yellow and magenta film separation positives.

Each mosaicked band normally is enlarged by using a copy camera and is exposed through a halftone screen to produce halftone image film separates for printing. The halftone image separates are punch registered to line and lettering separates, and a prepress four-color proof is prepared as a quality-control guide for printing (fig. 5). Alternatively, the enlargement and halftoning of film separations may use a graphic arts scanner-plotter.

Graphic Arts Electronic Scanning Procedures

Electronic scanning and digital enlargement of Landsat imagery results in significantly improved quality compared to photographic processing. The sensitometric conversion from photographic to lithographic reproduction, enlargement in the x and y dimensions, and generation of the halftone dot structure can be controlled precisely by using an electronic

scanner. Landsat MSS halftone separates can be produced at scales as large as 1:250,000. The same procedures apply to the use of TM imagery for electronic scanning and the production of halftone image separates at scales as large as 1:100,000.

The basic electronic scanning purpose on a graphic arts color drum scanner-plotter is to analyze a spot of light about 0.002-inch in diameter either reflected or transmitted by the material being scanned. The light is split into four paths. Photomultiplier tubes that have red, green, and blue separation filters create cyan, magenta, and yellow signals, respectively. The fourth light path activates another photomultiplier that is used for edge enhancement. The scanner stores one line at a time.

Each spot of light passing through the four photomultipliers is analyzed for color correction, gray value, halftone dot shape, enhancement, and size change. The analyzed data are stored during the first revolution of the linked scanner and plotter drums. The stored data are output signals from the computer and are used to make sequential exposures on film during the second revolution of the plotting drum, when a second line from the scanning drum is stored. The laser exposing light is split into six paths to produce halftone dots as small as 0.0006 inch.

The use of a Color Fire 240 transparency on the scanner-plotter allows the extraction of a black separate from the three color signals along with the yellow, magenta, and cyan. The cyan aimpoint for step 14 is altered to an 80-percent dot to adjust for the black separate. The scanner-generated black separate is 4 percent at step 8, 31 percent at step 11, and 59 percent at step 14. Color contamination in the transparency is removed from six color-control scales

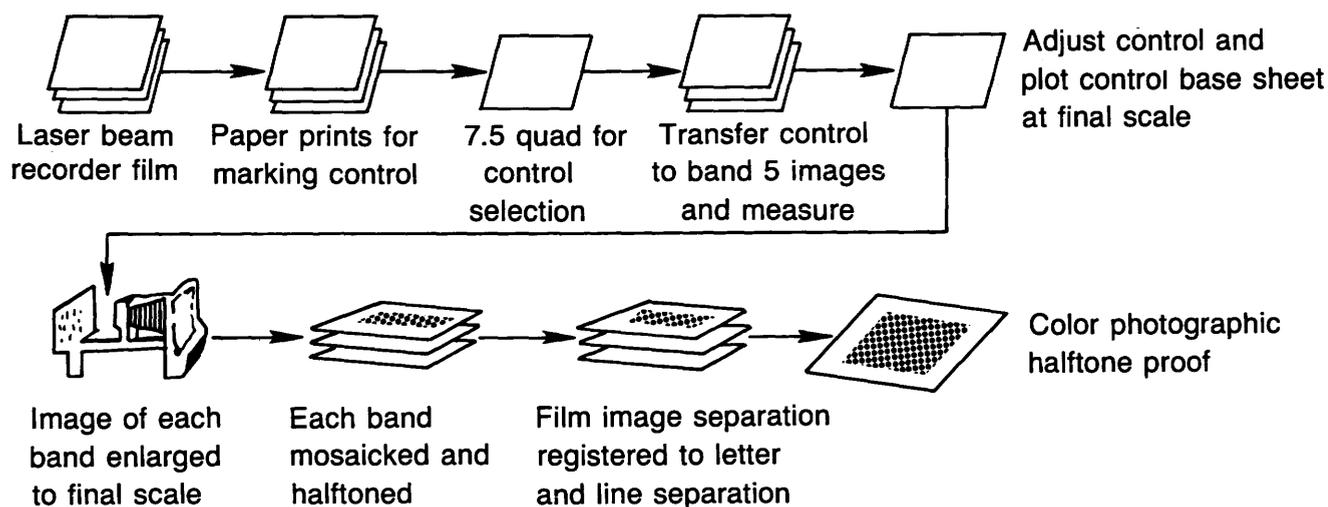


Figure 5. Photomechanical mosaic procedure (Warren, 1980).

(green, cyan, blue, magenta, red, and yellow) by use of the scanner's hue and saturation controls to achieve clean color. Additional edge enhancement by the scanner operator is used for visual improvement of feature sharpness without causing a "blocky" appearance.

Landsat imagery on black-and-white film is scanned after setting a dial for cyan, magenta, and yellow for MSS bands 7, 5, and 4, respectively. Each band is scanned and plotted on film at 1 inch per minute along the x axis to produce 175-line-per-inch halftones.

Before scanning images that have unique color requirements, a lithographic control color chart is visually compared with the photographic color proof, and density aimpoints are selected for production of halftones; aimpoints are labeled on the continuous-tone density scale of the black-and-white print of band 7. The aimpoints are similar to those shown in figure 3. The normal human eye is excellent for

direct side-by-side color comparison. Sensitometers are used for process calibrations and film-press control targets.

A film positive of the map projection is prepared at publication scale for the scanner operator to use for image registration and enlargement. The scanner operator also is supplied with LBR film separates of bands 4, 5, and 7 that have been digitally mosaicked in quadrangle format, a publication-scale photographic color proof, x-y scale factors for each band, and a band 7 print (cyan separation) that has density aimpoints labeled.

The scanner is calibrated by using densitometric controls for enlargement and for density aimpoints on the cyan separation (see fig. 6). The cyan half-tone positive is plotted at publication scale by scanning the band 7 film positive.

The magenta and yellow halftones are scanned to plot in gray balance with the cyan half-tone and have a 30-degree half-tone pattern rotation angle

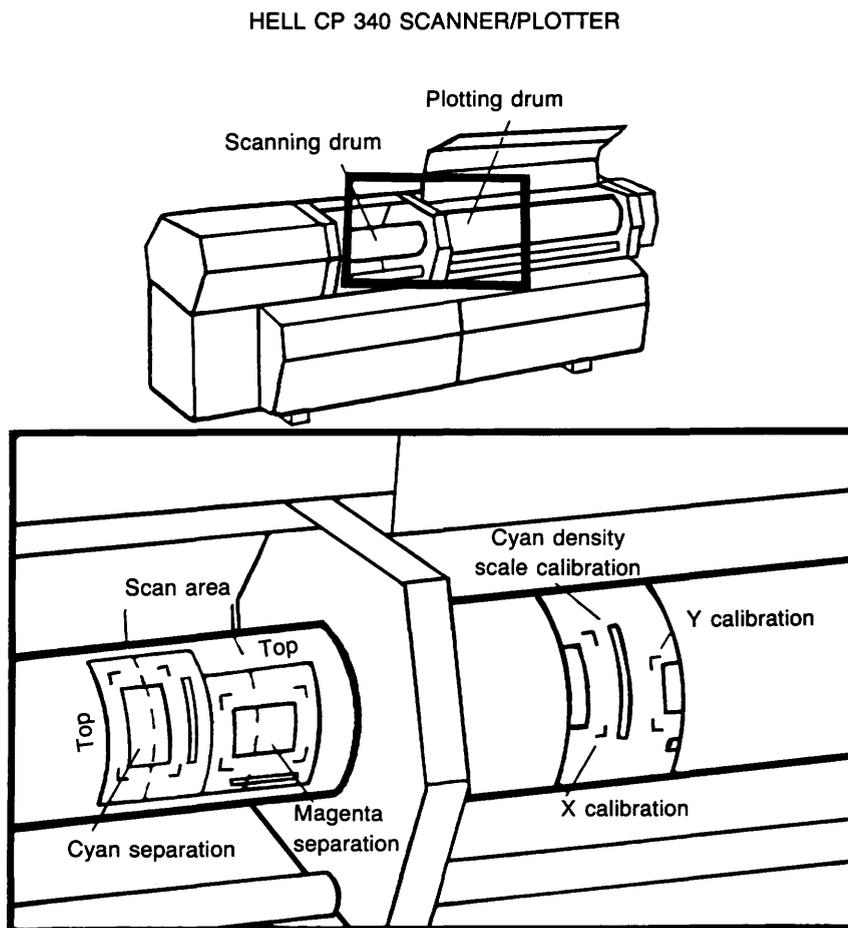


Figure 6. Procedure used to calibrate the enlargement scale factor for x and y dimensions and for cyan density scale aimpoints by scanning a few inches of the cyan and magenta separations for plotting.

between each separation to eliminate interference patterns. A four-color-half-tone proof is made from the final-scale scanner half-tone positives as a quality-control color guide for printing the final image map. Control targets on the proof are read by using a densitometer to meet printing standards.

The scanner-supplied half-tone positives are punch-registered with line-lettering separates for preparation of half-tone image negatives used for negative working pressplates. During the photographic conversion from positive to negative film, the half-tone image negatives are sharpened (dot size reduced) by 5 to 10 percent in the midtone to compensate for press dot gain. A second four-color proof from the half-tone negatives is made to be used as a guide to check registration of imagery and linework (fig. 7). The proof made from the first-generation scanner film will represent the final printing.

Consistency in the printing of image maps is achieved by standardizing the materials and procedures that affect the image detail and the color balance of the final product. Densitometric standards are established by printing control targets using the pressplates, inks, pre-press proofs, and paper to be used for printing (Kidwell and others, 1983). The printing standard provides aimpoints and limitations for image processing. The control targets used in developing printing standards are digitally processed with the map imagery to provide control from digital processing through lithographic printing.

CONCLUSION

Satellite image map research at the USGS has progressed from the development of photographic techniques for preparation of Landsat mosaics to the refinement of computer techniques for large-area

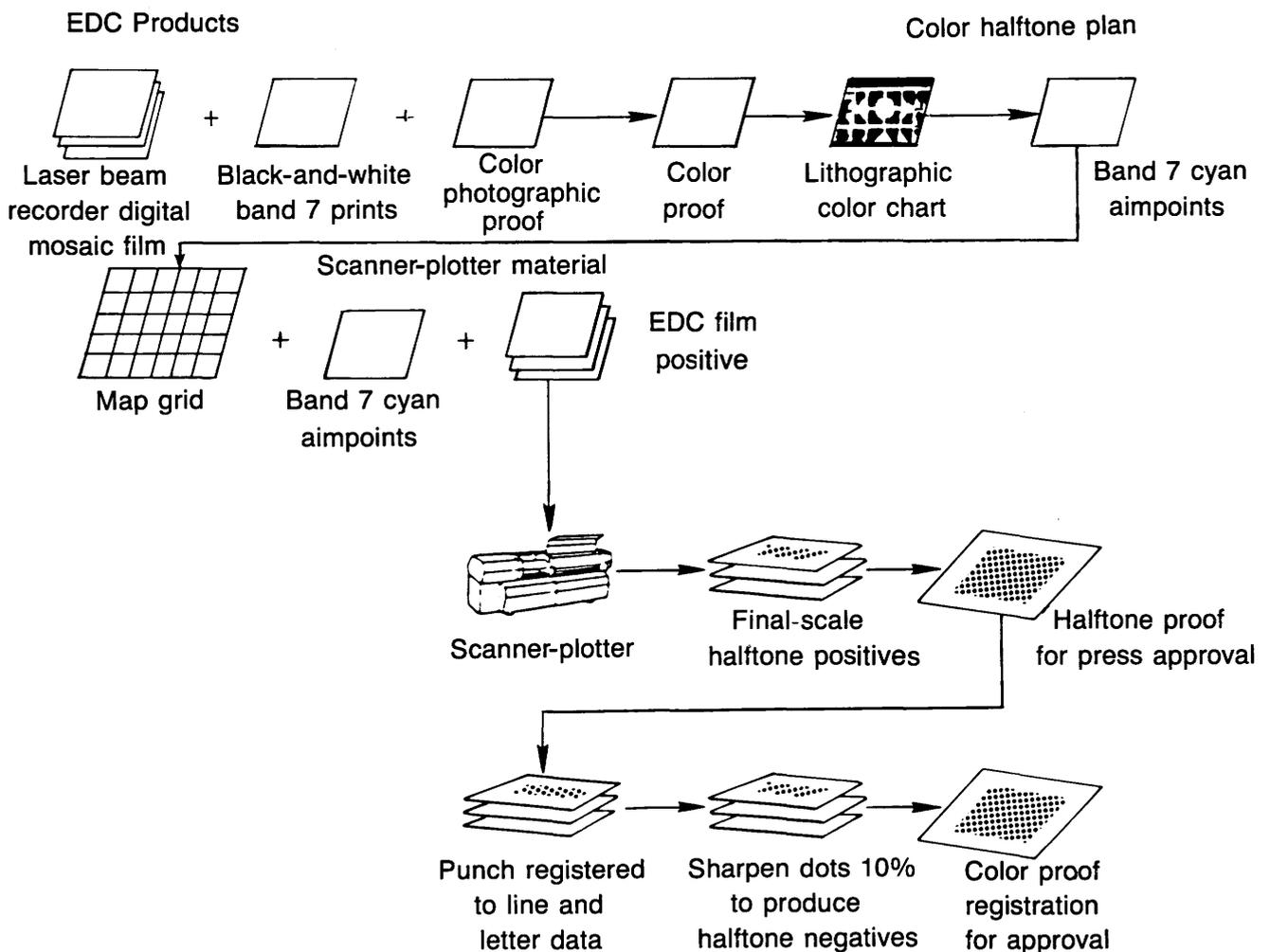


Figure 7. Graphic arts electronic scanner procedure.

mosaicking of Landsat film digital data and subsequent extraction of quadrangle-formatted mosaics for lithographic production. Techniques developed for the electronic production of publication-scale halftone image separates and selection of optimum density aimpoints for lithographic printing have significantly improved image quality, color balance, and repeatability in map production. Technical instructions have been developed to cover current procedures for digital image processing, photographic and electronic production of halftone separates, and lithographic printing (U.S. Geological Survey, 1987). Future plans include development of techniques for the production of publication-scale halftone image separates directly from Landsat digital data and the production of image maps from digitally merged Landsat and SPOT data.

REFERENCES

- Dahlberg, R., Luman, D., Warren, A., Stohr, C., 1987, A satellite image mosaic of Illinois: Falls Church, Va., Proceedings, American Society of Photogrammetry and Remote Sensing, p. 91–101.
- Illinois Department of Energy and Natural Resources, 1985, Landsat TM satellite image map of Illinois: Illinois State Geological Survey, 1:500,000 scale.
- Kidwell, R. D., and McSweeney, J. A., 1985, Art and science of image maps: Falls Church, Va., Proceedings, American Society of Photogrammetry, p. 771–781.
- Kidwell, R. D., McSweeney, J. A., Warren, A., Zang, G., and Vickers, E., 1983, Experiments in lithography from remote sensor imagery: Falls Church, Va., Proceedings, American Society of Photogrammetry, p. 384–393.
- U.S. Geological Survey, 1984, Landsat TM Washington, D.C., and vicinity image map: U.S. Geological Survey, 1:100,000 scale.
- 1986, Landsat MSS Anchorage, Alaska, image map: U.S. Geological Survey, 1:250,000 scale.
- 1986, Landsat MSS Denali National Park and Preserve, Alaska image map: U.S. Geological Survey, 1:250,000 scale.
- 1987, Landsat and Spot Image Display Phoenix, Arizona, area: Geological Survey Miscellaneous Investigation Map I-1741.
- 1987, Procedure manual for preparation of satellite image maps: U.S. Geological Survey Open-File Report 86-19, 144 p.
- Warren, A., 1980, Photomechanical mosaicking, *in* Manual of Photogrammetry, (4th ed.): Falls Church, Va., American Society of Photogrammetry, p. 769–775.
- Welch, R., and Ehlers, Manfred, 1987, Merging multiresolution SPOT HRV and Landsat TM data: Athens, University of Georgia, Photogrammetric Engineering and Remote Sensing, p. 301–303.
- Williamson, S. J., 1983, Light and color in nature and art: New York, John Wiley and Sons, p. 488.

CHAPTER D

Connecting Local- and Wide-Area Networks for Distributed Access to a Microcomputer Application

By RICHARD A. HOLLWAY and DENISE A. WILTSHIRE

U.S. GEOLOGICAL SURVEY BULLETIN 1841

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES

CONTENTS

Abstract	D1
Introduction	D1
Purpose and scope	D2
Acknowledgments	D2
Approach	D2
Results and discussion	D4
Summary and conclusion	D6
References	D6

FIGURES

1. Configuration of the local-area network showing heterogeneous components **D3**
2. Detailed configuration of Computer Systems Information Network local-area network **D5**

TABLE

1. Information-retrieval systems accessed by the Computer Systems Information Network **D2**



Connecting Local- and Wide-Area Networks for Distributed Access to a Microcomputer Application

By Richard A. Hollway and Denise A. Wiltshire

Abstract

In 1983, the U.S. Geological Survey formed a Bibliographic Special Interest Group to guide the development and the implementation of a system for accessing computerized information data bases. The objectives of the Bibliographic Special Interest Group were twofold—to provide a gateway that connects to multiple and distributed information systems and to provide access to the gateway system by way of a private telecommunications network for a potential user community of nearly 2,000 professionals conducting investigations in the earth sciences.

After exhaustive software evaluations, the first objective was met by the development of a customized gateway system, known as the Computer Systems Information Network. Although a growing number of offices within the Survey were procuring microcomputers, many did not own the specific hardware for the Computer Systems Information Network software package. Hence, a need arose to develop a method to remotely access the microcomputers on which Computer Systems Information Network software was available. The solution was to connect several microcomputers to a local-area network and to provide access to the local-area network through the Survey's wide-area network, Geonet, which is available to most potential Computer Systems Information Network users. The purpose of this paper is to describe the method of access and the problems that arose during implementation.

INTRODUCTION

The evolution of distributed data processing at the U.S. Geological Survey (USGS) during the early 1980's resulted in a need for special interest groups (SIG's) to guide the development of software applications. A Bibliographic SIG was established to design and implement an intermediary front-end system for conducting computerized bibliographic searches of public and commercial data bases. The Bibliographic

Use of trade names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

SIG was charged with providing a gateway that connects multiple and distributed information-retrieval systems and providing access to the gateway system by way of a customized telecommunications network for a potential user community of nearly 2,000 scientific and technical professionals.

The SIG met its first objective by conducting exhaustive software evaluations and by designing and implementing a customized gateway system, known as the Computer Systems Information Network (CSIN) (Wiltshire and Hollway, 1987, p. 483–492). CSIN provides gateway access to seven information-retrieval systems containing bibliographic and numeric data (table 1). The CSIN software is written in "C" programming language for a microcomputer that runs the Microsoft Disk Operating System (MS-DOS) version 3.0 or greater.

The implementation of CSIN for a microcomputer environment eliminated direct access to the system by potential users who would not have the required hardware. Because most of the potential users had access to an existing nationwide telecommunications network, it was proposed that a cluster of microcomputers be connected to a wide-area network commonly known in the Survey as Geonet. Geonet, which is a private telecommunications subnetwork of the public telecommunications network Tymnet, links computer hardware in Survey offices in the 50 States and Puerto Rico. A variety of terminals and microcomputers are connected to minicomputers in each of these offices. The minicomputers serve as primary nodes on Geonet. Cost considerations prohibited developing CSIN for a minicomputer environment.

At the time that the Bibliographic SIG was designing the CSIN project, the Survey also was conducting a project to determine suitable applications for a recently procured local-area network (LAN). Representatives from the Bibliographic SIG initiated a feasibility study to evaluate the effectiveness of connecting several CSIN microcomputers to a

Table 1. Information-retrieval systems accessed by Computer Systems Information Network

Name of system	Acronym	Description of system
Bibliographic Retrieval Service	BRS	Multidisciplinary bibliographic information.
Chemical Information System	CIS	Toxicity data.
Dialog Information Services	DIALOG	Multidisciplinary bibliographic information.
National Library of Medicine	NLM	Medical, toxicity, and chemical data.
Occupational Health Services, Inc.	OHS	Hazardous substances data.
Pergammon ORBIT InfoLine	ORBIT	Multidisciplinary bibliographic information.
Scientific and Technical Network	STN	Bibliographic and numeric data.

LAN located at the Survey Headquarters at Reston, Virginia (Angier and Hoehl, 1986, p. 23–26). This LAN would be connected to Geonet, thus serving as a network link for clients at remote sites to access CSIN.

Purpose and Scope

This report describes results of a study conducted from June 1985 through April 1987 to test the feasibility of accessing microcomputers that were running CSIN software connected to a LAN by way of a wide-area network of minicomputers. The scope of this feasibility study included the following elements:

1. Connecting the LAN to Geonet,
2. Connecting CSIN microcomputers to the LAN,
3. Executing CSIN remotely through a serial communications port rather than a microcomputer console,
4. Accessing commercial information-retrieval systems by way of a public telecommunications network in conjunction with Geonet, and
5. Capturing the results of searches of remote data bases and transferring these results to the user's minicomputer.

Acknowledgments

Special thanks are extended to Daphne L. Chinn of the U.S. Geological Survey for extensive telecommunications testing and for providing assistance with the LAN configurations. In addition, the authors thank John Burruss of Bolt, Beranek, and Newman, Inc., for technical support.

APPROACH

The CSIN project staff was given the opportunity to test the prototype LAN located at the Survey's headquarters in Reston, Virginia. The proposed testing plans called for implementing CSIN software application by using existing telecommunication hardware and software. A description of the tele-

communications hardware and software that compose the LAN configuration is given below.

The Ethernet-based LAN is comprised of communications servers manufactured by Bridge Communications, which are connected by coaxial cables (Estrin and Malone, 1984, p. 200). Figure 1 illustrates the heterogeneous components of the LAN including word processors, laser printers, communications servers, minicomputers, and microcomputers.

The microcomputers are connected to the LAN through an asynchronous communications server. A standard asynchronous communications port on the microcomputer connects to the LAN. The LAN is programmed to connect to an inactive microcomputer. The LAN connects to Geonet and to Tymnet through a network interface, which translates between the LAN and the packet-switching protocol of the wide-area network. The network interface connects to a Tymnet communications minicomputer called an engine. The LAN's unique network address is registered with Tymnet. Tymnet and the LAN software must be configured to recognize each other. Minicomputers are connected to Geonet through a Tymnet engine. In most cases, when the distances between the minicomputers and engines exceed a few tens of feet, dedicated telephone lines are used. Each minicomputer has a network interface and networking software registered by Tymnet. Although the above interconnections between heterogeneous hardware and software are complex, a user on a Prime minicomputer establishes the connection to a CSIN microcomputer by the simple command,

NETLINK -TO CSIN.

Usually, access to microcomputer programs is controlled by the user from the console keyboard, and output is directed to the monitor. Remote, a proprietary software package (Fersko-Weiss, 1985, p. 172–173) that is on each microcomputer, was obtained to provide remote access to the microcomputers. Remote allows the execution of CSIN on the microcomputer using commands transmitted through the asynchronous communications port.

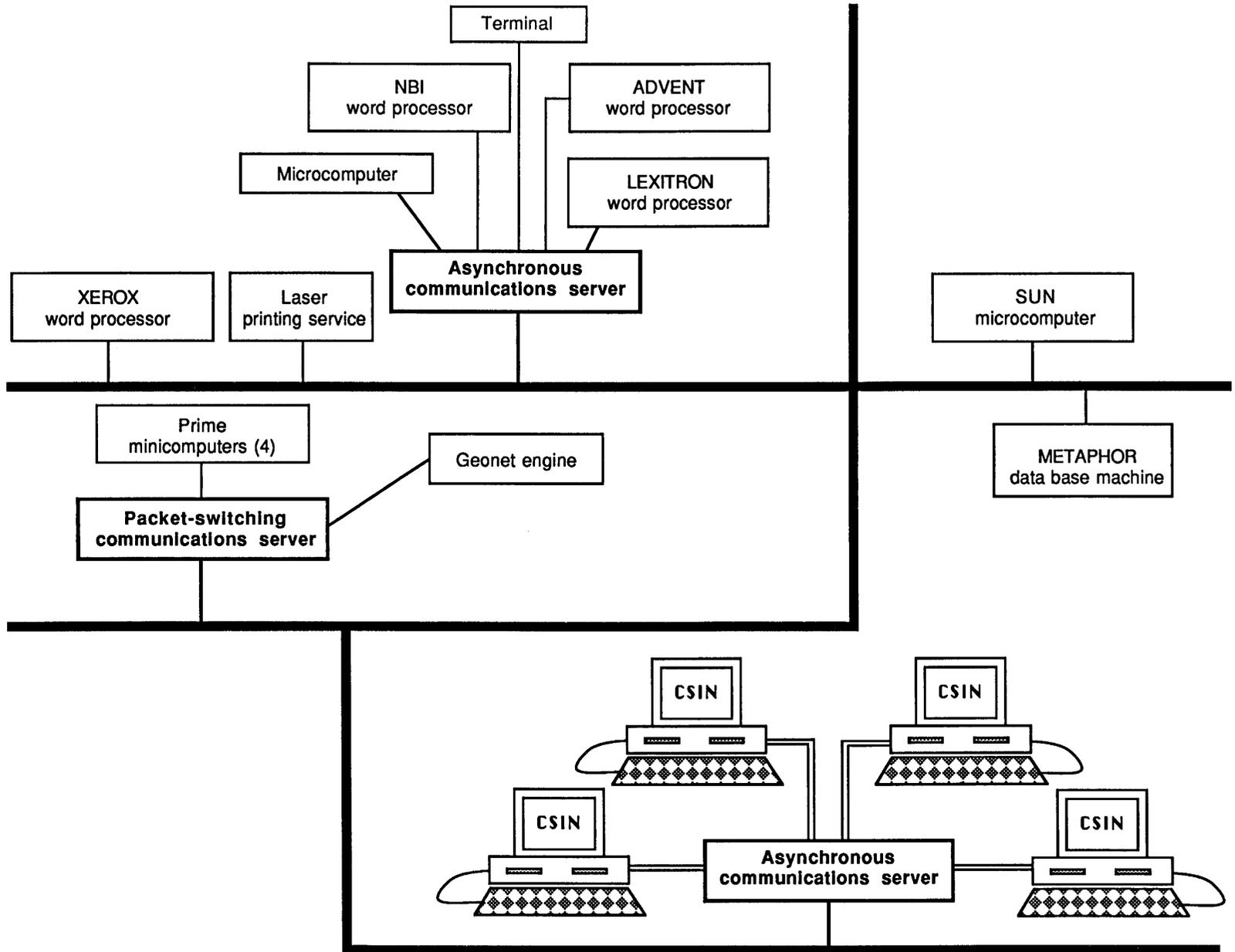


Figure 1. Configuration of the local-area network showing heterogeneous components.

Input and output originates from the client's remote terminal instead of from the microcomputer keyboard.

Each CSIN microcomputer has an additional serial-communications port connected to the LAN. The stand-alone version of CSIN uses dial-out modems connected to this port to access the information-retrieval systems. To avoid additional telephone line costs, CSIN was modified to use this second communications port to access the packet-switching communications server mentioned above so that it can connect with the remote data base systems.

All but one of the commercial information-retrieval systems that CSIN connects to reside on Tymnet's public data network. Access to these systems is by way of the communications route shown in figure 2.

Bibliographic and numeric search results retrieved by CSIN from the information systems listed in table 1 are stored initially in a disk file on the microcomputer. To move this disk file to the client's originating computer, CSIN invokes Kermit software, a terminal emulator and file-transfer utility. By way of the same communications port used to access the remote information-retrieval systems, CSIN connects the client to the host minicomputer through Geonet. By activating Kermit on the host computer, the client can transfer files containing search-retrieval results from the CSIN microcomputer to the host computer. While the results are being transferred, a user has two connections to the microcomputer—one that controls execution of CSIN and one that transfers the retrieval results.

RESULTS AND DISCUSSION

The performance of the CSIN software system has been consistent for the duration of the project. However, problems, many of which are related to the complex heterogeneous design of the LAN, have been evident, and some of these are unsolved.

Microcomputers that run MS-DOS are designed for a single user who interacts with the computer by way of the console keyboard and monitor. Complex software compatibility problems may arise when applications are controlled by way of a serial communications port instead of the console keyboard; for example, when microcomputers occasionally halt, they require a restart, and it is impossible to restart a microcomputer remotely.

The keyboard layout of a CSIN microcomputer and the protocol for controlling images on the monitor are significantly different from a terminal generally used by clients accessing Geonet. In most cases,

the terminal used by these clients operates according to the American National Standards Institute (ANSI) X3.64 standard (American National Standards Institute, 1979, p. 27–S). Fortunately, CSIN does not require the use of any of the microcomputer's unique function keys (Stone, 1987, p. 260), and, therefore, equivalent keystrokes are available on the client's terminal.

The CSIN software partitions a 25-line screen into three independent scrolling regions and bypasses standard MS-DOS display routines to attain high performance. Emulation of these three windows on a 24-line ANSI X3.64 terminal requires the generation of a complex pattern of screen control commands. Remote (Stone, 1985, p. 130–131), which is the same software package that provides access to CSIN by way of the serial port, performs this emulation. Whenever the contents of the display change, Remote transmits only the modified portion of the display to the client's terminal to minimize the number of transmitted characters. Remote's screen emulation can work efficiently when CSIN updates isolated areas of the display, such as a window containing a single line of data (Derfler, 1985, p. 180). Inefficient transmission occurs during screen scrolling when each line essentially is altered and an entire screen image is sent for every new line displayed at the bottom of the monitor.

Another problem related to Remote's screen emulation is unsolved. After several menus are selected, the user's screen may become scrambled. Although Remote appears to correctly emulate ANSI X3.64 terminal protocol, the display characters sometimes are altered along the communications pathway from the microcomputer to the client's terminal (Getts and Markowitz, 1986, p. 220). Once a terminal's display is out of synchronization with Remote, recovery is impossible because Remote never sends the entire screen, only screen changes. The characters are (1) transmitted through the serial communications port to the LAN communications server, (2) packeted by the X.25 communications server, and (3) transmitted by way of Geonet to the remote minicomputer. Finally, in the case of a host Prime minicomputer, Netlink software directs the sequences to the client's terminal. The CSIN project staff, even with the assistance of Geonet and LAN telecommunications specialists, have been unable to isolate the point of alteration in this complex heterogeneous hardware and software system.

Results of the LAN implementation also indicate additional problems associated with telecommunications pathways. Use of the packet-switching com-

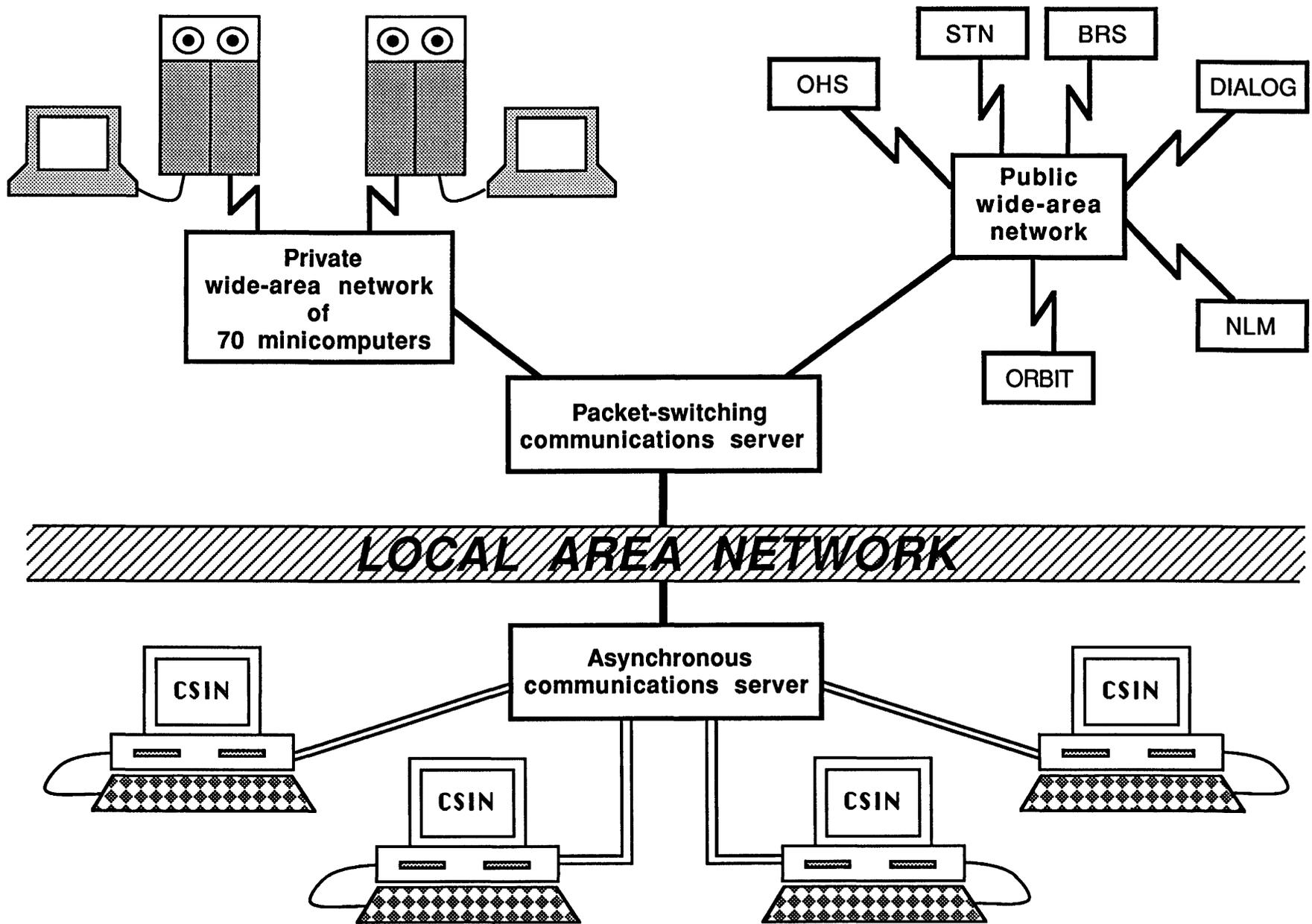


Figure 2. Detailed configuration of Computer Systems Information Network local-area network.

munications server to access the Tymnet public data network, in lieu of dial-out modems, has two limitations.

First, the LAN emulates a variety of protocols in a heterogeneous environment but is not always successful. Data loss to the microcomputer from the client's remote terminal and between the microcomputer and information-retrieval systems is a common occurrence, which leads to frustration and repeated attempts to complete a single bibliographic search. Multiple attempts at reconfiguring the LAN telecommunications pathways by the CSIN project staff have been unsuccessful. Because CSIN is not the sole application on the LAN, reconfiguring an aspect of the telecommunications may affect any LAN application. Adjusting the telecommunications environment is often troublesome and laborious because of these side effects.

Second, one of the information-retrieval systems that CSIN connects to, known as the Chemical Information System (CIS), does not have an address on the Tymnet public data network and, therefore, is inaccessible by this telecommunications approach. Fortunately, CIS contains data bases composed mostly of chemical-toxicity information and is not widely used by Survey personnel.

Although session activity can be captured in a file on the host Prime system, the screen emulation described above renders the resulting captured data file useless without extensive editing (Hansen, 1986, p. 86). To eliminate this extra burden on the user, Kermit is incorporated into CSIN. Kermit can transfer the search results that are captured on the microcomputer to the remote host Prime without transmission errors. If an alternative software package can be found that provides remote access to the microcomputers but does not use the same method of terminal display transmission as Remote, then the use of Kermit to transfer files containing search results will not be necessary. The CSIN project staff has investigated several alternatives to Remote without success.

SUMMARY AND CONCLUSION

A feasibility study conducted at the USGS by a Bibliographic SIG led to the design and implementation of a LAN connecting several CSIN microcomputers. Connecting this LAN to the Survey's wide-area network, Geonet, creates a telecommunications pathway to CSIN for users at remote sites who do not have access to a microcomputer.

The unresolved problems of screen emulation and data loss from retrieval systems render the

product not only difficult to use, but may compromise the accuracy of bibliographic search results. At present, the CSIN project staff continues to investigate methods for solving the telecommunications problems associated with the LAN experiment.

The occurrence of numerous telecommunication problems during the development of the LAN version of CSIN indicates a need for extensive monitoring for troubleshooting purposes. Telecommunications monitoring equipment and software were not readily available to the authors. Costs associated with the hardware and the software and with planning and designing the LAN dictate substantial commitment to the application.

Results of the study indicate that the application of CSIN is not suitable in a heterogeneous environment of hardware and software interconnecting a LAN and a wide-area network. The use of homogeneous versus heterogeneous hardware is a critical decisionmaking factor. For multivendor, heterogeneous environments, applications should be restricted during development to the area of operability common to all hardware and software elements.

REFERENCES

- American National Standards Institute, 1979, Additional controls for use with American national standard code for information interchange: New York, American National Standards Institute, Standard no. X3.64-1979, p. 27-51.
- Angier, J. J., and Hoehl, S. B., 1986, Local area networks (LAN) in the special library, pt. 1—A planning model: *Online*, v. 10, no. 6, p. 19-28.
- Derfler, F. J., 1985, Long-distance computing at full power: *PC Magazine*, v. 4, no. 13, p. 176-180.
- Estrin, J., and Malone, J. P., 1984, Software-transparent front-end processor links hosts, terminals to Ethernet: *Mini-Micro Systems*, v. 57, no. 13, p. 199-208.
- Fersko-Weiss, H., 1985, The Remote possibility: *Personal Computing*, v. 9, no. 5, p. 172-173.
- Getts, Judy, and Markowitz, Michael, 1986, Remote quadrille: *PC World*, v. 4, no. 9, p. 218-227.
- Hansen, Augie, 1986, Controlling from afar: *PC Tech Journal*, v. 4, no. 4, p. 84-87, 89 ff.
- Stone, M. D., 1985, Remote: *PC Magazine*, v. 4, no. 24, p. 130-131.
- 1987, Computing at a distance: *PC Magazine*, v. 6, no. 6, p. 249-262.
- Wiltshire, D. A., and Hollway, R. A., 1987, Design and implementation of a customized gateway system for information retrieval from public and commercial databases, in *National Online Meeting*, 8th, New York, May 1987, Proceedings: Medford, N.J., Learned Information, Inc., p. 483-492.

CHAPTER E

Generic File Structures for Selected Applications of Water-Resources Data

By MARK L. FARMER

U.S. GEOLOGICAL SURVEY BULLETIN 1841

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES



CONTENTS

Abstract	E1
Introduction	E1
Purpose and scope	E1
Definition	E1
Background	E1
Structure knowledge	E2
Application program transfer files	E2
Interface programs	E3
Summary	E4
References	E4



Generic File Structures for Selected Applications of Water-Resources Data

By Mark L. Farmer

Abstract

Generic file structures are rapidly becoming necessary for efficient data storage, retrieval, and analysis operations in the U.S. Geological Survey. Many programs have been written for the Geological Survey's National Water Information System data bases that facilitate the need and the use of generic file structures; however, no nationally supported guidelines for the preparation or use of these structures exist at this time. Presently, few conceptual philosophies from which to select a model design are available, but efforts are in progress within the Automated Data Processing System programming effort that could have a significant impact on the future role of these file types. This paper discusses the generic file structure concept and definition and focuses on the need for standardization of these structures in the National Water Information System data bases. The paper also explores the development of the Application Transfer Format file structure, as well as interface subroutines currently in development to handle writing and reading data to and from these file types.

INTRODUCTION

Currently, no nationally supported guidelines exist for the use, format, or definition of generic file structures within the U.S. Geological Survey's National Water Information System (NWIS), but a need for data commonality does exist among the different data bases of the NWIS, which include the Ground Water Site Inventory, Quality of Water, Automated Data Processing System (ADAPS), and Water Use. Efforts in progress within the ADAPS development may help to delineate guidelines for possible standardization of these structures.

Purpose and Scope

The purpose of this paper is to discuss the generic file structure concept and definition, focusing on the need for standardization of these structures in the NWIS data bases. The paper also explores the

Application Transfer File (APTRAF) structure and the development of interface subroutines to handle writing and reading data to and from these file types.

Definition

Generic file structures may vary in form and in content from one application to the next, but a few basic elements are common to all generic structures. The most fundamental characteristic of these files is that the structure contains data to be processed by one or more subsequent applications. This implies that the structure should not be application specific. Moreover, the file should be usable with a variety of program applications, as well as a variety of computer systems. This capability is what ultimately constitutes the term "generic." The primary types of data subject to generic structure implementation within the NWIS are linear and multidimensional arrays of numeric, character, and logical data. These are the standard storage structures that are the most convenient, as well as the most comprehensible, to use in a hydrologic environment. Arrays and tables are easily human and machine readable, and a generic interpretation needs to possess the same readability.

Background

Historically, data have been transferred to and from the Geological Survey's National Water Data Storage and Retrieval System (WATSTORE) by means of "card image" records, which are written to the specifications of the WATSTORE applications that read and write the data files. The final card image data file produced contains the arbitrary number of records written to the prescribed formats demanded by the WATSTORE program generating the file. The records commonly consist of a variety of different card types, each having a special purpose for the application using the file. The records are

written in structured formats that can be read easily by the postprocessing application. This type of file generation is often used with the Daily Values Retrieval Program in the WATSTORE system. If the user requests card image output, then the program will generate a series of daily values code (Type 2) cards and daily values data (Type 3) cards. If specified, the program also will generate password-agency (Type Z) cards, station header (Type H) cards, and station name (Type N) cards. Each card type has a specific function and a prescribed format for all data items generated on the card. To define these functions and formats, the user must refer to the documentation for the program in the WATSTORE User's Guide (Hutchinson, 1975, p. IV-B 27). Technically, these card image files fit the basic definition of a generic file structure in that they contain data to be processed by one or more subsequent applications; however, these files are strictly WATSTORE specific. Any application that uses these files requires advance knowledge of the file structure to efficiently process these data.

STRUCTURE KNOWLEDGE

It would be impossible to eliminate all aspects of the programming environment from any portable file structure. Any subsequent processing program needs to have some knowledge of structure before accessing the data in the file. This knowledge does not have to be of data forms and column placement, but can consist of a brief set of rules governing how the structure may "self-define" the data in the file. If row and column data are to be accessed in this file, then some method of identifying what the rows and columns represent needs to exist in the file; data types in the file also should be indicated. The generic structure concept can be expressed easily as a method where the data file is readily accessible to the post-processing programs with a minimum amount of parsing and searching for a specific data item. If properly defined, then the data in the file can be read by an application without consuming large amounts of processor time "massaging" the data. As long as the structure knowledge remains as information only about the structure entity and is not dependent on the application using or generating the structure, a generic environment can be achieved in a multiprocessing data system.

Structure knowledge also can facilitate the use of interface programs to coordinate the flow of data to and from the generic structure. These routines can save a programmer from the variable analysis required to read a generic structure. The interface

programs are not necessarily "generic" by definition but would be used specifically to handle the overhead involved with the input and the output to the generic file. Interface programs are under development in the ADAPS programming effort of the NWIS that are functionally similar to the types of programs described above. The Water Resources Division's Office of Surface Water has designed a utility program to read APTRAF, and "generic table file writing subroutines" are being written by personnel in the New Jersey District Office of the Geological Survey. These programs process tabular data produced from statistical programs in the ADAPS, as well as other sources.

APPLICATION PROGRAM TRANSFER FILES

The Daily Values Statistical Program (DVSTAT) and the Daily Values Monthly and Annual Summary Program (DVMAS) are currently producing optional APTRAF as part of the available output selections in the ADAPS. APTRAF, as mentioned above, were developed by the Office of Surface Water to generalize the production of statistical output tables and arrays in a suitable generalized format for subsequent programs to access.

The original proposal for development of the APTRAF suggested that the files be suitable for a variety of types of hydrologic data from several systems, including time-series data, water quality, extreme events, and water use. The proposal also suggested that the structure be a stand-alone file containing all required information necessary for analytical purposes and that it be usable by various NWIS production-level programs and by a variety of ad hoc user-written applications. Finally, it was suggested that the programs accessing these files be processed in unpredictable sequences, each program using an APTRAF generated from a preceding application and writing the results to an APTRAF to be used by one or more subsequent programs (G. R. Dempster and W. H. Kirby, U.S. Geological Survey, written commun., 1985).

The structural knowledge required for APTRAF implementation consists of knowing the APTRAF record definitions and their structural components. Note that this knowledge is file specific and not program specific; for example, APTRAF describes each data using one or more lines of title information in a table format. The APTRAF data set also consists of real-valued numerical data arranged in a tabular design of rows and columns. The columns in the table represent observed or computed

values of a single variable, and each row is expressed as a set of these variables. Each data set also may contain optional auxiliary records to further identify the tabular data. The APTRAF record types are generally free format, with the exception of the record-type keywords beginning in column 1. The data elements are also free format unless a \$FORMAT record exists.

The APTRAF data set consists of specific record types to define the various descriptors and data elements written in the file. A basic implementation of this structure can be exemplified in the DVSTAT program of the ADAPS. The following is a linear APTRAF table of supplementary statistical characteristics calculated from a flow-duration analysis:

```

$COM ADAPS/DVSTAT - WED, APR 29 1987 - 1 6:12:12
$title1 Station 07056000 BUFFALO RIVER NR ST. JOE, ARK.
$title2 Parameter code 00060 - DISCHARGE
$title3 Statistic code 00003 - MEAN
$title4 Supplementary Statistical Characteristics
$title5 For period OCT - SEP
$PARM MEAN
$PARM STD Standard Deviation
$PARM CV Coefficient of Variation
$PARM SKEW Coefficient of Skew
$DATA
    2.59460 0.67213 0.25905 -0.13321
$END

```

Note that the \$TITLE records, which contain title information, contain information pertaining to parameter codes, statistical codes, and the period of record. This information could have been included in the actual data table if the necessary variable definition had been included. The \$COM records provide comment lines to the table. The \$PARM records identify the respective columns of the row and do the actual variable assignment from the \$PARM record name. These numerical data are defined by the \$DATA statement and are concluded with the \$END statement. These data are written in a free format design in which the actual position of individual data items is unimportant, providing they are placed in the respective order of the \$PARM record defining them.

A general purpose subroutine, Application Program Transfer File Reader (APTRAN), has been developed to facilitate the reading of APTRAF. This subroutine extracts tables, one at a time, from an existing APTRAF and stores the information in the subroutine argument list. APTRAN, written in Fortran 77, scans the APTRAF, identifies the different record types, extracts these data from the records, and stores it in internal storage arrays previously

defined by the calling program (W. H. Kirby, U.S. Geological Survey, written commun., 1986).

The Pearson frequency programs under development in the ADAPS will be implementing the APTRAN subroutine to process APTRAF generated from DVSTAT and DVMAS. Personnel in the Arkansas District of the Geological Survey have been testing the APTRAN routine since October 1986.

INTERFACE PROGRAMS

Personnel in the New Jersey District of the Geological Survey are developing interface routines to be used to write table-type files that have particular emphasis on P-STAT (Buhler and Buhler, 1986,) and Statistical Analysis System (SAS) (SAS Institute, Inc., 1982,) applications. These routines can be used for time-series, site, parameter, measurement, and rating-file data produced from the ADAPS (J. L. Fulton, U.S. Geological Survey, written commun., 1987). The approach to the file structure definition is similar to the way APTRAF perceives a data table. These routines operate on a few basic assumptions about the table structure as follows:

1. It contains a defined number of rows and columns,
2. The table will be written by rows,
3. All data in a single column must be of the same type,
4. The allowable data types are character, real, double precision, integer*2, integer*4, logical*2, and logical*4,
5. Each column must have a label that is a legal variable name for the type of data being written, and
6. The table can have any number of header and footer lines describing the data.

Three basic subroutine types will be required for the interfaces to open the table file, to write a row of data to the table file, and to close the table file. The objective of the interfaces is to provide routines general enough to be used not only within the ADAPS, but with other NWIS subsystems as well.

Although the interface routines will be used initially to create P-STAT files and flat files, they will be expanded in the near future to optionally create Data Interchange Format files, APTRAF and Prolog (McCammon, 1986) data-base files. The table-file software is general enough so that interface programs that use the software will not need modifications to work with new output formats.

SUMMARY

Generic file structures are rapidly emerging as solutions to problems of data flow between various data subsystems and software applications of the NWIS. By definition, these structures are row- and column-oriented "table-type" files and are designed specifically to be usable in one or more subsequent applications. Although progress has been made in the development of the APTRAF structure and interface writing routines, no nationally supported specifications or guidelines for generic structures within the NWIS exist at this time. However, the progress does provide us with prototype systems that can be considered for the possible data standardization of these files. The emphasis on structure knowledge and file-specific characteristics, as opposed to program-specific guidelines, increases the file's potential transportability among a variety of different applications. Interface programs also may be used to coordinate

the flow of data to and from these files, which relieves users and programmers of these data "massaging" tasks presently required in many applications. The standardization and support of these types of routines and file structures could greatly enhance data manipulation within the subsystems of the NWIS.

REFERENCES

- Buhler, Shirrell, and Buhler, Roald, 1986, P-STAT user's manual: Edison, N.J., J. Schiller, Inc., p. 3.
- Hutchinson, N. E., 1975, WATSTORE—National water data storage and retrieval system of the U.S. Geological Survey: User's Guide, p. IV-B 27.
- McCammon, R. B., 1986, The muPROSPECTOR mineral consultant system: U.S. Geological Survey Bulletin 1697, p. 3.
- SAS Institute Inc., 1982, SAS user's guide—Basics: Cary, N.C., SAS Institute Inc., p. 54.

SELECTED SERIES OF U.S. GEOLOGICAL SURVEY PUBLICATIONS

Periodicals

Earthquakes & Volcanoes (issued bimonthly).

Preliminary Determination of Epicenters (issued monthly).

Technical Books and Reports

Professional Papers are mainly comprehensive scientific reports of wide and lasting interest and importance to professional scientists and engineers. Included are reports on the results of resource studies and of topographic, hydrologic, and geologic investigations. They also include collections of related papers addressing different aspects of a single scientific topic.

Bulletins contain significant data and interpretations that are of lasting scientific interest but are generally more limited in scope or geographic coverage than Professional Papers. They include the results of resource studies and of geologic and topographic investigations; as well as collections of short papers related to a specific topic.

Water-Supply Papers are comprehensive reports that present significant interpretive results of hydrologic investigations of wide interest to professional geologists, hydrologists, and engineers. The series covers investigations in all phases of hydrology, including hydrogeology, availability of water, quality of water, and use of water.

Circulars present administrative information or important scientific information of wide popular interest in a format designed for distribution at no cost to the public. Information is usually of short-term interest.

Water-Resources Investigations Reports are papers of an interpretive nature made available to the public outside the formal USGS publications series. Copies are reproduced on request unlike formal USGS publications, and they are also available for public inspection at depositories indicated in USGS catalogs.

Open-File Reports include unpublished manuscript reports, maps, and other material that are made available for public consultation at depositories. They are a nonpermanent form of publication that may be cited in other publications as sources of information.

Maps

Geologic Quadrangle Maps are multicolor geologic maps on topographic bases in 7 1/2- or 15-minute quadrangle formats (scales mainly 1:24,000 or 1:62,500) showing bedrock, surficial, or engineering geology. Maps generally include brief texts; some maps include structure and columnar sections only.

Geophysical Investigations Maps are on topographic or planimetric bases at various scales; they show results of surveys using geophysical techniques, such as gravity, magnetic, seismic, or radioactivity, which reflect subsurface structures that are of economic or geologic significance. Many maps include correlations with the geology.

Miscellaneous Investigations Series Maps are on planimetric or topographic bases of regular and irregular areas at various scales; they present a wide variety of format and subject matter. The series also includes 7 1/2-minute quadrangle photogeologic maps on planimetric bases which show geology as interpreted from aerial photographs. Series also includes maps of Mars and the Moon.

Coal Investigations Maps are geologic maps on topographic or planimetric bases at various scales showing bedrock or surficial geology, stratigraphy, and structural relations in certain coal-resource areas.

Oil and Gas Investigations Charts show stratigraphic information for certain oil and gas fields and other areas having petroleum potential.

Miscellaneous Field Studies Maps are multicolor or black-and-white maps on topographic or planimetric bases on quadrangle or irregular areas at various scales. Pre-1971 maps show bedrock geology in relation to specific mining or mineral-deposit problems; post-1971 maps are primarily black-and-white maps on various subjects such as environmental studies or wilderness mineral investigations.

Hydrologic Investigations Atlases are multicolored or black-and-white maps on topographic or planimetric bases presenting a wide range of geohydrologic data of both regular and irregular areas; principal scale is 1:24,000 and regional studies are at 1:250,000 scale or smaller.

Catalogs

Permanent catalogs, as well as some others, giving comprehensive listings of U.S. Geological Survey publications are available under the conditions indicated below from the U.S. Geological Survey, Books and Open-File Reports Section, Federal Center, Box 25425, Denver, CO 80225. (See latest Price and Availability List.)

"**Publications of the Geological Survey, 1879- 1961**" may be purchased by mail and over the counter in paperback book form and as a set of microfiche.

"**Publications of the Geological Survey, 1962- 1970**" may be purchased by mail and over the counter in paperback book form and as a set of microfiche.

"**Publications of the U.S. Geological Survey, 1971- 1981**" may be purchased by mail and over the counter in paperback book form (two volumes, publications listing and index) and as a set of microfiche.

Supplements for 1982, 1983, 1984, 1985, 1986, and for subsequent years since the last permanent catalog may be purchased by mail and over the counter in paperback book form.

State catalogs, "List of U.S. Geological Survey Geologic and Water-Supply Reports and Maps For (State)," may be purchased by mail and over the counter in paperback booklet form only.

"**Price and Availability List of U.S. Geological Survey Publications**," issued annually, is available free of charge in paperback booklet form only.

Selected copies of a monthly catalog "New Publications of the U.S. Geological Survey" available free of charge by mail or may be obtained over the counter in paperback booklet form only. Those wishing a free subscription to the monthly catalog "New Publications of the U.S. Geological Survey" should write to the U.S. Geological Survey, 582 National Center, Reston, VA 22092.

Note.--Prices of Government publications listed in older catalogs, announcements, and publications may be incorrect. Therefore, the prices charged may differ from the prices in catalogs, announcements, and publications.

