Selected Papers in the Applied Computer Sciences 1990

U.S. GEOLOGICAL SURVEY BULLETIN 1908

U.S. DEPARTMENT OF THE INTERIOR
March 3, 1849

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 U.S. 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 U.S. 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 1990

Edited by D.A. WILTSHIRE

This volume is published as chapters A–F.
These chapters are not available separately

U.S. GEOLOGICAL SURVEY BULLETIN 1908

DEPARTMENT OF THE INTERIOR
MANUEL LUJAN, Jr., Secretary

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

UNITED STATES GOVERNMENT PRINTING OFFICE: 1990

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 1990 / edited by D.A. Wiltshire.
p. cm.—(U.S. Geological Survey bulletin ; 1908)
Includes bibliographical references.
1. Electronic data processing. I. Wiltshire, Denise A. II. Series.
QE75.B9 no. 1908 [QA76] 557.3—dc20 [004] 89–600309 CIP

FOREWORD

The role of computer technology in earth science investigations has evolved into an integral aspect of the research process. Today, computer methods are used to simulate and analyze complex earth processes. Geographic information systems have been developed to analyze the spatial relations among scientific data sets, such as topography, land use, geology, and water resources. Hardware and data base technology have advanced such that researchers have the capability to store, manipulate, and disseminate very large data sets, thus providing innovative methods for archiving historical information. Recent advances in laser printing and publishing software provide new opportunities for producing complex geoscience reports and maps by using desktop computer systems.

This compendium of papers reports on the technical advances in the applied computer sciences that were developed to support earth science research. This second volume, published in the U.S. Geological Survey Bulletin series, comprises short papers that address the following aspects of the applied computer sciences:

- Transformations in computer graphics,
- Production of computerized color graphics for publications,
- Color separation techniques for geologic maps,
- Digital spatial data technology,
- Geographic information system analysis of indoor radon values, and
- A water-use data system.

The purpose of this series is to provide a forum for the discussion of innovative computer techniques for the earth sciences; hence, comments are encouraged. Readers may submit comments on these papers and those published in the first volume (1988) to:

Editor, Selected Papers in the Applied Computer Sciences
U.S. Geological Survey
801 National Center
Reston, VA 22092.

CONTENTS

[Letters designate the chapters]

CHAPTER A

Applications of User-Supplied Transformations In Computer-Graphics Programs

By STANLEY A. LEAKE

U.S. GEOLOGICAL SURVEY BULLETIN 1908

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES 1990

CONTENTS

FIGURES

Applications of User-Supplied Transformations In Computer-Graphics Programs

By Stanley A. Leake

Abstract

User-supplied transformations of plot vectors are a feature of the CA–DISSPLA[1] software package. The transformations, which involve changing the coordinates of the plot vectors by using a set of rules established by the user, are useful for applications in the presentation of scientific information and in the development of graphical art. One useful application of user-supplied transformations is the contouring of data from irregular finite-difference grids. Transformation of plot vectors results in contours that retain the smoothness and detail of the original surface obtained from the finite-difference grid. An application in the development of graphical art is the transformation of straight lines into wavy or curved lines.

INTRODUCTION

The software package CA–DISSPLA is a collection of FORTRAN routines that can be used in a wide range of computer-graphics applications (Computer Associates, 1985). Because the package is extensive, some of its capabilities and features are not well known by graphics programmers. One of the lesser known features of CA–DISSPLA is an option to allow final plot vectors to be transformed by a set of rules established by the user. This option is referred to as user-supplied transformations. The capabilities of user-supplied transformations extend far beyond those of the more familiar transformations of scaling, rotation, and shifting of plot vectors.

User-supplied transformations have been applied in the presentation of scientific information by programmers in the U.S. Geological Survey (USGS). These transformations are useful in the creation of graphical art as well.

USE OF TRANSFORMATIONS IN GRAPHICS PROGRAMS

Two types of user-supplied transformations are available in the CA–DISSPLA software package. The first type of transformation, ATRANS, transforms only the end points of plot vectors (fig. 1*A*). A straight line is drawn from the starting point of the vector to the transformed end point. This type of transformation is signaled with a call to the CA–DISSPLA subroutine ATRANS. The program must include a FORTRAN subroutine, TRANS, that has three arguments. The first two arguments are real variables that contain the untransformed coordinates of the vector end point upon entering the subroutine and the transformed coordinates upon exiting the subroutine. The units of the coordinates passed to subroutine TRANS are inches, not the units specified by the current graph axes. The third argument is an integer flag that indicates whether the pen is up or down.

The second type of transformation, BTRANS, transforms intermediate points along the vector as well as the end point (fig. 1*B*). All other aspects of operation are identical to that of ATRANS except that a call to the CA–DISSPLA routine BTRANS is required to signal the transformation. BTRANS selects intermediate points along the vector by breaking the original untransformed vector into as many smaller vectors as is needed to give a resolution of 0.05 inch.

The results of the ATRANS option are acceptable if plot vectors are relatively small or if little or no curvature results from the transformation. The BTRANS option requires more computer work than ATRANS because considerably more points are transformed. Some desired effects of transformations cannot be attained by transforming only end points of vectors.

Neither ATRANS nor BTRANS transforms plot vectors that make up text characters or graph symbols. For text strings, only the location of the start of the string is transformed. Similarly, for graph symbols, only the location of the symbol is transformed. Also, areas in a drawing that are blanked prior to invoking a user-supplied transformation remain untransformed. A method is available to transform all vectors including text characters and symbols. The method involves saving the untransformed drawing in a metafile and using the CA–DISSPLA routines GETMET and PUTMET along with the desired transformation in a

Manuscript approved for publication September 9, 1988.

[1]CA–DISSPLA is a registered trademark of Computer Associates International, Inc.

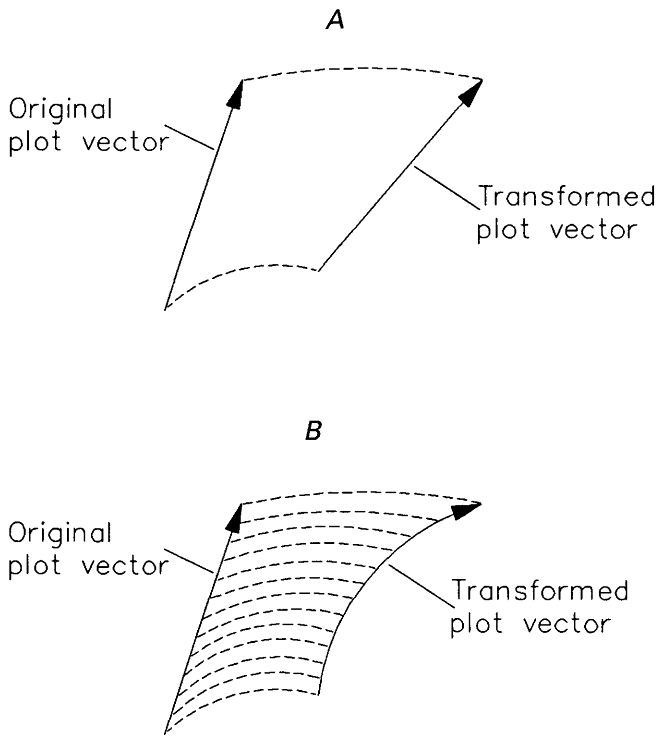


Figure 1. Types of user-supplied transformations available in the CA–DISSPLA software package (modified from Computer Associates, 1985, fig. B 28–1). *A*, Action of ATRANS causes only end points of vectors to be transformed. *B*, Action of BTRANS causes end points and intermediate points along vector to be transformed.

program that redraws the metafile. By this method, vectors that make up text characters and symbols are not distinguished from any other vectors in the new drawing and therefore are also transformed.

APPLICATIONS FOR PRESENTATION OF SCIENTIFIC DATA

CA–DISSPLA and other software packages provide tools for standard graphical presentation of scientific data. Some specialized applications, however, require features that are not part of most graphics packages. One application developed within the USGS is contouring surfaces from irregular finite-difference grids.

Several widely used ground-water flow models allow for the computation of head values for a flow system by using a rectangular finite-difference grid. Contours of the gridded surface provide a useful graphical tool for the interpretation of model results. Surfaces of hydraulic head computed by using ground-water models generally are smooth and well defined by the gridded model results. Finite-difference grids may be regular or irregular. A regular grid is one in which the grid spacing is uniform in each direction, although the grid spacing in one direction is not necessarily the same as in the other direction (fig. 2). An irregular grid has nonuniform spacing in at least one of the directions.

The CA–DISSPLA software package provides routines for contouring data from regular grids. Data from irregular finite-difference grids can be contoured by generating a new regular grid of data values by using CA–DISSPLA capabilities to create regular grids from randomly spaced data. This approach, however, is undesirable for several reasons. First, the data from the irregular grid are not randomly spaced, and to treat the data as such is inefficient use of computer resources. Second, the process of generating a new regular grid commonly results in a surface that is not as smooth as the original surface obtained from the ground-water model. Furthermore, where the spacing of the new uniform grid is larger than the spacing of the irregular grid, detail of the surface is lost.

A better approach is to contour the data set as though it was from a regular grid. Prior to plotting the contours, however, the plot vectors can be transformed into the locations dictated by the original irregular grid. For the shaded block (row 3, column 2, numbered from the lower-left block) in the regular grid shown in figure 2, plot vector coordinates X and Y can be transformed into coordinates x and y for the corresponding block in the irregular grid by using the FORTRAN equivalent of the following transformation:

$$x=x_2+\frac{(X-X_2)\cdot \Delta x_3}{\Delta X},$$

$$y=y_1+\frac{(Y-Y_1)\cdot \Delta y_2}{\Delta Y},$$

where x_2, X_2, Δx_3, ΔX, y_1, Y_1, Δy_2, and ΔY are dimensions shown in figure 2. A computer program written by the author for use within the USGS uses this transformation procedure to contour results from the ground-water flow model developed by McDonald and Harbaugh (1988). Resulting contours retain the detail and smoothness of the original surface obtained from the finite-difference grid.

Other applications of transformations for the presentation of scientific information include Piper diagrams for the display of chemical-quality data and probability graphs that are not included in the CA–DISSPLA software package. Also, transformation routines can be used for the numerical integration of smoothed curves (Computer Associates, 1985, pt. B, p. 28–6).

APPLICATIONS FOR GRAPHICAL ART

The CA–DISSPLA software package provides programmers with extensive capabilities for the creation of computer drawings. Many programmers in the USGS use

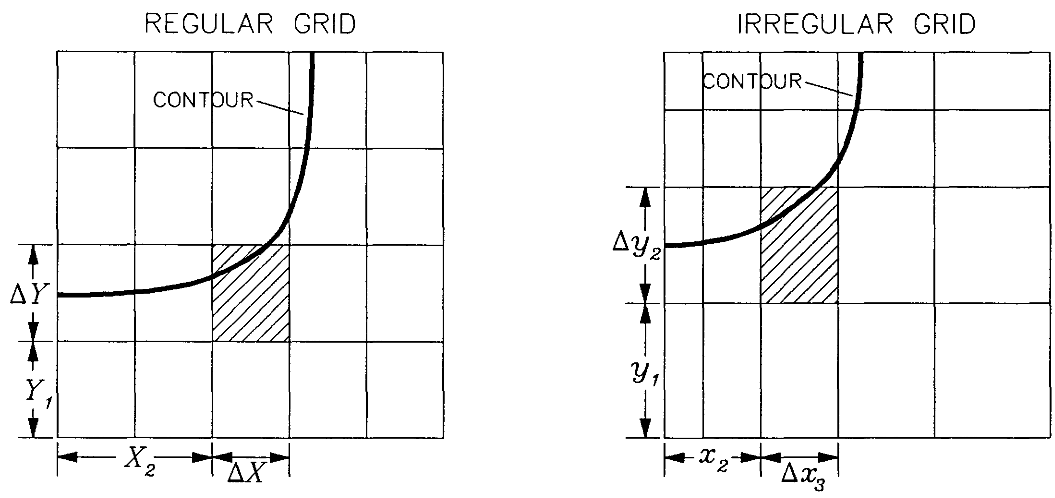


Figure 2. Regular and irregular finite-difference grids.

CA–DISSPLA to prepare graphical art for report illustrations, posters, and presentation slides. Interesting and unusual visual effects can be added to drawings by applying user-supplied transformations.

An example of a simple graphical illustration is given in figure 3*A*. A few of the standard capabilities of the CA–DISSPLA software package were used to prepare the illustration. The character set SWISSL was used to display "NCTM '88," an abbreviation for the 1988 National Computer Technology Meeting of the USGS. Character blanking was used to make the text more prominent against the shaded square in the background.

Suppose that a desired enhancement for the illustration is to have the background made of wavy lines rather than straight lines. Wavy lines are not part of the standard shading capabilities of CA–DISSPLA; therefore, a user-supplied transformation is needed. The plot vectors for the background shading can be transformed by invoking BTRANS after the text is drawn. Subroutine TRANS must include the FORTRAN equivalent of the transformation from original plot vector coordinates X and Y to coordinates x and y:

$$x=X+A \cdot \sin(P \cdot Y),$$
$$y=Y+A \cdot \sin(P \cdot X),$$

where A is a parameter that specifies the amplitude of the waves and P is a parameter that specifies the period of the waves. The results (fig. 3*B*) demonstrate that the transformed plot vectors are not drawn in areas that were blanked in the original coordinate system.

If the desired effect is to have the text characters transformed as well as the background shading, then a metafile of the untransformed drawing (fig. 3*A*) needs to be transformed in a separate program. Application of the previous transformation for wavy lines results in the unusual illustration shown in figure 3*C*.

For a final example, suppose that the desired effect is to have the illustration appear distorted as though it were displayed on the curved surface of a video screen. Because the text and background are to be transformed, the metafile of figure 3*A* needs to be transformed in a separate program. Option BTRANS needs to be invoked with an appropriate transformation given in subroutine TRANS. The following transformation of plot vector coordinates X and Y to coordinates x and y produces the desired effect:

$$x=X+\exp(C \cdot R) \cdot (X-X_0),$$
$$y=Y+\exp(C \cdot R) \cdot (Y-Y_0),$$

where exp is the exponential function, C is a parameter that controls the apparent curvature, R is the distance to the center of the drawing, and X_0 and Y_0 are the coordinates of the center of the drawing. The value of R may be computed as:

$$R=[(X-X_0)^2+(Y-Y_0)^2]^{0.5}.$$

An example of this transformation, shown in figure 3*D*, was made by using a value of C equal to 0.8. Different transformations may exist to produce similar visual effects.

Another interesting application of user-supplied transformations is to use subroutine TRANS to change the color of plot vectors as they are being drawn. Hue, saturation, and intensity can be computed as a function of plot vector coordinates. For this application, color changes will be limited to the resolution of increments generated by BTRANS. In general, applications of user-supplied transformations in graphical art are limited only by the ability of the programmer to formulate the desired effects in subroutine TRANS.

SUMMARY

User-supplied transformations are a useful feature of the CA–DISSPLA software package for the presentation of

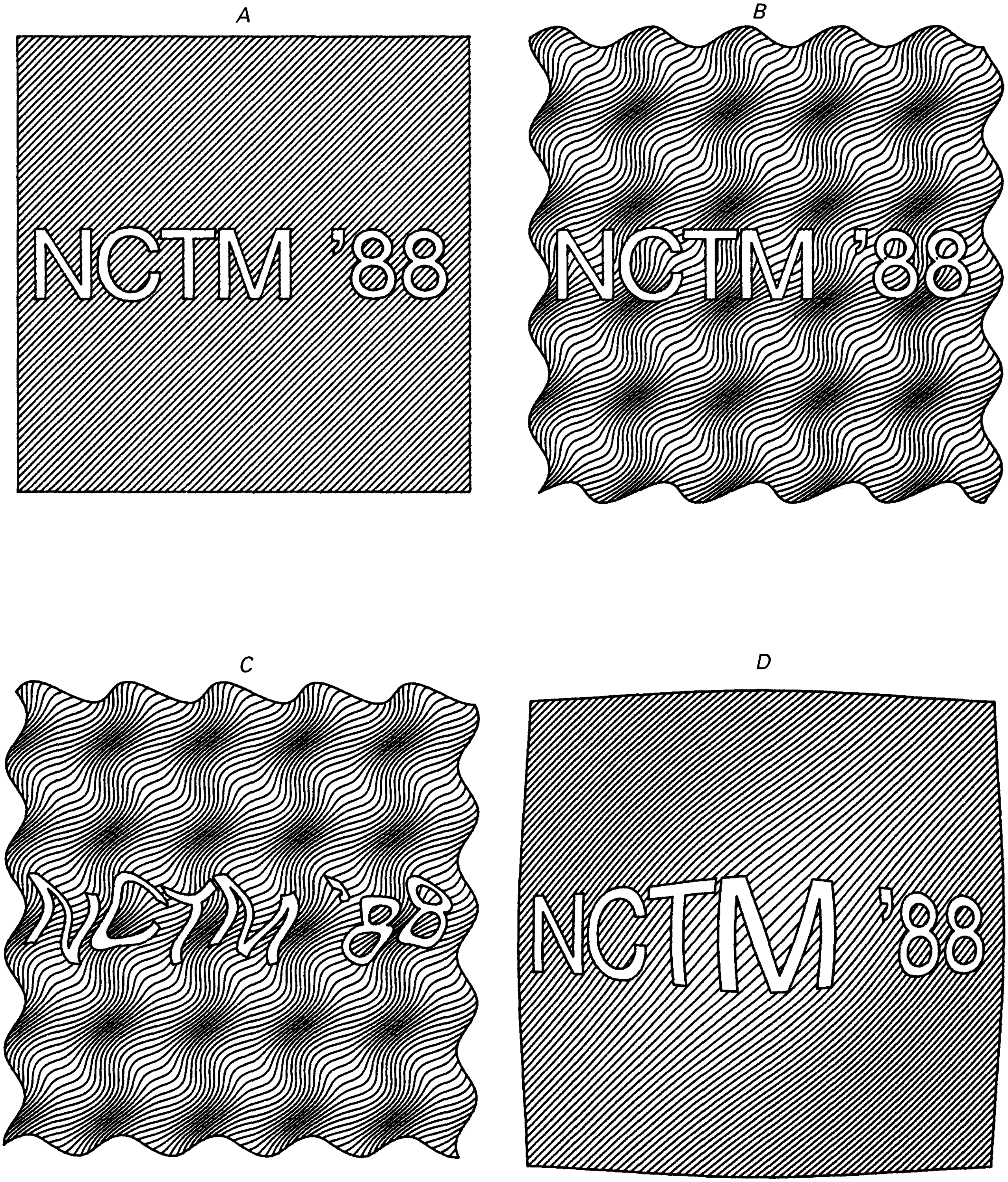


Figure 3. Example applications of user-supplied transformations in graphical art. *A*, Untransformed drawing. *B*, Background shading transformed from straight lines to wavy lines. *C*, Entire drawing transformed to wavy lines. *D*, Entire drawing transformed to appear as the curved surface of a video screen.

scientific information and the development of graphical art. Two types of transformations are available. The first type, ATRANS, transforms only end points of plot vectors. The second type, BTRANS, transforms increments along the vector as well as the end points. The BTRANS option requires more computations by the program than ATRANS but creates effects that would not be possible by transforming only end points. For either type, a user-supplied subroutine called TRANS needs to be programmed to include the transformation.

One useful application of user-supplied transformations is the contouring of data from irregular finite-difference grids. Transformation of plot vectors results in contours that retain the smoothness and detail of the original surface obtained from the finite-difference grid. User-supplied transformations can be applied in the development of graphical art to produce desired effects.

REFERENCES CITED

Computer Associates International, Inc. (formerly Integrated Software Systems Corporation), 1985, DISSPLA user's manual, version 10.0: San Diego, Calif. [Computer Associates], variously paged.

McDonald, M.G., and Harbaugh, A.W., 1988, A modular three-dimensional finite-difference ground-water flow model: U.S. Geological Survey Techniques of Water-Resources Investigations, book 6, chap. A1, 586 p.

CHAPTER B

Linking Digital Technology to Printing Technology for Producing Publication-Quality Color Graphics

By GREGORY J. ALLORD and KERIE J. HITT

U.S. GEOLOGICAL SURVEY BULLETIN 1908

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES 1990

CONTENTS

FIGURES

Linking Digital Technology to Printing Technology For Producing Publication-Quality Color Graphics

By Gregory J. Allord *and* Kerie J. Hitt

Abstract

The U.S. Geological Survey has published thousands of high-resolution computer-generated graphs and maps in the Water-Supply Paper series the National Water Summary (U.S. Geological Survey, 1984, 1985, 1986, 1988, 1990). The production of publication-quality computer graphics is accomplished by linking existing digital technology to pre-press technology. For example, Prime minicomputers running CA–TELLAGRAF[1] and ARC/INFO software can be linked to output devices such as the Scitex and the Linotype L300. The hardware, software, and techniques used to prepare illustrations have evolved to keep pace with the increasing quantity and high complexity of the graphics required in each subsequent water summary. Examples of design criteria, choice of hardware and software, and methods of production used to prepare each report were chosen to illustrate this evolution. The goal has been to design computer data bases that are conducive to analysis and to effective displaying of data.

INTRODUCTION

In 1983, the Secretary of the Interior directed the U.S. Geological Survey (USGS) to prepare a National Water Summary. In response, the USGS initiated a series of reports that describe the conditions, trends, availability, quality, and use of the Nation's water resources.

Each water summary report is organized into three main sections. The first section describes seasonal hydrologic conditions and events of the current water year. The second section contains articles that give background information to help the reader understand the feature topic of the report. These first two sections also explain principles and terms and give examples to help people interpret the information contained in the State summaries, which form the third and largest (70 percent) section of the report. Within each report, the State summaries follow a consistent format so that the information they contain can be compared from State to State.

Each article and State summary is illustrated with figures consisting of many components, such as maps, graphs, and diagrams. Overall, several thousand individual graphics are required for each water summary report.

PRE-PRESS TECHNOLOGY

Because of the quantity and complexity of the figures, preparing the illustrations by traditional manual methods became overwhelming. To overcome this problem, USGS managers and illustrators explored methods to link digital geographic information system data bases (Moody and Lanfear, 1988) to pre-press technology. The objective was to maintain the same high standards of visual accuracy and aesthetics as traditional methods of preparation while taking advantage of the computer's ability to quickly perform repetitious tasks and to process massive amounts of data and computations.

Preparing publication-quality graphics by using computerized methods involves three interrelated elements: the information that is to be portrayed, the desired design and format of the graphic, and the software and hardware that output the graphic. The size, complexity, and required resolution of the final graphic dictate the kind of hardware and software that must be used. The choice of hardware and software depends on the type of data and the equipment's ability to symbolize that data in a way that meets the design standard for the graphic.

Ideally, the output device can produce point, line, area, and text types of data on film negatives, but until recently (1988) such a device was not available for producing the water summary. Vector pen plotters, such as Hewlett Packard models 7475 and 7586 and Calcomp 1039, can symbolize publication-quality lines, but not points, area fill, or text. Vector film plotters, such as Gerber and Kongsberg, can adequately symbolize points and lines, but not area fill or text. Raster film plotters, such as Scitex and Linotype, can symbolize points, lines, area fill, and text. The Scitex and the Linotype also can prioritize features so that, where items overlap, the most important features block out the less important ones on final negatives.

Manuscript approved for publication September 21, 1989.

[1]CA–TELLAGRAF is a registered trademark of Computer Associates International, Inc.

The type and complexity of data can limit the choice of hardware devices; consider a dot map showing population distribution. In areas of high population density, where many dots coalesce, a vector pen plotter cannot symbolize the data because the dots are of poor quality and the paper tends to wear out before all of the dots are drawn. However, the pen plotter can effectively symbolize lines, such as those used in a map of rivers. To symbolize a large quantity of dots, a film plotter is needed.

EVOLUTION OF GRAPHICS ILLUSTRATION IN THE NATIONAL WATER SUMMARY REPORTS

An examination of the hardware, software, and techniques used to prepare the graphics for the various National Water Summary reports illustrates how traditional cartographic production techniques can be merged with new digital procedures. Computer methods have evolved to keep pace with the changing types of data and design criteria for the graphics.

1983 National Water Summary

The 1983 National Water Summary (USGS, 1984) was prepared entirely by manual methods. Each State summary contains one water-issues map showing a few site locations, line symbols, area fills, and counties. Symbols for the maps were generated by traditional cartographic techniques and printed in combination with existing base maps.

1984 National Water Summary

The 1984 National Water Summary (USGS, 1985) was prepared mostly by manual methods. Map data showing principal aquifers, physiographic diagrams, and generalized cross sections were not available in computerized format. The design of these graphics did not warrant the use of computer methods because the data were not complex. Because data for areal distribution of ground-water withdrawals included 50 or fewer site locations per State map, the maps were produced manually.

To compile the data for the graphs of ground-water withdrawals, the authors accessed and updated USGS computer files. The illustrators used the data base and pen plotters to make the graphs. Unfortunately, the data base contained many errors that had to be corrected before the graphs could be drawn. To control the quality of the data in subsequent reports, data bases have been built from data supplied by the authors. In this way, the quality of the data and its relation to the graphic presentation are maintained, and a computerized data base is available for creating future illustrations.

1985 National Water Summary

Preparing the graphics for the 1985 National Water Summary (USGS, 1986) report involved a combination of computerized and manual methods. This report contains a large number of illustrations having a consistent design, and the data for plotting the illustrations were available in a computerized format. Manually preparing these graphics would have been tedious and time consuming. Film or paper positives were prepared by using vector pen plotters, adding the area fill and text by hand, and then creating printing negatives.

Graphs.—Each 1985 State summary contains three precipitation graphs, three runoff graphs, and six annual-flow graphs. These graphs were prepared by using CA–TELLAGRAF software, the Prime Computer, and Calcomp pen plotters for linework and traditional photomechanical techniques for printing negatives (fig. 1).

Authors transmitted the data for the graphs electronically to files on the Prime computer. The files were organized into several subdirectories. One subdirectory contained the incoming files. The original data set from each district was preserved. The working data sets were copied into another subdirectory for processing into plot files. The finished plot files were maintained in a third subdirectory. This organization allowed the many data sets to be tracked during processing and preserved the original data.

The basic format of each type of graph was designed, and generic CA–TELLAGRAF command files were written to match the design. One CA–TELLAGRAF file template was prepared for the precipitation graphs and another template was prepared for the runoff graphs. For these graphs a line editor was used to load station data into the template. Then the line editor was used to modify the elements of the graph specific to each station (maximum value on the *y* axis, station name). The discharge graphs were prepared by using a program called BANDS.F77. This program was used to calculate the 15-year moving average flow from the input data and to write the discharge and moving average flow data and the CA–TELLAGRAF commands to an output file. BANDS.F77 was used to calculate the maximum value of the *y* axis and to write the station name (K.J. Lanfear, USGS, written commun., 1985). BANDS.F77 is more flexible than the line editor because plot files can be regenerated from new data on command and the program automatically calculates site-specific information.

Each graph was previewed on a monitor to check the data and labels. The graphs were fine tuned to resolve any data problems and to position the labels so that they did not extend past the frame of the graph or overlap the data. Each CA–TELLAGRAF file contained several subplots—each subplot drew a color separate; that is, all of the elements for one color printed on one page. A final subplot contained registration tics that were plotted on every color separate.

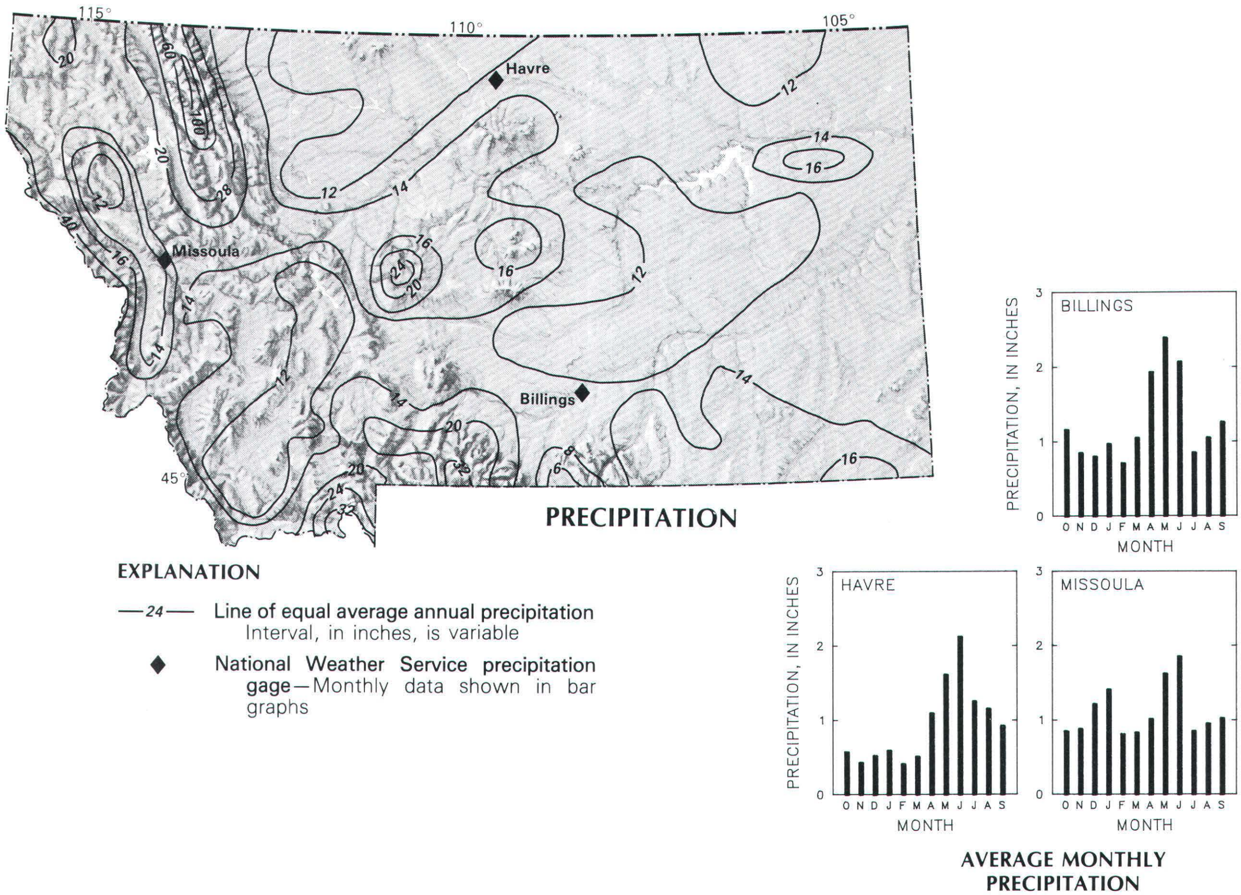


Figure 1. Graphs and a map from the 1985 National Water Summary originally generated at twice publication size on a pen plotter (U.S. Geological Survey, 1986).

Final plots were generated at twice publication size on a Hewlett Packard 7475 pen plotter that has refillable drafting pens. Bond paper or clay-coated paper resulted in better output than vellum. Once each color separate was plotted on paper, the image was photographically reduced by 50 percent, and the negatives were combined to make the final art. Reducing the image size gave finer linework and obscured irregularities in the ink.

Maps.—Each 1985 State summary contains maps of physiographic divisions, relative discharge, precipitation, runoff, and principal river basins. The physiographic and relative discharge maps were prepared manually, and the other maps were prepared by a combination of manual and digital techniques.

State outlines and rivers were derived from 1:2,000,000 digital line graph (DLG) files converted to ARC/INFO format. For display purposes, the river network was thinned by using ARCEDIT. The contour data for the precipitation and runoff maps were prepared from national data bases compiled especially for the National Water Summary. Both maps were digitized, and parts of them are included in the State summaries.

Final plots were generated at twice publication size on a Calcomp 1039 pen plotter, which has liquid ink drafting pens. These plots were photographically reduced to publication size. The type and area fill were done manually (fig. 1).

Discussion.—Using mechanical vector pen plotters had several disadvantages. No publication-quality area fill, points, or text could be symbolized on the maps. The quality of the linework was inconsistent because of irregularities in the ink. Area fill on the graphs was not uniformly dense and had to be fixed manually. The area fill and the text on the maps had to be done by hand because neither could be symbolized by the vector pen plotter without a loss in quality. An extra step was necessary to photograph the plots and create the negative separates. All of these factors combined required a great deal of effort by the graphics artists to finish the artwork.

Several advantages of using pen plotters were the low cost, ready availability, and direct interface of hardware to available software. Also, some of the graphics artists were more familiar with pen and ink media and could design the graphics destined for this type of output more efficiently.

1986 National Water Summary

The 1986 National Water Summary (USGS, 1988) marked the first extensive use of computer-generated graphics output directly to film negatives. Although traditional methods still were used to finish the artwork, much of the preparation was computerized. As in the 1985 report, the 1986 State summaries have many similar-looking graphs that lent themselves to being computer generated. The report also contains several maps showing large numbers of points and maps showing area-fill patterns that vector pen plotters cannot symbolize. The type of data and the design of the graphics made the use of a raster film plotter imperative to the production of the report. In cooperation with Environmental Systems Research, Inc. (ESRI), the USGS developed an interface from software available on the Prime to the Scitex electronic laser plotter, which can create prescreened film separates directly from computer plots (K.A. Ryden, ESRI, written commun., 1987). The film separates were prepared on the large format Scitex plotter, and then text was stripped in manually.

At the time the 1986 water summary was being prepared, no direct interface between CA–TELLAGRAF or ARC/INFO and the Scitex was available. To link ARC/INFO to the Scitex, the ARCPLOT file was converted to an intermediate plot file in the Standard Interchange Format (SIF). The SIF file then was converted to Scitex format. The specifications for the Scitex symbol set and the prioritization sequence had to be predefined on the Scitex by using a design set supplied by a cartographer (W.J. Danchuk, USGS, written commun., 1987). The ARCPLOT linecolor and markercolor linked the symbols in ARCPLOT to symbols on the Scitex. For example, linecolor 1 in ARCPLOT can specify a 100 percent black line, width 0.008 inches. Markercolor 1 in ARCPLOT can specify a filled dot, 100 percent magenta, 40 percent yellow, diameter 0.035 inches. For shade patterns, the color of the marker can convert to an area fill that fills the entire polygon; for example, markercolor 20 in ARCPLOT can specify a 10 percent cyan tint.

Graphs.—Each 1986 State summary contains five box-and-whisker diagrams showing water-quality data for different chemical constituents. Each of 24 different chemicals portrayed in the State summaries has a unique graphic format. The format of the graph for each chemical is the same throughout all State summaries to enable the reader to compare the graphs for one chemical from State to State.

The box plots were prepared by using CA–TELLAGRAF to generate the linework and text. The shading within the boxes was done manually to match the shading on the map of principal aquifers (fig. 2).

For the graphs, the authors submitted data files on the Prime computer. The files were organized in directories for incoming data, files for setting up the plots, and files for the finished plot. The CA–TELLAGRAF plots were created by

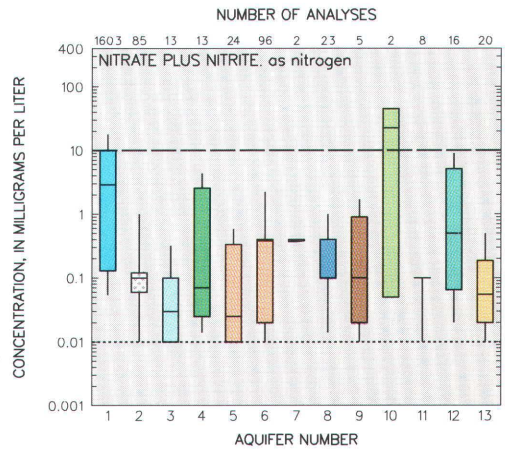


Figure 2. Box and whisker diagram from the 1986 National Water Summary generated on a large format electronic laser plotter (U.S. Geological Survey, 1988).

a program called GENPLOT.F77, which drew the box-and-whisker diagrams by using the submitted percentile values as input data. The individual box-and-whisker symbols were created in CA–TELLAGRAF by drawing two curves and connecting the data points to make the upper and lower rectangles and whiskers. GENPLOT.F77 read in the parameters for the unique format for each chemical from separate files set up for each chemical. These files listed the specific values for each chemical, such as the maximum and minimum value on the *y* axis, and drinking water standards. If the input data changed during the review process, the operator used GENPLOT.F77 to generate a new plot file from the revised parameters.

The CA–TELLAGRAF plots were designed originally to be output at twice publication size on a pen plotter because CA–TELLAGRAF had no interface to a raster film plotter. By using the USGS TELPOP program (S.D. Bartholoma, USGS, written commun., 1986), the CA–TELLAGRAF metafiles were converted to ARCPLOT files and scaled down by half using the CA–DISSPLA postprocessor, and then the ARCPLOT files were converted to SIF files for the Scitex. The Scitex outputs the linework for the plots by keying on linecolor to specify the characteristics of the line. The Scitex could not produce the shading patterns because converting software shading in CA–TELLAGRAF to hardware shading in ARCPLOT was not possible. Consequently, the linework was positioned manually and then used to make traditional photomechanical peelcoats for the area fill.

Maps.—Each 1986 State summary contains maps of principal aquifers, key locations, population distribution, and waste sites. The aquifer maps were prepared manually by using reproducible materials from the 1984 water summary. The linework, points, and screens for the other maps were done on the Scitex. The State outlines were derived

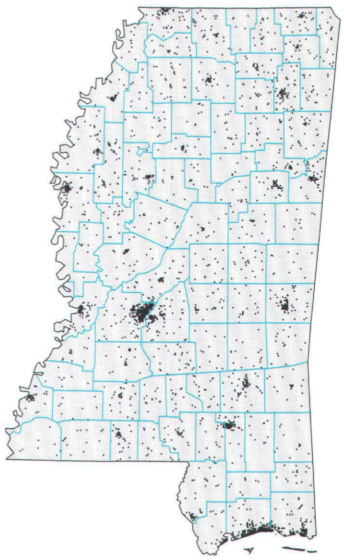


Figure 3. Population dot map from the 1986 National Water Summary generated on a large format electronic laser plotter (U.S. Geological Survey, 1988).

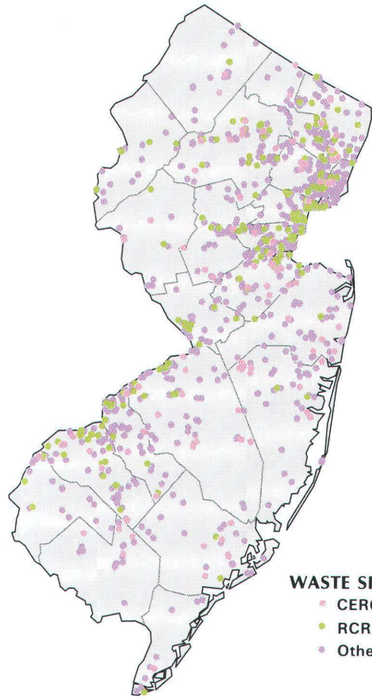


Figure 4. Hazardous-site map from the 1986 National Water Summary generated on a large format electronic laser plotter (U.S. Geological Survey, 1988). CERCLA, Comprehensive Environmental Response, Compensation, and Liability Act; RCRA, Resource Conservation and Recovery Act.

from 1:2,000,000 DLG files converted to ARC/INFO format that were edited for display purposes. The digital population data came from the Bureau of the Census, and the waste sites data were supplied in Prime files by authors of the State summaries. Prescreened negatives of the maps were generated at publication size and position by using the Scitex electronic laser plotter to symbolize the points, lines, and areas on the maps (figs. 3, 4). The type was stripped in by hand.

Discussion.—Producing the graphics with the Scitex plotter made it possible to portray lines, points, and area patterns that could not have been depicted by using a pen plotter. Prioritizing the waste-site data was done by the Scitex automatically. Outputting directly to film saved the extra step of photographing the pen plots. Because lines are drawn with a beam of light that is not subject to mechanical variations, the linework was of a consistent high quality. However, the laser is extremely precise, and the fine detail exposes mistakes that are disguised by pen and ink methods. Another drawback to the use of the Scitex is the numerous symbol specifications that must first be defined. The specifications cannot be updated easily to accommodate changes in or additions to the symbol set. No proofing device was available to check the plots before sending them to the Scitex. Therefore, traditional proofing methods had to be employed to check the plots after the negatives came off the Scitex. Translating the CA–TELLAGRAF plots to ARCPLOT introduced errors. The software-generated text on the boxplots was of poor quality because the CA–DISSPLA postprocessor used to scale down the plots omitted some of the vectors in the letters.

1987 National Water Summary

Computer-generated graphics are almost exclusively in the 1987 National Water Summary (USGS, 1990), although traditional proofing and finishing techniques are still required. Most of the data for the 1987 report already existed in the National Water-Use Information Program's data base on Estimated Use of Water in the United States (EUOWITUS). Routines to access the water-use data base were adapted to generate the input data needed for the water summary report instead of having the authors of the State summaries provide the data. However, the authors did supply data on water withdrawals by aquifer. The 1987 report contains graphs, maps having area fill and large numbers of point symbols, and pie diagrams. A Linotype L300 imagesetter was purchased to process the raster film output of points, lines, text, and area fill. The L300 is compatible with the PostScript page description language and can produce Adobe typeset fonts and prescreened negatives. Preparing the illustrations for the 1987 water summary requires PostScript files to be output directly to film negatives from the L300.

ARC/INFO has a PostScript postprocessor that converts an ARCPLOT file to a flat file of PostScript commands and generates one PostScript file for each color separation. An interpreter in the output device translates the PostScript code to the device-specific instructions that form the graphic at the resolution of the device. Because PostScript is device independent, the same plot can be sent to a low-resolution device, such as the Apple LaserWriter, for proofing and to the high-resolution L300 for final negatives. Two methods of proofing the plots are available. The first is to send the image to an Apple LaserWriter connected by an asynchronous multiline controller to the Prime, and the second is to download the PostScript files to a Macintosh and print them on an Apple LaserWriter connected to the Macintosh.

Specifications for the L300 symbols are set up in a flat translation file on the Prime. The ARCPLOT pen number links the symbols in ARCPLOT to the symbols in the translation file. The PostScript postprocessor converts the ARCPLOT symbols to PostScript symbols for lines, areas, and text, which are output by the L300. For example, linecolor 1 can specify an 80 percent black line, width 0.004 inch. The same symbol will specify an 80 percent black area fill if called as shadesymbol 1. Marker symbols such as circles currently cannot be defined as hardware characters. Symbols are stroked characters generated by the software.

Graphs.—Each 1987 State summary contains two graphs (population growth and reservoir storage) and a schematic detailing water use and disposition. The format of the graphs is the same throughout the State reports. The population and reservoir storage data were in existing computer files on the Prime. No special file processing routines were needed. The EUOWITUS data base was the data source for the water-use diagram.

The population and reservoir storage graphs were prepared by using CA–TELLAGRAF, which generated the linework and text. At the time the graphs were created, CA–TELLAGRAF had no PostScript driver. CA–TELLAGRAF metafiles were converted to ARCPLOT files by using the TELPOP program, and then the ARCPLOT files were converted to PostScript by using the ARC/INFO postprocessor (fig. 5).

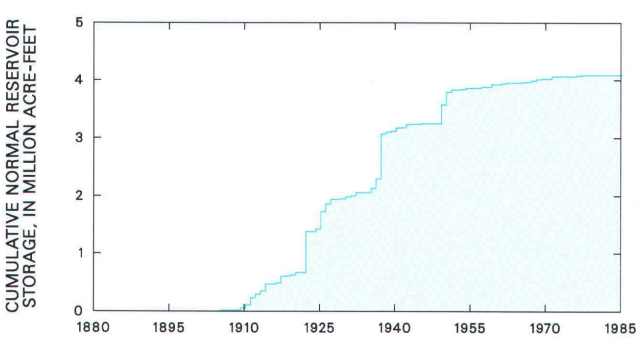


Figure 5. Graph from the 1987 National Water Summary generated on an electronic imagesetter (U.S. Geological Survey, 1990).

CA–DISSPLA routines accessing the EUOWITUS data base prepared the linework for the water-use diagram. CA–DISSPLA metafiles were converted to ARCPLOT files by using TELPOP. Another program converted the ARCPLOT files to polygon coverages in ARC/INFO (K.A. Ryden, oral commun., 1988). The polygons were labeled so that shade symbols were assigned to the areas, and new ARCPLOT files were created from the polygon coverages. These plot files were converted to PostScript by the ARC/INFO postprocessor. Text labels were applied manually and the computer-generated text was used as a data check.

Maps.—Each State summary contains choropleth county maps, maps showing river basin or aquifer boundaries, and a set of pie diagrams. Because a national set of digital aquifer boundaries is not available, the aquifer boundaries were prepared manually. The linework, screens, and some of the text on other parts of the maps were done on the L300. Most of the text was positioned by hand.

The ARC/INFO coverages derived from 1:2,000,000 DLG's were used to define State and county outlines. The river basin boundaries were drawn from hydrologic cataloging units or combinations of cataloging units available in ARC/INFO format at the 1:2,000,000 scale. Linework and shaded fill for the river basins and the shade patterns for county choropleth maps of water use (fig. 6) were prepared on the L300.

Each pie chart of water withdrawals by water-use sector was an ARC/INFO polygon coverage generated by a FORTRAN program that read the input data and converted it to the ARC GENERATE command format. The linework and shaded fill for the pie charts and some of the text labels came from the L300. Many labels could be positioned properly by ARCPLOT, but others were not located correctly for a good visual presentation. A cartographer adjusted the labels by hand and used the computer output as a data check. Prescreened negatives of the maps were generated at publication size and position by using the L300 laser plotter. Most of the type was stripped in by hand. The explanation block for the pie diagrams was set up on the L300.

Discussion.—Producing the graphics on the L300 conveyed the inherent advantages of a laser plotter—the ability to go directly to film negatives and to produce hardware text (characters produced by the output device), fill patterns, and precise linework. The amount of manual effort needed to process the graphics continues to decrease. Changing the symbol specifications for the L300 is much easier than changing the symbols on the Scitex. Being able to change the specifications easily gives more flexibility in changing the design of the graphics. Any PostScript compatible printer can serve as a black and white proofing

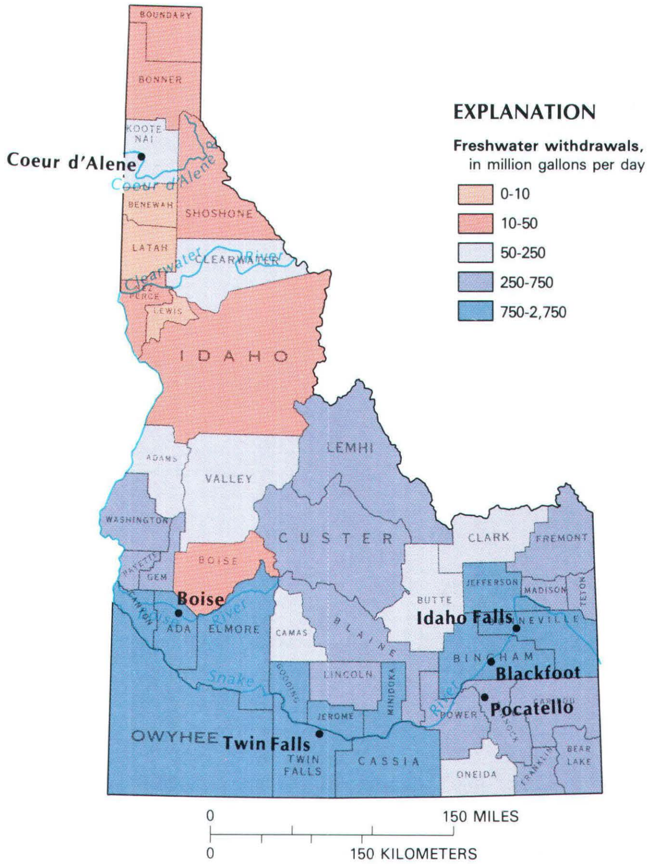


Figure 6. County choropleth map from the 1987 National Water Summary generated on an electronic imagesetter (U.S. Geological Survey, 1990).

device for files destined to be output on the L300. Manual methods still are needed to strip in punch registry and to prepare color proofs. The use of color electrostatic plotters (raster devices) for proofing is being explored.

SUMMARY

No direct link from computer graphics software running on the Prime computer to pre-press devices was available for producing the illustrations for the National Water Summary until the Linotype L300 raster imagesetter became available to output PostScript files on film negatives. Stopgap measures were used to produce the illustrations until the interface from ARC/INFO to PostScript was developed. One of these measures was to produce positive output on paper by using vector pen plotters. The pen plot then was photographed to make the negatives for the printer. This method was the least efficient. Another measure was to use an intermediate file format (SIF) that could be understood by both ARC/INFO and the Scitex electronic laser plotter. This indirect link between the graphics software and the output device was more efficient than using the vector pen plotters but still was not a viable solution. Using the interface from ARC/INFO to PostScript devices was the best method for producing the illustrations for the National Water Summary.

REFERENCES CITED

Moody, D.W., and Lanfear, K.J., 1988, Use of geographic information for long-range water planning and management, *in* Niemann, B.J., and Sullivan, J.G., eds., Seminar on multipurpose land information systems: Application of information technology for natural resource planning, management, and monitoring: Madison, Wis., University of Wisconsin, Wisconsin Land Information Report 3, Institute for Environmental Studies Report 132, p. 59–70.

U.S. Geological Survey, 1984, National water summary 1983—Hydrologic events and issues: U.S. Geological Survey Water-Supply Paper 2250, 243 p.

———1985, National water summary 1984—Hydrologic events, selected water-quality trends, and ground-water resources: U.S. Geological Survey Water-Supply Paper 2275, 467 p.

———1986, National water summary 1985—Hydrologic events and surface-water resources: U.S. Geological Survey Water-Supply Paper 2300, 506 p.

———1988, National water summary 1986—Hydrologic events and ground-water quality: U.S. Geological Survey Water-Supply Paper 2325, 560 p.

———1990, National water summary 1987—Hydrologic events and water supply and use: U.S. Geological Survey Water-Supply Paper 2350, 553 p.

CHAPTER C

Generating Color Separations of Geologic Maps On the Computer

By ALEX V. ACOSTA and JANET M. BARRETT

U.S. GEOLOGICAL SURVEY BULLETIN 1908

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES 1990

CONTENTS

FIGURES

TABLES

Generating Color Separations of Geologic Maps On the Computer

By Alex Acosta *and* Janet M. Barrett

Abstract

Manual and computerized methods of producing color separations of geologic maps start with an author's original compilation (brownline) of a geologic map. The various steps performed in the manual procedure are conceptually duplicated in the computer. The brownline is prepared and scanned to produce a raster image. The raster image is then processed by using a line-thinning algorithm to produce one-pixel boundaries of lines that enclose polygons of geologic units. The image is then edited on a computer display to ensure that all polygons are closed. Further necessary processes include tagging, vectorizing, and filling in the polygons. The vectorized polygons can be input to a geographic information system and used in that environment with other spatial data sets. The filled-in polygons represent a computerized raster peelcoat that is peeled, screened, and composited in the computer to generate the yellow, magenta, cyan, and black separations that are required for printing. The computerized procedure can be implemented on a low-cost workstation that can provide the basis for a desktop mapping system.

INTRODUCTION

The color of an area on a lithographically printed geologic map is controlled by the size of minute color dots (Wesner, 1974) that cover a percentage of the area. The color dots are yellow, magenta, cyan, or black and can be combined to make other colors. Four lithographic plates, one for each color of printing ink (yellow, magenta, cyan, or black), are used to print a map. The plates, which contain the minute dots, are derived from film transparencies called color separations. The color separations are photographically printed on metal. Then, a resist is applied to protect the dots, and the surrounding metal is removed by etching to produce the four printing plates. Maps printed by the U.S. Geological Survey (USGS) contain 120 dots per inch. The repetitive dot pattern is rotated relative to a horizontal base line for each of the four colors to eliminate moire patterns (Bryngdahl, 1974) in the printed map. The dot harmonics on the printed map are so small that the eye cannot resolve the patterns; thus, the visual effect is of continuous color.

Manuscript approved for publication October 7, 1988.

Purpose

The manual procedure currently used to produce color separations is labor intensive and requires a significant amount of photographic materials. This procedure can be computerized to save labor and material costs. The computerized method can be implemented on a low-cost, 32-bit workstation that is a part of a desktop mapping system.

Acknowledgments

The authors thank the staff of the USGS in Flagstaff, Ariz., for their discussions and assistance on the manual method of producing color separations for geologic maps. Specifically, we thank Lynda Bellisime, Jana Ruhlman, and Kirsten Rea for testing the computerized procedure and producing several of the illustrations in this paper. Also, we appreciate the expedient and helpful reviews by Roger Carroll, Hugh Thomas, and Phil Davis.

MANUAL PROCEDURE

The manual procedure for generating color separations is as follows:

- generate scribecoat,
- generate color guides,
- generate peelcoats,
- generate open-window negatives,
- generate color separations, and
- generate color proof.

Generate Scribecoat

The linework in the author's original compilation, or brownline (fig. 1), generally is of various widths and does not meet mapping standards. To comply with the standards of uniform line weight, the linework is scribed on a scribecoat (fig. 2). A scribecoat is an actinically opaque, stable polyester film that is coated on one side with a surface especially made for scribing. The brownline is photographically duplicated or composited on the scribecoat prior to scribing. Other bases, such as shaded relief or topography, also may be composited on the scribecoat. The draftsman uses the composited bases as an aid in scribing the geologic unit lines. The lines must create enclosed polygons. Lines of constant width are produced by using scribers (fig. 3), which have precisely sized points; scribing is tedious.

Generate Color Guides

After the scribecoat is complete, color guides (fig. 4) are produced. Map color guides are necessary for peelcoat production to determine which geologic-unit polygons should be peeled. A color guide is generated by compositing the scribecoat on a mylar sheet, which has a coated surface that can be marked by pen or pencil. The number of color guides generated is based on the number of geologic units on the map; a map of 21 geologic units would require 3 color guides. No more than nine geologic units are represented per guide. Each geologic unit is assigned a distinct color. Color guides are visually checked for errors, such as unscribed geologic units or one color used for two geologic units. If errors are detected, the scribecoat and color guides are retouched.

Generate Peelcoats

The peelcoat (fig. 5) is a polyester film coated on one side with an actinically opaque, light-sensitive, peelable surface. A peelcoat is produced by photographically transferring the scribecoat information onto the coated film. One peelcoat is required for each of the color percentages that are used to produce the colors on the map. For example, if a geologic map contains units that require 2 different percentages (dot sizes) of yellow, 5 of magenta, 6 of cyan, and 4 of black, then 17 peelcoats are required to produce the 4 color separations.

The color guides and the brownline are used as guides to peeling all the polygons (fig. 5) that are to have the same percentage of a particular color ink. The remaining lines that delineate polygons of different colors are masked by a wipe-on coating that makes them opaque to subsequent photographic processing. The resultant peelcoat, which contains open areas of polygons, is called an open-window negative.

Generate Open-Window Negatives

One open-window negative is required for each percentage of each color. Next, an open-window negative for a particular percentage is composited with a line screen (fig. 6*A*) of the corresponding percentage to produce a screened polygon area (fig. 6*B*). Then, another open-window negative for a different percentage is composited with its corresponding line screen and the previously screened areas to generate an intermediate composite that contains two sets of differently screened polygons. Similarly, the remaining open-window negatives are composited with their corresponding line screens and previously screened areas to generate the final composite.

Generate Color Separations

The final composite of open-window negative and linescreens is chemically developed to produce the color separation (fig. 7). Four color separations (yellow, magenta, cyan, and black) are usually generated and are used by a printer to generate the lithographic plates used to print the geologic map.

A line screen is a film sheet containing dots of a particular size, spaced at a particular distance, that repeat in a specified direction relative to a horizontal baseline. The size of the dots on the screen determines how much ink of a particular color will appear on the printed map. The size is usually called a percentage, such as 8 percent yellow, 13 percent magenta, or 20 percent cyan, and refers to the percentage of ink that will cover a unit square. The spacing of the dots or the side dimension of the unit square is related to the resolving power of the eye to distinguish the dot pattern. The eye cannot resolve the dot pattern if the spacing is small; thus, the visual effect of continuous color on the printed map is created. Unfortunately, small spacing requires high-quality paper and ink that increase the cost of the printed map.

A precise orientation of the dot pattern is necessary to prevent moire patterns in the printed map. It has been determined experimentally that dot patterns of dark colors such as magenta, cyan, and black must be 30° apart on the printed map. The angles most commonly used are 90° for yellow, 75° for magenta, 105° for cyan, and 45° for black.

Generate Color Proof

The color separations are used to produce a color proof that is similar to a printed map. The colors on the proof closely resemble printed colors, and a proof can be mistaken for a printed map. The proof is checked to ensure that all geologic units have the proper color. Due to the complexity of the map, it is usually necessary to generate at least three color proofs (Roger Carroll, USGS, oral

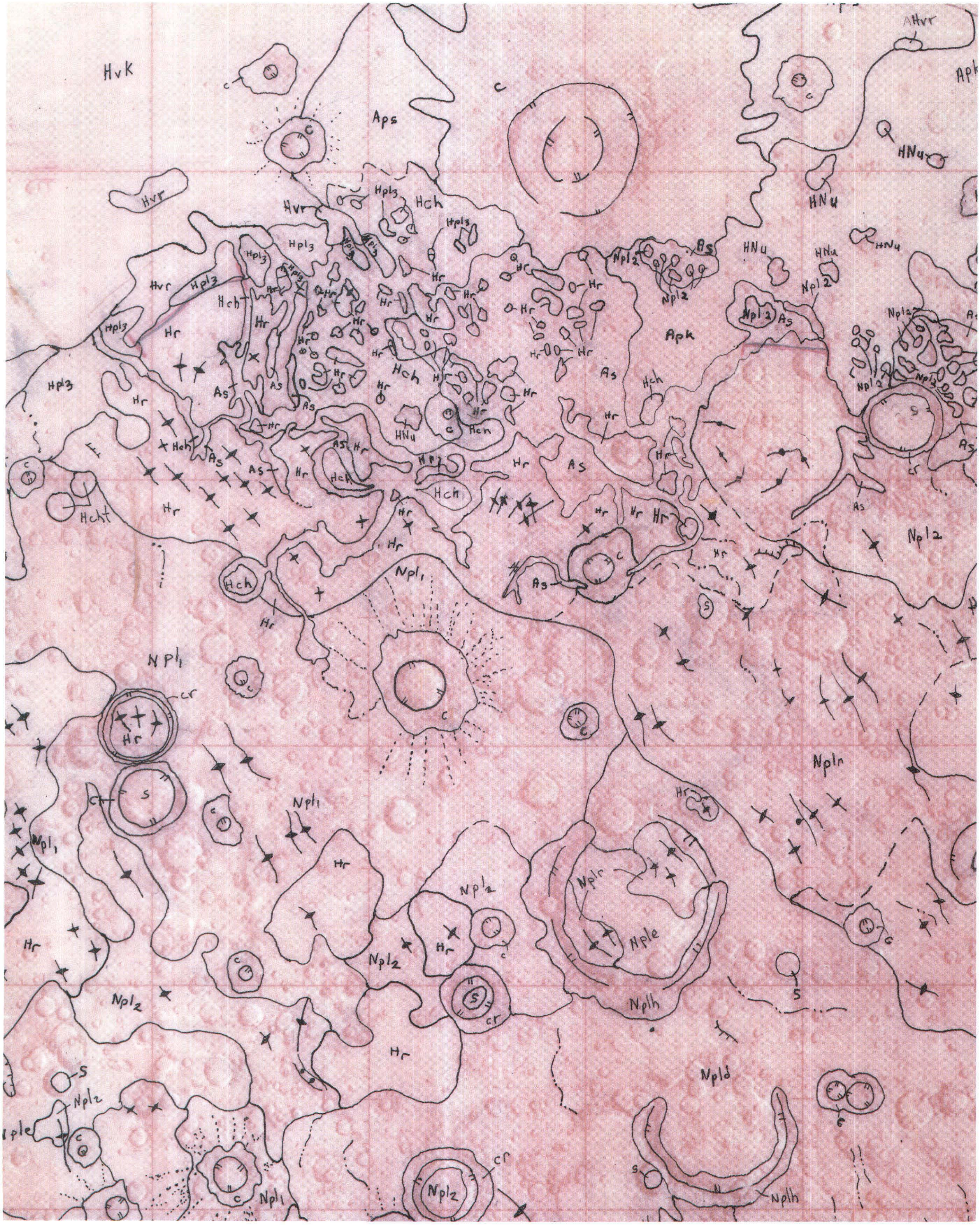


Figure 1. Photographically enlarged subarea of author's original line drawing on a shaded-relief base map (Greeley and Guest, 1987). North is to the top.

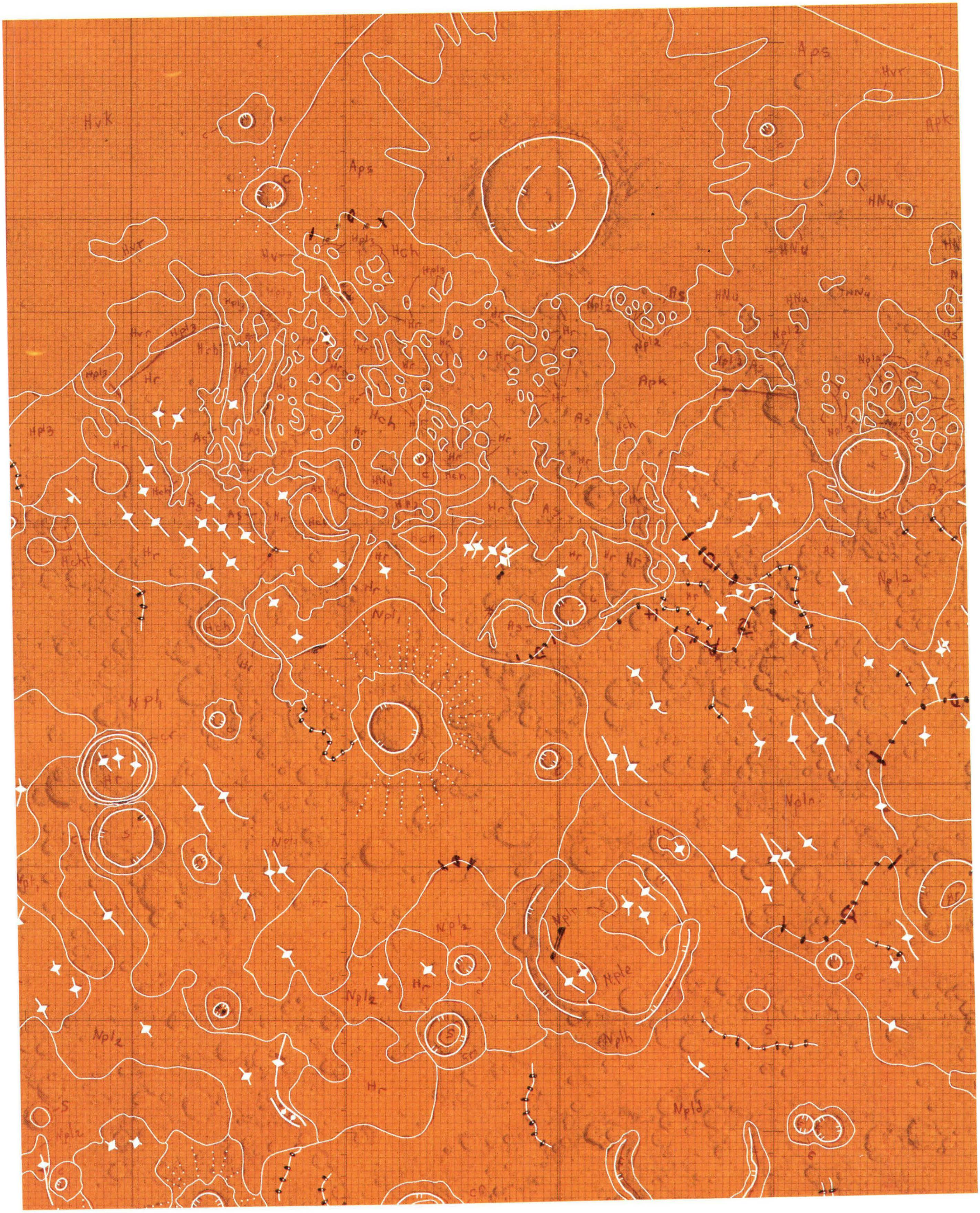


Figure 2. A scribecoat produced by compositing the author's original line drawing and a shaded-relief map on the scribecoat to assist the draftsman in scribing.

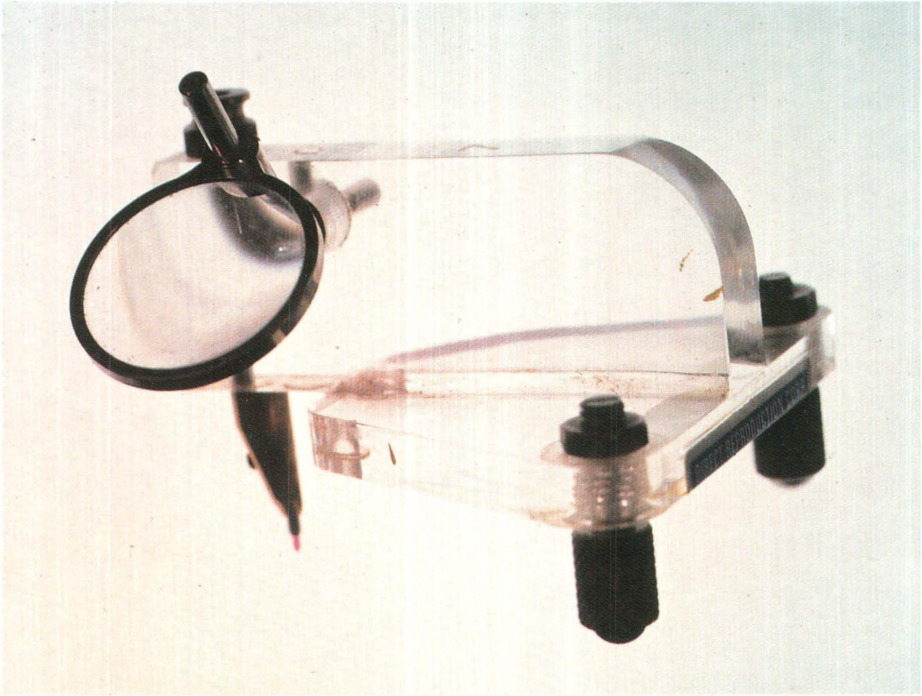


Figure 3. A scriber. Precisely sized points are used with the scriber to produce lines of constant weight on a scribecoat.

commun., 1988) to correct the errors on a single geologic map. Color proofs are prepared in a photographic laboratory by using whirl-on light-sensitive color emulsions and proofing stock, which are exposed to ultraviolet light (UV) through a color separation negative and then developed with water. The color emulsion remains in the areas exposed by the UV light. The color of each emulsion corresponds to the color of each separation.

COMPUTERIZED PROCEDURE

The computerized procedure for generating color separations resembles the manual method but requires less time and little photographic processing. A computerized version of a scribecoat is generated by raster scanning a prepared version of the brownline. The resultant raster image contains multishaped polygons of geologic units. To associate a polygon with a geologic unit, each polygon must be identified by a number (a tag value placed within the polygon; see Tagging Polygons) that is unique to that geologic unit. The lines of all polygons are then converted to vectors (see Vectorizing Polygon Boundary Data). The vector data are used to generate a raster image that contains polygons whose areas have been filled with the corresponding tag value. The vectorized polygon data are a useful byproduct of the computerized procedure and can be translated into a format (for example, Digital Line Graph) that is acceptable to a geographic information system. The raster image of filled-in polygons is a computerized peelcoat that can be peeled, screened, and composited digitally to generate the required color separations. The computerized procedure is as follows:

- prepare brownline,
- scan prepared brownline,
- edit scanned image and convert to binary,
- tag and vectorize polygons,
- generate computerized peelcoats,
- generate and screen computerized open-window negatives,
- generate computerized color separations, and
- generate color proof.

Prepare Brownline

The brownline supplied by the author (fig. 1) is not suitable for raster scanning (Burrough, 1986; Leberl and Olson, 1982) because it contains special symbols and descriptive text that does not meet USGS standards. A negative copy of the brownline can be manually edited quickly with an opaquing pen to prepare it for scanning (fig. 8). A positive reproduction of the opaqued negative is further edited with a pen to connect all dashed lines that represent geologic boundaries, to darken lines that are too light, and to define the outside borders of the map to ensure that all geologic units are enclosed polygons (fig. 9). Ideally, start with a version of the geologic map containing

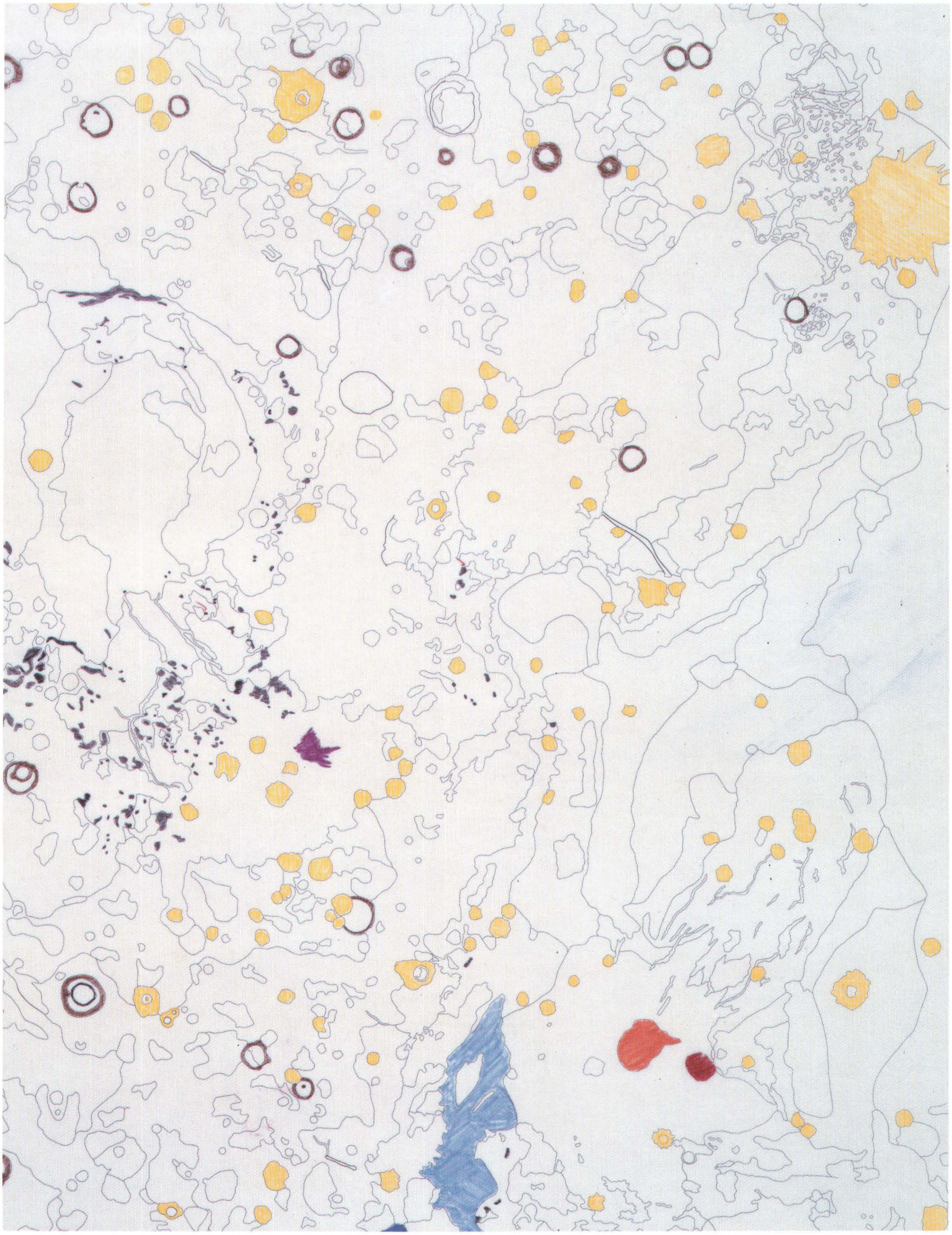
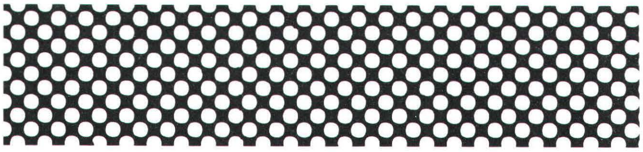
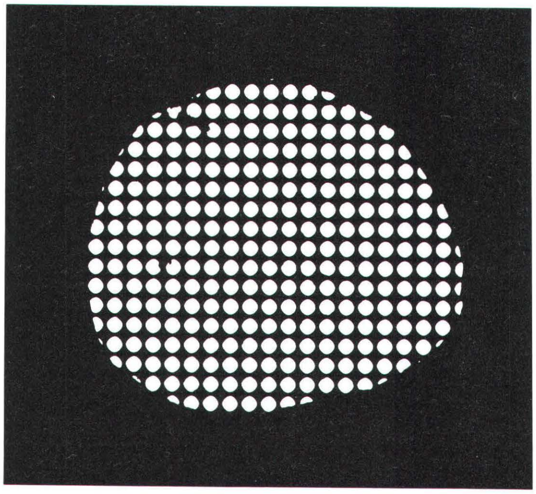


Figure 4. A color guide. Geologic units are colored with pencils. The guide is used as an aid in peeling polygons.

Figure 5. A peeled peelcoat. The white areas, produced by peeling an overcoating, are polygons representing geologic units. Lines delineating other polygons are covered with a wipe-on coating.



A



B

Figure 6. *A*, A line screen used to control the amount of ink that will be applied in a particular area. The dot diameter is varied to control the amount of ink. The spacing between dots is very small (screen magnification ×12). *B*, Screened open-window negative produced by compositing line screens with open-window negatives.

only polygons of geologic units. The preparation for scanning is minimized, and no opaquing is necessary. If all line weights are consistently dark, the only preparation necessary is to connect the dashed lines of the boundaries and to define the outside borders of the polygons.

Scan Prepared Brownline

The prepared version of the brownline is raster scanned. The brownline is usually large; therefore, a large-format (40 by 40 in. or larger) scanner is necessary. If the only scanner available is small format, the prepared brownline must be photographically reduced to the maximum format allowed by the scanner. Some maps cannot be reduced for scanning because they contain small polygons that will reduce to dots or will be lost in the reduction. Photographic reduction can be applied to any map to reduce the data volume.

The sampling interval or raster setting used to scan the prepared brownline should be small enough to sample the area enclosed by the smallest polygon at least once but large enough to minimize the data volume. For example, if the smallest polygon encloses an area that is 0.5 mm (0.02 in.) wide at its narrowest point in a horizontal or vertical direction, then the sampling interval should be less than or equal to half of 0.5 mm. The raster data produced by the scanner will consist of 0's and 1's if the scanner is binary, or the data will consist of integers from 0 to 255 if the scanner is continuous tone.

Edit Scanned Image and Convert to Binary

The rasterized brownline requires visual evaluation and possibly interactive editing before it can be processed further. If the raster setting selected for scanning is too large, small polygons may appear as black spots (fig. 10*A*) and a close pair of lines may appear merged and thus erroneously define a new polygon or obscure a valid polygon (fig. 10*B*). Rescanning the prepared brownline at a smaller raster setting will resolve the associated problems. Besides the raster setting problems, lines defining valid polygons may be disconnected due to opaquing. Interactive editing on a display is the easiest method for connecting broken lines of valid polygons.

Numerous scanners are available that can generate binary (0's and 1's) or continuous-tone (usually 0 to 255) data. If the scanner generates continuous-tone data, the raster image must be converted to a binary image for subsequent processing. Continuous-tone scanners generate variable numbers for lines because of the raster setting and variable line densities on the brownline. These variable numbers are analyzed, and a threshold value is selected that will not cause a break in any line when the continuous-tone data are converted to binary data. In the conversion, values from 0 to the threshold are assigned to 0. The converted binary image contains 0's that define lines and 1's that define background. A line-thinning algorithm (Naccache and Shinghal, 1984) is applied to the binary image to convert thick lines (fig. 11*A*) to thin lines (fig. 11*B*). Thin lines are required for the vectorizing process (see Vectorizing Polygon Boundary Data).

Tag and Vectorize Polygons

The binary image has thin lines and is analogous to a scribecoat. All lines have the same weight, one pixel defines the line boundary, and all geologic units are defined by polygons. The polygons in this image are obvious when displayed, but processing cannot continue until polygonal areas are associated with geologic units. This association is done by tagging each polygon with an 8-bit value that has been chosen to represent a particular geologic unit (see Tagging Polygons). For example, if a set of polygons in a binary image represents a particular alluvial unit, a number

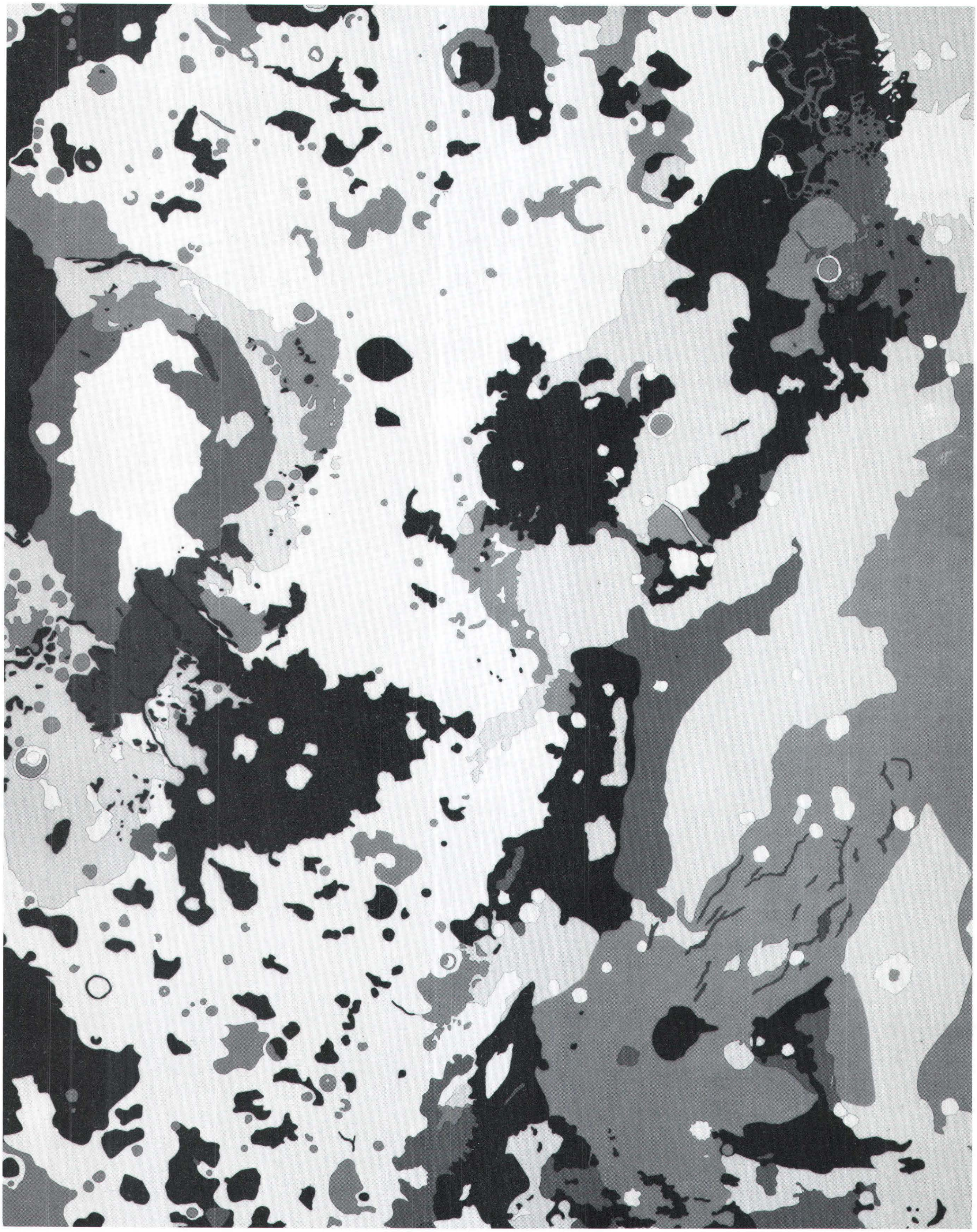


Figure 7. A photographically reduced version of a yellow color separation. The gradation in gray tone exemplified by this version of the original is induced by the limits in the eye's resolving power to detect the dots. Light-tone areas will be yellow in the printed map.

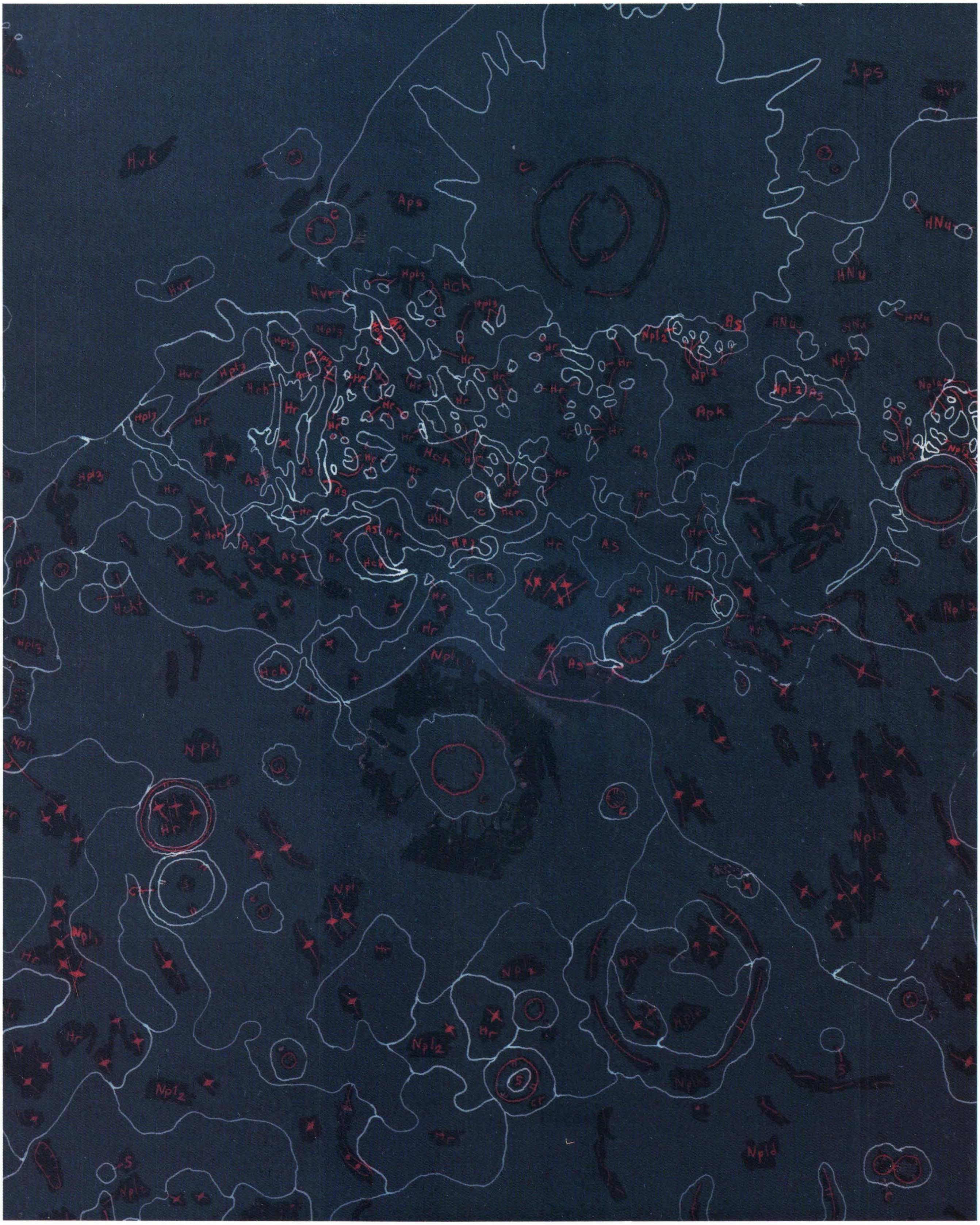


Figure 8. An opaqued negative. The text and special symbols in the author's original compilation do not delineate geologic units and are not in color on the printed map. They are manually opaqued prior to scanning. Manual editing in this case is more efficient than computerized editing.

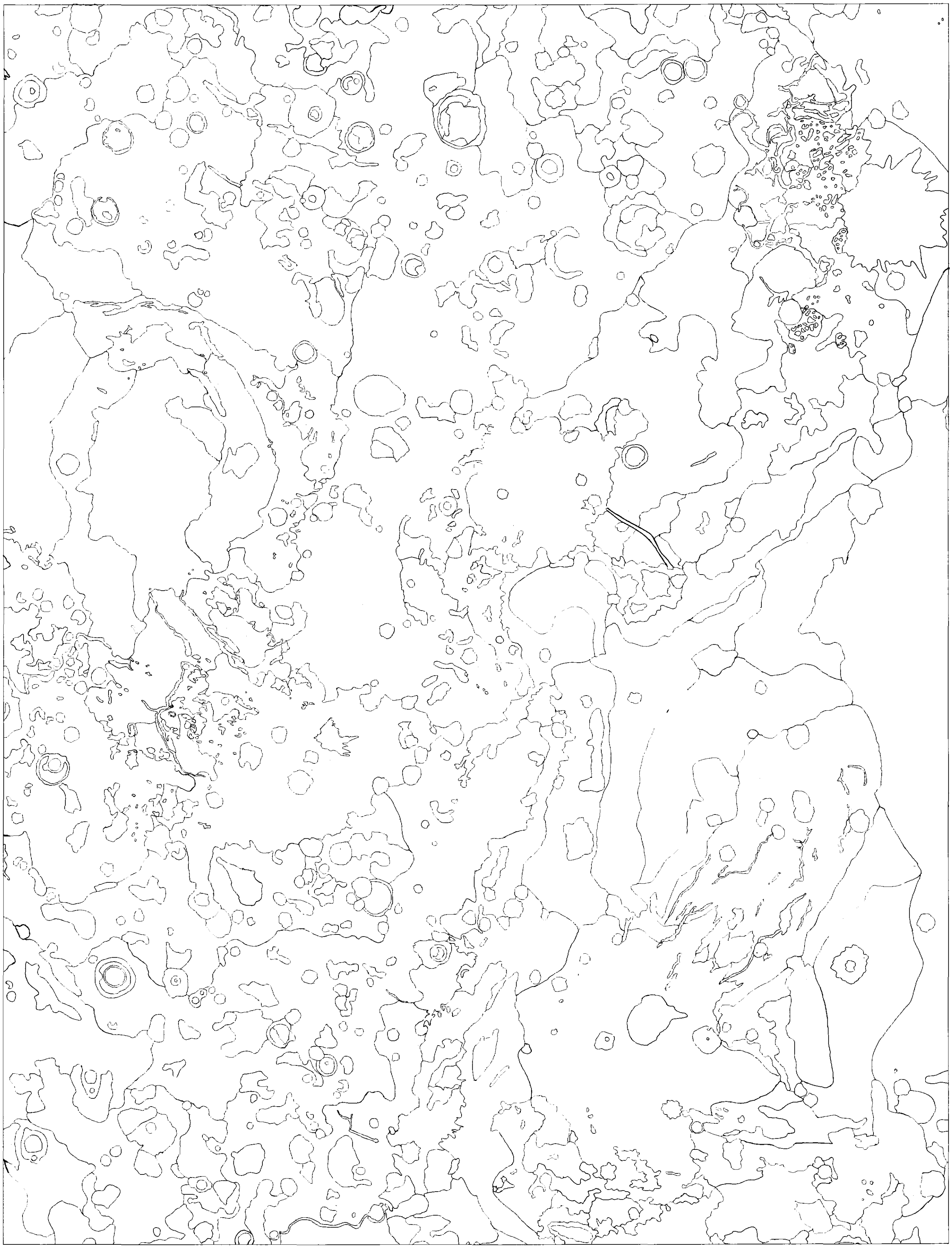
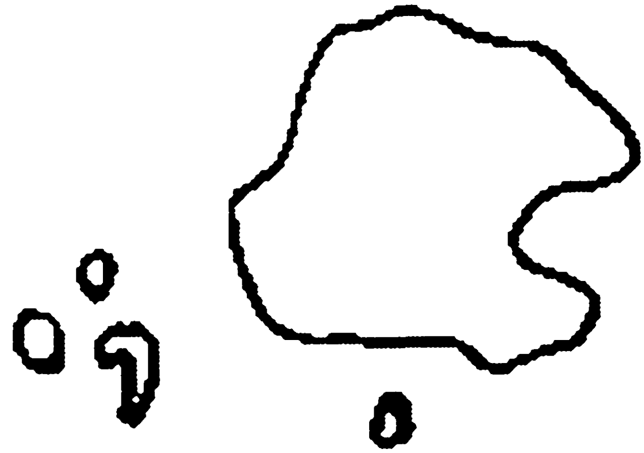


Figure 9. A prepared brownline ready to be scanned. All linework delineates geologic units.

A

B

Figure 10. Scanning errors caused by too large a sampling interval. *A*, Polygon in center appears as a black object. *B*, Line pairs in center are merged.



A

B

Figure 11. Raster image lines. *A*, Thick lines defined by more than one pixel. *B*, Thin lines defined by one pixel.

is chosen (for example, 250) to represent that alluvial unit, and each polygon in the binary image that represents that alluvial unit is tagged with 250. Each tag value has associated line and sample values, or pixel coordinates, that are used as a starting point to vectorize the related polygon. All tagged polygons (some may be untagged due to operator error) are vectorized; that is, the line boundary of a polygon is converted to pixel coordinates (see Vectorizing Polygon Boundary Data). The generated vector coordinates are subsequently used to find untagged polygons and to generate a computerized peelcoat. The coordinates can also be used by translator programs to produce data that are acceptable to a geographic information system.

Next, a binary image is generated from the vector coordinates to check for untagged polygons. This image is identical to the line-thinned image except for those polygons that are erroneously untagged. The untagged polygons can be easily detected on a color cathode ray tube (CRT) display if the line-thinned image is displayed in the green channel and the binary image is displayed in the red channel. The untagged polygons appear as red on the display. These untagged polygons are then tagged and vectorized, and the vector data are appended to the first vector data set to form a complete set. The vector data are then used to produce a raster image that is a computerized peelcoat (see Vectorizing Polygon Boundary Data).

Generate Computerized Peelcoats

A computerized peelcoat consists of polygons whose areas have been filled. All polygons that are associated with a geologic unit have their areas filled with the same value. Various derivative products can be generated from the peelcoat. For example, all the numbers in a peelcoat may be reassigned to add contrast or remain the same to produce gray-tone photographic prints of the geologic map (fig. 12). Also, the peelcoat can be colored on a color CRT display to check for tagging errors and to produce a color photographic print of the geologic map (fig. 13). Because all polygons in a peelcoat are filled with an identifying number, subsets of numbers can be reassigned in the computer to generate open-window negatives.

Generate and Screen Computerized Open-Window Negatives

A masking operation is applied to the peelcoat to generate a computerized open-window negative of those colors of geologic units that will include the same percentage of a particular ink. For example, if the colors of two different geologic units include 8 percent cyan, then the two numbers associated with the geologic units are reassigned to 1 and all other numbers are reassigned to 0. The result is a computerized version of figure 5. If the filmwriter (a device used to produce film transparencies from digital data) used to produce the color separations provides halftone capabilities (a required percentage of dots at a specified angle), then the computerized open-window negative is used as input by the filmwriter to produce appropriately screened areas on film. For example, the filmwriter can produce on film an 8 percent dot pattern angled at 105° in the areas that are open in the computerized open-window negative. If the filmwriter has no halftone capabilities, then the computerized open-window negative must be rotated and screened in the computer. First, the negative is rotated appropriately depending on the required color (90° for yellow, 75° for magenta, and so on). This rotation is necessary because the dot patterns on computerized screens are generated in a horizontal pattern (see Line Screens and Continuous Color), and an angled dot pattern is needed to avoid moire patterns on the printed map. The open-window negative is then composited by using the AND operation (table 1) with the appropriate computerized screen to produce a screened version of the negative similar to figure 6*B*.

Generate Computerized Color Separations

The screened open-window negative is one part of a color separation and is generated for each percentage of a particular color to generate a set of screened polygon areas for that color. The screened areas are then composited by using the OR operation (table 1) to produce a computerized color separation. Because the compositing is done in the computer, no registration problems arise. The computerized steps of peeling, screening, and compositing are repeated for each of the remaining colors to produce the required separations.

Table 1. The AND and the OR operations

AND			
Case	*x*	*y*	*x* AND *y*
1	0	0	0
2	0	1	0
3	1	0	0
4	1	1	1

OR			
Case	*x*	*y*	*x* OR *y*
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	1

The computerized color separations are written to film by a filmwriter to produce the transparencies used by a printer to print the geologic map. The filmwriter used to produce the transparencies must be large format (40 by 40 in. or larger) because the map scales used by the USGS are large. Also, the output raster setting of the filmwriter must allow for writing the color separations at 120 dots per inch (see Line Screens and Continuous Color).

Generate Color Proof

Because the data is in digital form, other options, besides the manual whirl-on procedure, exist for generating color proofs for error checking. One option is to generate an enlarged color print from film transparencies of the computerized color separations. Visually, the colors on the proof should approximate the colors that will be printed on the final map. Another choice is to produce a color proof by using the computerized color-separation data on a color printer, a computer peripheral that uses yellow, magenta, cyan, and black inks.

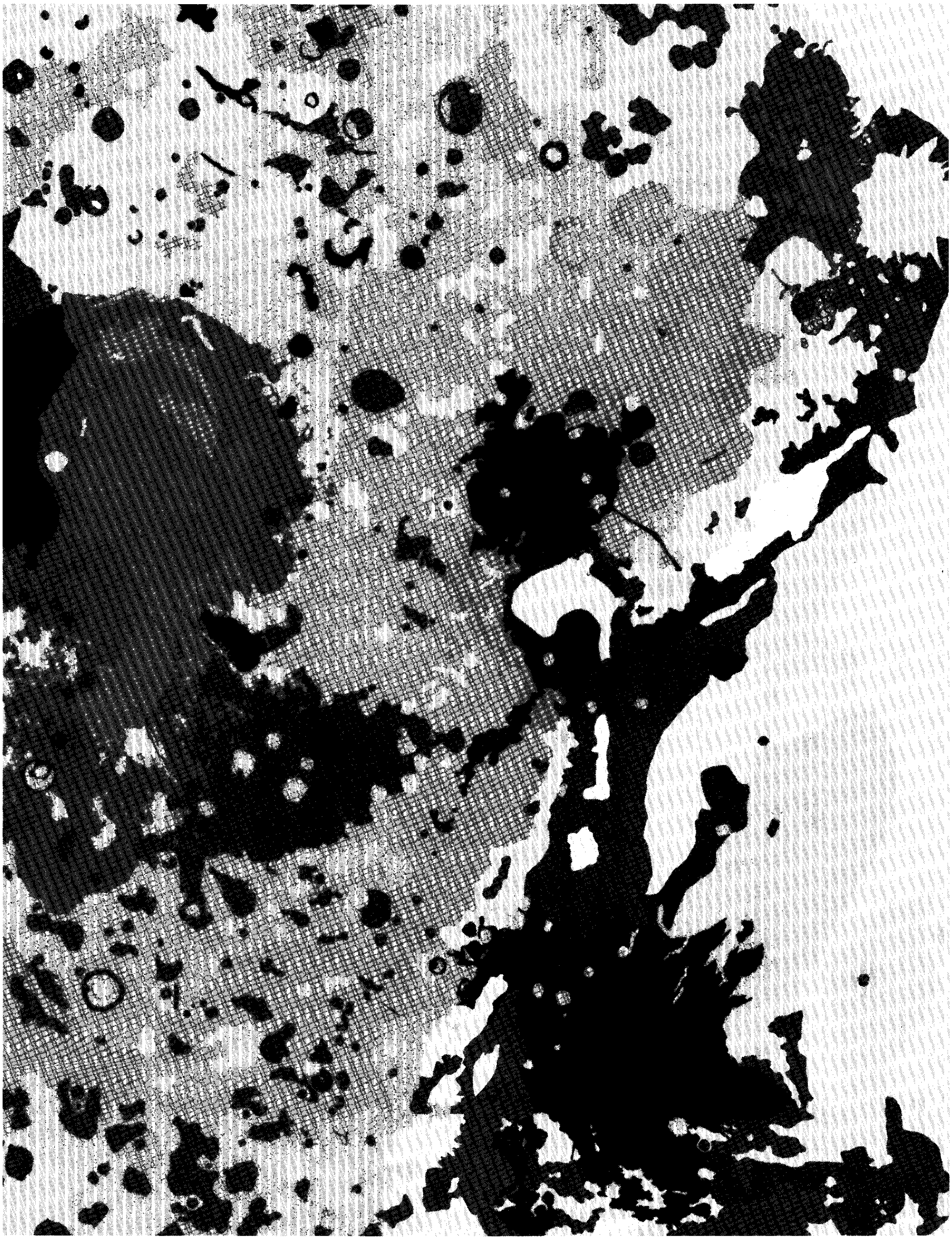


Figure 12. A computerized peelcoat in which the areas of all polygons have been filled with corresponding tag values. Open-window negatives are generated by assigning tag values in the computer.

TAGGING POLYGONS

Binary images derived from scanned geologic maps clearly depict polygons of geologic units when displayed, but in the computer no association exists between the polygons and the geologic units. To form this association, each polygon is tagged with a number that represents a geologic unit. Tagging a polygon means placing a number

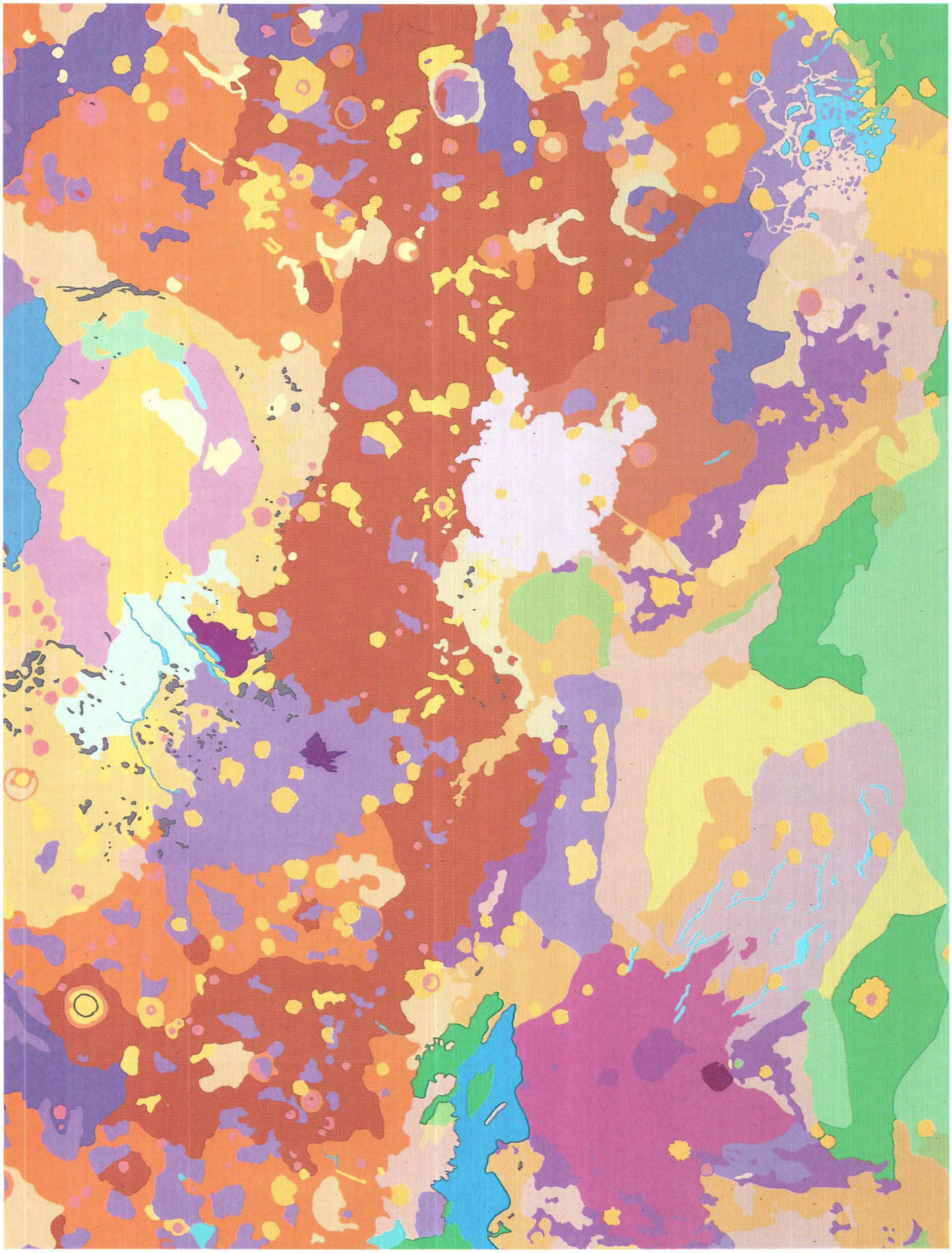


Figure 13. A color geologic map produced by interactively adding color to the computerized peelcoat.

or tag value within the polygon area. The tag value becomes part of the binary image and thus has pixel coordinates (Condit and Chavez, 1979). If the tag value is placed near the polygon boundary or edge so that no other boundary exists between the tag value and the polygon edge, then the tag value coordinates can be used as a starting point to search for the boundary. The edge pixel coordinates can be vectorized, and thus the polygon boundary, tag value, and area can be associated with a geologic unit.

The method used to tag polygons involves a digitizer and a paper copy of the author's original compilation in large format (about 30 by 30 in.). Four control points that can be easily found on the corresponding scanned binary image are initially digitized. Choosing control points that are near each of the four corners will minimize errors in the tagging process. Once the control points have been digitized, each geologic unit in the paper copy is tagged by digitizing the tag value within a geologic unit. The geologic unit is marked as tagged on the paper copy. The tagging process is repeated until all geologic units have been tagged.

The (*x*, *y*) coordinates generated by the digitizer must be converted to pixel coordinates on the corresponding raster image. The pixel coordinates of the four control points are determined by visual inspection. An association is established between the (*x*, *y*) digitizer coordinates of the four control points and their pixel coordinates on the raster image. This information is then used to apply two surface splines (Goshtasby, 1988) to the tag value (*x*, *y*) digitizer coordinates to determine their pixel coordinates in the corresponding raster image. The surface spline mapping places each tag value within a polygon in the binary image that is associated with the appropriate geologic unit on the paper copy of the map.

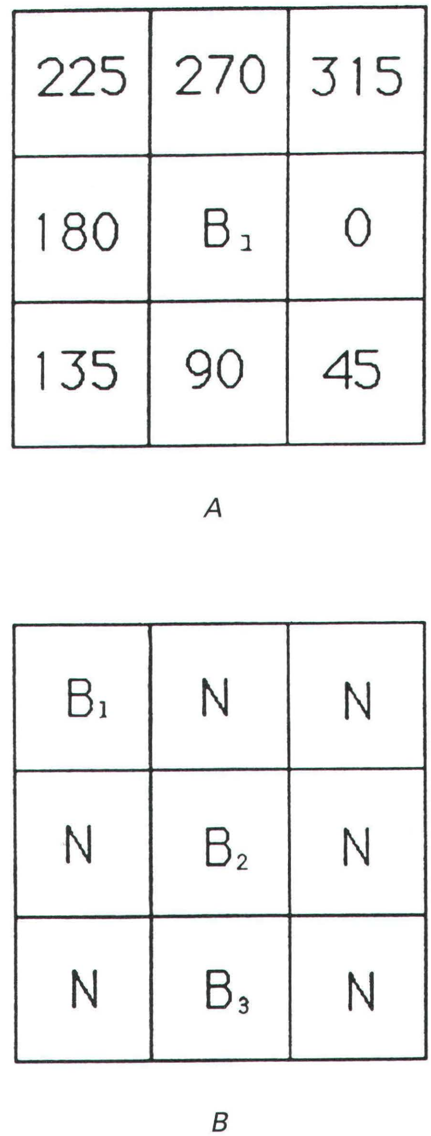


Figure 14. Boundary pixels. *A*, A boundary pixel and directional angles that can be associated with the pixel. *B*, Boundary and nonboundary pixels denoted by B and N, respectively.

VECTORIZING POLYGON BOUNDARY DATA

The vectorization algorithm automates the extraction of the boundaries of polygonal areas from a raster representation of a geologic map. A rasterized geologic map must meet two criteria before the algorithm can be applied successfully. First, the map must be binary; that is, the map consists of 0's for the boundary pixels and 1's for nonboundary pixels. Second, the boundaries should not exceed a width of one pixel. Vectorization of boundary data includes three steps: boundary detection, boundary tracing, and data compression. After polygon boundaries are vectorized, a polygon fill algorithm is applied to the data to produce a gray-tone image map.

The boundary of a polygon is located by using the pixel coordinates (Condit and Chavez, 1979) of a tag value as a reference position. A boundary can be located by examining each pixel's intensity to the left of the tag value and tracing the same line coordinate until a boundary intensity value is encountered. The tag value must be placed within the polygon area so that no other polygon boundary exists between the tag value and its polygon boundary. Once an initial starting boundary pixel (B_1 on fig. 14*A*) is obtained, the pixel coordinates are saved and a search is made for the next boundary pixel. The search for the next pixel is based on the eight neighboring pixels. Each of the neighboring pixels is assigned a directional angle relative to the current pixel, B, which also has an associated directional angle. The associated directional angle of the initial starting boundary pixel is defined as 90°. If there are no breaks in the boundary of the polygon, then one of the neighboring pixels is the next boundary pixel, B_2. The first neighboring pixel examined is at a directional angle 90° less than the current directional angle. If no boundary pixel exists there, the search continues clockwise until a boundary pixel is found or until all directional angles have been searched. For example, in figure 14*B*, if the current boundary value is the center pixel and its associated directional angle is 45°, then the first pixel examined is at a directional angle of 315° (45−90=−45°, or 315°). No boundary pixel exists there and so the search continues clockwise through directional angles 0, 45, and 90° until the next boundary pixel, B_3, is found; B_3 becomes the current boundary pixel and has an associated angle of 90°. The save and search process continues until the starting boundary pixel is reached or until no boundary pixels exist. In the former case, the polygon has been vectorized. In the latter, the polygon boundary is not connected, indicating an error. This search process ensures that bifurcation points generated by the intersection of two polygon boundaries will not cause the algorithm to erroneously vectorize the boundary of the second polygon. That is, the search for a boundary pixel is always toward the center of the polygon that is being processed.

The vectorized boundary pixels are used to fill in the areas of the corresponding polygon that has the tag value. To fill the polygon properly, two special cases that involve changes in the directional angles may occur and must be noted. The boundary pixels associated with the changes are called vertices. One of the changes occurs when the directional angles of previous boundary pixels are 0 to 180°; then the directional angle for the current boundary pixel changes to 180 to 360°. The other change occurs when the directional angles of previous boundary pixels are 180 to 360°; then the directional angle for the current boundary pixel changes to 0 to 180°. After the locations of boundary pixels and vertices have been collected, the data are compressed and stored.

The interconnected nature of the boundary data allows the data to be easily compressed by using chain coding. Data structure is chain coded by designating the starting pixel coordinate of the boundary data as the origin of the chain. All other boundary data are represented in the chain as a sequence of numbers from 0 to 15. The number used to represent a boundary pixel in the chain depends upon the directional angle from the previous pixel in the chain and whether or not the pixel is a vertex. The numbers 0 to 7 are respectively associated with the directional angles 0 to 315° (0=0, 1=45, 2=90, . . . , 7=315). If a pixel defines a vertex, the number associated with its directional angle from the previous pixel is incremented by 8 before it is entered into the chain structure. After all the data for a boundary are coded into a chain, the chain is compressed to 4-bit fields and is stored. A table describing the polygon attributes defined by each chain structure is also maintained. The information stored in the table includes the tag value assigned to the polygon, the line coordinate in the rasterized map of the topmost pixel on the polygon, and the location of the chain within a storage structure. Each vectorized polygon is an entry in the table, which is saved for further processing.

The vectorized polygons and the table are used to generate a raster image that contains polygons whose areas are filled with the corresponding tag value. The polygon-fill algorithm assigns a priority to each polygon and then uses the vector data to fill the polygons according to the priority. Polygons are ranked on the basis of the order in which the topmost pixel of each polygon appears in the raster data. A polygon whose topmost boundary pixel occurs near the top of the raster image has a higher priority than one whose topmost pixel occurs near the bottom. This priority system ensures that polygons internal to other polygons are filled after the containing polygon is filled.

All pixels defining the interior region of a polygon are set to the tag value by using a scan conversion algorithm. In the algorithm, pixels are determined to be inside or outside the polygon by examining the boundary pixels that occur on a raster scan line. Beginning at the extreme left of a scan line and continuing right, a tally, initially set to 0, is kept to determine interior pixels along the scan line. The tally is incremented by 1 for all boundary pixels that are not vertices and incremented by 2 when a vertex is encountered. When the tally is 0, the first boundary pixel that occurs on the line is designated as a starting fill position, and the tally is incremented accordingly. A search is then made for the next boundary pixel that occurs on the line, and the tally is incremented. The tally is then checked. If the tally is a multiple of 2, the location where the boundary pixel occurred on the line is designated as an ending fill position. At this point, all pixels between the starting and ending fill positions are set to the tag value. The tally is then set to 0, and the process continues across the scan line. All scan lines are processed until the polygon has been filled. Polygons are filled until all polygons in the raster image have been filled.

LINE SCREENS AND CONTINUOUS COLOR

Line screens (fig. 6*A*) are used to control the amount of ink that will be printed in a certain area of a printed map. The amount of ink is determined by the diameter of white (clear) dots on a black background spaced a small distance apart in a repetitive pattern. The spacing between dots is so small that the eye cannot resolve the color dot pattern; thus, continuous color is perceived. The same area of a printed map may also contain dots of a different diameter, at the same spacing and in a different repetitive pattern, that control the amount of another color of ink, such as cyan (see any color printed map under magnification). Because the eye cannot resolve either dot pattern of the two inks, the visual effect is the combination of the two inks; for example, yellow and cyan visually generate green. The dots not only vary in diameter, but also repeat at different angles relative to a horizontal baseline. The angle prevents objectionable moire patterns on the printed map. Line screens are readily available for use in the manual method of producing color separations because they can be photographically duplicated from a master. For use in the computerized

Table 2. Percentage of pixels within grid cells

Grid cell size	Pixels in grid cell	Percentage							
		8	13	20	30	40	50	60	70
2	4	0.32	0.52	0.80	1.20	1.60	2.00	2.40	2.80
3	9	.72	1.17	1.80	2.70	3.60	4.50	5.40	6.30
4	16	1.28	2.08	3.20	4.80	6.40	8.00	9.60	11.20
5	25	2.00	3.25	5.00	7.50	10.00	12.50	15.00	17.50
6	36	2.88	4.68	7.20	10.80	14.40	18.00	21.60	25.20
7	49	3.92	6.37	9.80	14.70	19.60	24.50	29.40	34.30
8	64	5.12	8.32	12.80	19.20	25.60	32.00	38.40	44.80
9	81	6.48	10.53	16.20	24.30	32.40	40.50	48.60	56.70
10	100	8.00	13.00	20.00	30.00	40.00	50.00	60.00	70.00
11	121	9.68	15.73	24.20	36.30	48.40	60.50	72.60	84.70
12	144	11.52	18.72	28.80	43.20	57.60	72.00	86.40	100.80
13	169	13.52	21.97	33.80	50.70	67.60	84.50	101.40	118.30
14	196	15.68	25.48	39.20	58.80	78.40	98.00	117.60	137.20
15	225	18.00	29.25	45.00	67.50	90.00	112.50	135.00	157.50
16	256	20.48	33.28	51.20	76.80	102.40	128.00	153.60	179.20
17	289	23.12	37.57	57.80	86.70	115.60	144.50	173.40	202.30
18	324	25.92	42.12	64.80	97.20	129.60	162.00	194.40	226.80
19	361	28.88	46.93	72.20	108.30	144.40	180.50	216.60	252.70
20	400	32.00	52.00	80.00	120.00	160.00	200.00	240.00	280.00

method, however, line screens must be generated in the computer.

Computerized line screens are not necessary if a large-format filmwriter is available to generate the color separations. The open-window negative for a particular color percentage can be used as a mask to instruct the filmwriter whether or not a dot pattern (the dot diameter, or percentage, and the angle of repetition specified with dials on the filmwriter) should be written. In this manner, the color separation can be generated directly on the filmwriter, and valuable disk storage will not be needed to save the computerized color separation.

If a large-format filmwriter is not available, line screens can be generated in the computer by using halftone techniques (Hearn and Baker, 1986) that implement the various percentages by turning on a certain number of pixels within grid cells of a particular dimension. Problems arise in determining which pixels to turn on depending on the percentage and the angle of repetition. Another problem is that the resolution of the filmwriter affects the dimension of the grid cell.

The percentages (8, 13, 20, 30, 40, 50, 60, and 70) of grid cell size used to print USGS maps are not equal to integral numbers of pixels to turn on for numerous grid cell sizes (table 2). The smallest grid cell size that has an integral number of pixels on for the specified percentages is 10. If 10 is used as a grid cell size for a computerized screen and if the computerized screen is written on a filmwriter at 25-micron (0.001 in.) spacing, the dots per inch of the output screen will be 101.6. Therefore, the filmwriter output will have to be reduced photographically by a factor of 0.8467 to obtain the 120 dots per inch required in USGS maps. The ideal spacing on the filmwriter for a grid of 120 dots per inch and a cell size of 10 is 21.167 microns (0.00083 in.). The next percentage, the 13th, which specifies an odd number of pixels to turn on, destroys the symmetry in the pixels that are to be on; thus, moire patterns can result in the printed map.

For the USGS percentages, the next grid cell size that has an integral number of pixels to turn on is 20 (table 2). The even numbers ensure shape symmetry of the pixels turned on for all the percentages (fig. 15); the symmetry eliminates moire patterns. At a grid cell size of 20, computerized screens generated on a filmwriter at 25 microns produce 50.8 dots per inch. The film output will have to be reduced photographically by a factor of 0.423 to produce 120-dots-per-inch line screens. To generate these screens, the ideal micron setting on a filmwriter is 10.583 microns.

The filmwriter generates computerized line screens that have dot patterns oriented horizontally. The angles required for the various colors (for example, 90° for yellow) are introduced to the open-window negatives in the computer by using a program that rotates a digital image prior to screening. The rotated and screened open-window negatives are composited in the computer to generate the

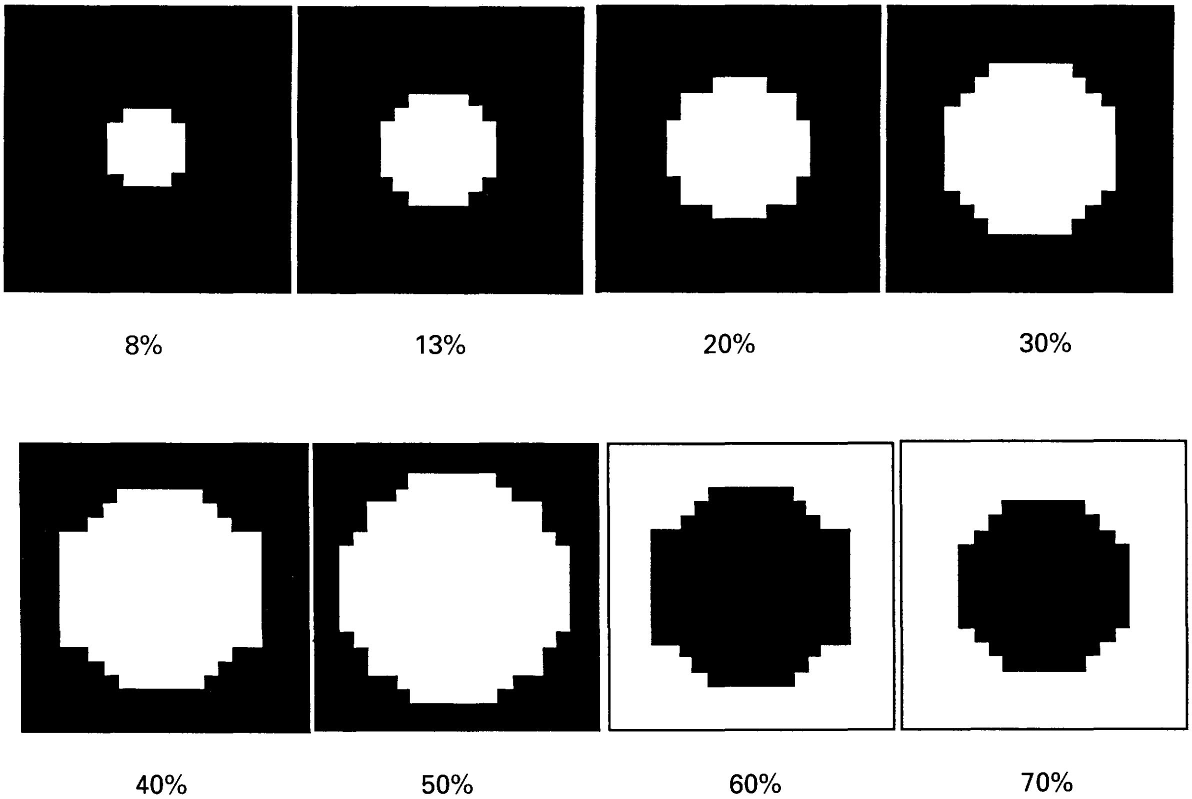


Figure 15. Computerized line screen dots generated by using a grid cell size of 20. The amount of white within a cell represents a percentage, shown below each cell.

required color separations. The computerized color separations are then written to the filmwriter, and the transparencies produced are manually registered. These registered transparencies can be used in a printer to produce a color printed map similar to figure 13.

SUMMARY

Color separations for printing geologic or other thematic maps may be generated by using manual or computerized methods. The manual method is labor intensive and requires several photographic products, such as scribecoats, peelcoats, color guides, screened peelcoats, and color proofs. The computerized procedure requires few photographic products to prepare the map for scanning. The only need for manual interaction in the computerized process is in preparing the author's original compilation, editing the scanned version of the map, and tagging polygons. Because peeling, screening, and compositing are done totally in the computer, material, labor, time, and registration errors are reduced or eliminated. The manual procedure is implemented by skilled draftsmen and photographers in a specialized cartographic unit. The computerized procedure is implemented by trained computer users on a low-cost personal computer in a typical office environment.

REFERENCES CITED

Bryngdahl, Olof, 1974, Moire: Formation and interpretation: Journal of the Optical Society of America, v. 64, no. 10, p. 1287–1294.

Burrough, P.A., 1986, Principles of geographical information systems for land resources assessment: New York, N.Y., Oxford University Press, p. 61–64.

Condit, C.D., and Chavez, P.S., Jr., 1979, Basic concepts of computerized digital image processing for geologists: U.S. Geological Survey Bulletin 1462, p. 3.

Goshtasby, Ardeshir, 1988, Registration of images with geometric distortions: IEEE Transactions of Geoscience and Remote Sensing, v. 26, no. 1, p. 60–64.

Greeley, Ronald, and Guest, J.E., 1987, Geologic map of the eastern equatorial region of Mars: U.S. Geological Survey Miscellaneous Investigations Map I–1802–B, scale 1:15,000,000.

Hearn, Donald, and Baker, Pauline, 1986, Computer graphics: Englewood Cliffs, N.J., Prentice-Hall, Inc., p. 286–289.

Leberl, F.W., and Olson, Dale, 1982, Raster scanning for operational digitizing of graphical data: Photogrammetric Engineering and Remote Sensing, v. 48, no. 4, p. 615–627.

Naccache, N.J., and Shinghal, Rajjan, 1984, SPTA: A proposed algorithm for thinning binary patterns: IEEE Transactions on Systems, Man, and Cybernetics, v. SMC–14, no. 3, p. 409–418.

Wesner, J.W., 1974, Screen patterns used in reproduction of continuous-tone graphics: Applied Optics, v. 13, no. 7, p. 1703–1710.

CHAPTER D

Digital Spatial Data Technology and Applications

By RANDLE W. OLSEN

U.S. GEOLOGICAL SURVEY BULLETIN 1908

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES 1990

CONTENTS

Digital Spatial Data Technology and Applications

By Randle W. Olsen

Abstract

Digital spatial data technology is rapidly becoming a common tool and resource for mapping and geographic analysis. This paper focuses on the collection, processing, and applications of digital spatial data. Data types, sources, and standards are described in the context of technology currently being used by organizations that are building multipurpose data bases. Applications of these data bases are described from the standpoint of users, many of whom are not in the earth science, mapping, or remote sensing professions. These applications pertain to computer-aided mapping and the use of geographic information systems. The challenge to mapping agencies is to provide not only data, but also related user assistance regarding appropriate technology and its limitations.

INTRODUCTION

The principal challenge in digital technology is to develop an integrated scheme that can be applied effectively in a society that has rapidly increasing digital spatial awareness. This scheme must address source data integrity and standards; the development of Federal, State, local, and private sector roles in building, maintaining, and disseminating multipurpose data bases; the definition of map accuracy standards for derivative graphic products; and the demonstration of spatial data applications to those unfamiliar with surveying, mapping, and remote sensing.

DATA TYPES

Digital spatial data are computer-readable, geographically referenced features or feature classes that are described by positions and attributes. Planimetric (horizontal) positions are referenced to coordinate systems, such as geographic latitude and longitude, the Universal Transverse Mercator coordinate system, or a State Plane Coordinate System. Height (vertical) positions are specified as elevations referenced to sea level. Attributes are defined characteristics of a feature and include thematic or feature code, textual label, numerical value, and (or) qualitative descriptor or code.

Digital spatial data can be organized in a data file, a file structure, or a data structure. A data file is a collection of spatial data that covers a defined category or feature class and is physically stored on an exchange medium, such as a magnetic tape or a floppy disk. A data structure describes the logical organization of data elements and the manner in which relations among data elements are defined. A file structure is the physical implementation of a data structure in the computing system environment. There are two basic types of data structure, vector and raster.

Vector Data

Vector data support positions and attributes explicitly related to points, lines, and areas. Point data represent objects or features positioned by single spatial coordinate positions, such as geodetic control points, section corners, and centroids of small cartographic symbols. Line data are ordered sets of points representing linear and curvilinear boundaries and features, such as roads, streams, contours, and utility lines. Area data are ordered sets of points or lines that encompass a closed area, such as county boundaries, land-use classifications, soil types, and land ownership.

The simplest structures of vector data are unstructured. Unstructured point data are lists of coordinates that have a group attribute or individual point attributes. Unstructured line data, often called spaghetti or graphics-formatted data, are best suited to graphics applications. With such data, however, it is difficult to establish relations among features and perform an integrated analysis with other data categories outside of a graphics environment. Other data structures are more complex and include feature networks and topologically structured vector data that employ a combination of points, lines, and areas and a cross-referencing scheme among adjacent data elements. These structures are more difficult to create but offer multipurpose applications, including feature networking, geographic information system analysis, graphics output, and data integration, geographically and between categories for a given area. The U.S. Geological Survey (USGS)

Manuscript approved for publication September 26, 1989.

Digital Line Graph is an example of topologically structured vector data. Vector data are especially suited to point digitizers, line-tracing digitizers, cartographic-quality displays, and network analyses.

Raster Data

Raster data support the classifying of gridded areal units based on a spatial reference system. Simple cartographic raster data structures are binary attributes sequentially arranged cell-by-cell along grid rows and columns. A zero (0) attribute represents a cell containing no features, and a one (1) attribute represents a cell containing a portion of a feature. Rows and columns are referenced to the spatial coordinate system. Digital elevation data are often structured as an array of elevations based on a uniform grid or lattice. More complex raster data structures include attributing schemes describing areal features or color or gray-scale values for pixels (picture elements) of digital imagery. The raster data type is especially suited to scanning digitizers and plotters, statistical analysis, graphics display of integrated data categories, and digital image processing.

DATA SOURCES

Sources of digital spatial data include geodetic surveys, remote sensors, photogrammetry, and maps. Integrated data bases combine these source groups for multidisciplinary applications.

Geodetic Data

Geodetic data are acquired by direct surveys of the Earth's surface and are numerically processed and adjusted. For conventional surveys, raw observed data, such as distances, horizontal angles, or azimuths, and vertical angles are acquired by using instrumentation, such as theodolites, electronic distance-measuring equipment, and levels. Modern surveying equipment, such as total station instruments, combine angular measurements with electronic distance measurements and can record, process, and adjust the raw data in real time.

An increasingly popular alternative to conventional surveys is Global Positioning System technology, which provides coordinates of points, in near real time, based on measurements between a ground-based receiver and numerous orbiting satellites that have known positions in time. Geodetic data bases are files of records that include point type spatial coordinate data, pertinent observed data, and descriptive attributes. Attributes include station name or feature identification and source and accuracy of data.

Remote Sensor Data

Remote sensor digital data are image based and are acquired by nonphotographic sensors. Although sensors can be carried in aircraft, satellite sensors are used most often. Satellites such as the Landsat series and SPOT (Satellite Pour l'Observation de la Terre) scan the Earth's surface from orbital paths and telecommunicate multispectral responses to receiving stations.

Selected facilities process and distribute formatted data files for geographically defined areas, for given dates and (or) orbits, and in one of several multispectral wavelength bands. The primary data elements are pixels arranged on a grid. Each pixel is attributed with a spectral-response or gray-scale value. Remote sensor digital data can be acquired also by scanning aerial photographs. This process results in raw digital image data having pixels arranged on a regular grid and attributed with gray-scale values.

Various levels of geometric rectification are available from distribution outlets. The more sophisticated outlets offer user-defined grid orientation and datum reference. These rectifications require resampling of the original data. Most applications of remote sensor digital data require compositing of registered bands of a given area. Two applications, for example, are automatic classification of land cover and preparation of false-color image graphics.

Photogrammetric Data

Photogrammetric digital data are acquired by digitizing the features on aerial photographs or other remote sensor imagery. Digital terrain data and planimetric data are acquired by using an analog photogrammetric stereoplotter interfaced to a three-axis digital recording device or to an analytical stereoplotter. Terrain data are collected as contours, profiles, elevation grids, and terrain features such as slope breaks, ridge lines, and drainage networks. Data base file structures for terrain data are commonly resampled grids of elevations referenced to a coordinate system. Planimetric compilation includes digitizing roads, hydrography, and other features identifiable on photographs. Data of this type are entered into a data base as vector data files to establish defined feature classes.

Cartographic Data

Cartographic data are acquired by digitizing maps. Three basic methodologies in use are manual digitizing, semiautomatic line tracing, and automatic scanning. Manual digitizing is done by pointing to or tracing cartographic features. An operator uses a cursor on a flat table that has a two-axis digital recording device. Attributes describing the features are entered by using a menu or by keying an alphanumeric code. Sophisticated manual digitizing systems include interactive editing and on-line data processing.

Semiautomatic line tracing involves automatic digitizing by tracking along features. The operator starts and assists the tracker and enters feature attribute codes. Automatic scanning digitizes features and produces a raster data file referenced to a high-resolution grid. Each grid cell is attributed with a binary or color code indicating the presence of a line, symbol, or background color within a grid cell. Scanning is rapid, but postprocessing is often required to convert the raw raster data to the vector data required by many applications. Attribute coding is done off line but may be simplified by prescan preparation and automatic symbol or color recognition.

DATA STANDARDS

Data standards establish digital spatial data specifications, which are important to systematize production processes and data exchange, to establish contract specifications, and to provide a description and level of reliability to the user community. Many aspects of digital spatial data are standardized and can be grouped into three basic components: data exchange, data quality, and feature classification and attribution. A proposed standard, published in the January 1988 issue of the *American Cartographer* (American Congress on Surveying and Mapping, 1988), describes these components. This proposed standard was prepared from two previous standards efforts conducted by the American Congress on Surveying and Mapping (ACSM) National Committee for Digital Cartographic Data Standards and by the Office of Management and Budget Federal Interagency Coordinating Committee on Digital Cartography. The standard is being reviewed and tested and will then be submitted to the National Institute of Standards and Technology for consideration as a formal Federal Information Processing Standard (FIPS).

Data Exchange

Specification of data exchange formats provides physical and logical file structures to producers or users of digital data. The physical file structure includes specification of the exchange medium, such as 9-track magnetic tape, cassettes, floppy disks, or hard disks. For tapes, specification of bits per inch, labeling scheme, file delimiters, and record format is required. The record format includes the type of characters, such as the American Standard Code for Information Interchange (ASCII) standard, and the record length and blocking scheme within a file. Options for specifications are limited by the computer hardware used to read or write digital data files. Exchange and transportability are facilitated by use of industrywide standards. A standard ASCII character set is specified by the American National Standards Institute, as ANSI X3.4–1977. A standard tape labeling specification is given in FIPS Publication 79 (1980), although unlabeled tapes are also often specified.

The logical file structure describes the specification of individual data elements within a file, including a sequential description of fields within records that defines field length and format for alpha characters, integers, real numbers, or real numbers in scientific notation. A logical file structure specification also describes the data type and structure, such as raster or vector data, the coding scheme, the topological relation (if any), and the geographic reference system.

To encourage data exchange and commonality of software, industrywide exchange format standards are being developed. An exchange standard, Information Processing Specification for a Data Descriptive File, has been established by the International Standards Organization as ISO–8211. The standard has recently been adopted by the National Institute of Standards and Technology as FIPS Publication 123 (1986). ACSM (1988) proposed a Spatial Data Transfer Specification, which is especially suited to raster or vector data. The transfer specification employs the FIPS Publication 123 as the implementation method.

Data Quality

Data quality can be described in terms of five characteristics: source and processing history, positional accuracy, attribute accuracy, logical consistency, and completeness. These characteristics have been proposed as sections of a standardized truth-in-labeling quality report by ACSM (1988).

Specification of source and processing history is a textual record describing the method of data acquisition and subsequent modifications or revisions, including data source, processing transformations, and reference systems. Specification of positional accuracy of coordinate data can be in terms of relative accuracy with respect to source materials and absolute accuracy with respect to the reference system. The accuracy can be determined by deductive estimating or by testing. Specification of attribute accuracy is more subjective than that of positional accuracy and can be described as a tolerance for misclassification. Again, the accuracy can be determined by either estimating or testing. Specification of logical consistency includes descriptions of topological validity internal to a file and compatibility with other files for edge joining or data integration of multiple categories. Specification of completeness includes a description of the feature or object selection criteria compared to the universe of such features or objects.

Data quality standards are set by producers or users of data who have specific applications requirements. There are so many diverse requirements and related economic considerations that industrywide standardization of data quality is not considered viable. However, standardizing data quality for a particular data base is employed by many producers and users of digital spatial data. Standardizing the contents of a quality report is also viable. Actual documentation of data standards can be in the form of published user's guides or included as a record within a digital file.

Feature Classification and Attribution

Feature classification and attribution have two characteristics: definition of features and feature classes common to a digital file and attribution of features or topological elements. ACSM (1988) proposed a list of cartographic features and definitions and a scheme for producing attribute values. This standard does not specify either a hierarchy of feature classes or specific coding and labeling values. As with data quality, there are so many established systems and diverse requirements and applications that standardizing industrywide is not viable. However, standardized classification and coding values or labels for a particular data base are employed by providers of digital spatial data. In many cases, there is sufficient coordination among Federal agencies to promote some of these classification or coding schemes as FIPS.

DATA BASE ISSUES

Archival Storage

One of the principal components of a data base is the archival storage of data considered to be the master copy. Small data bases can be archived as data records within the central processing unit of a computer; no storage is required beyond the use of computer memory. Large data archives currently require off-line storage of data files resident on computer-readable magnetic tapes, cassettes, or disks. In this case, access to data requires locating the appropriate storage device and providing computer access through the loading of tape or disk input units. In some cases, external storage devices are accessible directly by the computer through input channels. Examples include on-line disk units and laser-recorded, mass-storage optical disk units.

Factors to be considered when building a data base include sizing the ultimate archive, frequency of access, and resource availability for purchasing and staffing a data base computer.

Management Systems

Management systems for data bases are software routines on the host computer that provide for queries and for data file or record location in the archive. A data file management system is the simplest type of query and retrieval mechanism; queries such as location and data type are cross-referenced to an actual tape number. More sophisticated relational data base management systems operate with on-line access to data files or records, have full query and retrieval mechanisms, and perform limited data integration with other data sets, reformatting, and other post-retrieval data processing.

Maintenance and Revision

Maintenance of an archive of data requires procedures that ensure access and integrity of the master file. For large multiuser data bases, a data base manager is responsible for policing which files are changed or replaced and for providing file security by limiting user access and by periodically duplicating the archive. Well-defined rules are required to control changes to archival data. For off-line files such as magnetic tapes, the data base manager must change the retrieval and management system to modify the pointer for a new tape number. For on-line files, clear rules are required specifying who has authority to modify records or create new files. Authority to perform such modifications should be restricted to a data base manager, although users may submit new or revised data files to the manager.

Security procedures are required to protect a data base from accidental physical damage and human error and limitations of the storage media. Examples of physical damage include heat, moisture, unsanctioned access, and sabotage, which can destroy central archives. Accidental damage can occur through human error in maintaining a data base. Periodically backing up the data base protects the information. Backup copies should be stored off site, in a physically secure location. Maintenance procedures are required also to overcome limitations of the storage media. Magnetic tapes, for example, require periodic exercising because of limited shelf life. Tapes have a 3- to 5-year shelf life, which varies with tape quality and environmental factors. Exercising of tapes is done either through routine access or by off-line re-reeling of the tape itself. The shelf life of disk units and laser-recorded optical disks is longer.

Distribution

Data base distribution is the transfer of data files or records from the archive to a user. The data base management system provides query and retrieval capability, often sophisticated enough to direct broad queries into specific files or records. Security procedures may limit access, especially to personal or sensitive data files.

For small computers that have resident data bases, applications software can directly access and use data without employing a distribution function. For off-line, but resident, data archives, data files must be loaded onto a host computer for a user. For large data bases, data files are copied and mailed to a user or, more recently, copied and telecommunicated through a distributed network to a user's computer. Mass storage technology and telecommunication links will increasingly provide ready access to geographically distant data bases.

DIGITAL DATA PRODUCTS

Types of Data Products

Digital data products are computer-readable data sets generated from an established data base. They can be classified into three broad groups, according to the level of postprocessing. The first group of products replicates the archived data sets in terms of content, area of coverage, and format. The preceding descriptions of data types, sources, and standards can also be applied as characteristics of this group of products. Many data retrieval facilities offer some flexibility in tape copy specifications regarding tape labeling, data density, character sets, and blocking of records.

The second group, derivative products, requires some level of postprocessing, which uses established computer programs operating in a batch mode. Derivative products are created by changing characteristics of the archived data without adding new data. Examples include paneling and partitioning files to change an area of coverage, eliminating undesired data content, integrating data categories for a given area into a single file, adding or removing topological pointers, converting from vector to raster data types (or vice versa), and reformatting data records.

The third group, custom products, involves an interactive manipulation or addition of data by using human intelligence. While many of the operations are similar to those described in derivative products, custom products are created for a single application. Geographic information systems are often a convenient tool for developing custom products, and interactive editing workstations may also suffice.

Documentation

Documentation of digital data products is extremely important to a data user, especially if the user is not familiar with data collection and processing or if the user is from a different organization. The most common form of documentation is a data user's guide, which describes the record format, coding and geographic reference schemes, and content. Additional details such as source, accuracy, lineage of digitizing and processing history, and sample data sets are also desirable. Data user's guides should be brief enough, however, for economical routine distribution to inquirers and to purchasers of data.

Another form of documentation is the use of textual records, within the file itself, as headers to each record type. Documentation of derivative or custom products is more difficult. Many organizations include information on the software used to perform postprocessing or brief descriptions of derivative options as part of a data user's guide. The purchasers of derivative or custom products specify their own requirements and are already familiar with characteristics of the data.

User Services

Many data base distribution agencies and businesses have extensive marketing and user service functions. These functions include processing orders and inquiries, facilitating data base queries, distributing products, and providing technical assistance to users. Extremely technical or detailed assistance is referred to a technical specialist of the organization. User service organizations also frequently have a means to receive feedback and comments on the utility of available data and requirements that may not be met, either because of lack of data or deficiencies in the data or specifications. It is important that this type of feedback be relayed to the people responsible for setting standards and specifications as well as programmatic priorities for future data collection and data base activities.

Many agencies and businesses maintain data bases for internal use and do not actively market their data or devote resources to extensive documentation and user services. The difficulty in acquiring spatial data and locating potential sources often results either in applications going unmet or in expensive duplication of the digitizing effort.

COMPUTER-AIDED MAPPING

Computer-aided mapping is a process that converts digital spatial data into a graphic. The graphic is a map or chart and can be soft copy or hard copy. The soft copy is a graphic display of digital data on a screen, such as a cathode-ray-tube terminal or a video display. The uses of soft-copy graphics are short term and include verification, interactive editing, design of a hard-copy plot, and spatial data integration and analysis such as geographic information system applications. The hard copy is a graphic plotted onto permanent media such as film, paper, or mylar. The uses of hard-copy maps and charts are similar to those of the soft-copy graphics described above and include the production of maps that conform to established standards of accuracy, content, and cartographic design.

Graphics Workstations

Graphics workstations are used to prepare digital data for a particular graphics display device or hard-copy plotter.

The workstation is an integrated system comprising a display screen, a computer, resident software, an input device to receive data, and an output device to create a plot file. Workstations are required to perform many data-processing steps, such as the following:

- Vector-to-raster conversions for data bases, in vector format, requiring display or plotting on a raster device.
- Raster-to-vector conversions for raster data sets requiring plotting on a vector plotter. This conversion requires interactive editing for cartographic quality linework.
- Unstructuring vector data for data bases that are topologically structured into nodes, lines, and areas that have complex pointers.
- Feature filtering for selecting features that are based on attributes and inclusion rules.
- Coordinate transformations for converting the data base reference system to that of the desired map product.
- Generalization for aggregating or simplifying complex features that were digitized at a larger scale than the desired graphic.
- Horizontal integration for paneling or partitioning of one or more files into a desired geographical area.
- Vertical integration for compositing multiple files to combine features from separate category files into a single file.
- Symbolization for using rule bases to convert attributes to graphics symbols or line weights for selected classes of features.
- Displacement for displacing lines that are too close for distinct cartographic display. Displacement can be performed by reformatting digital records into the file structure and format required by the plotter or display device.

Vector Plotters

Vector plotters operate in a linear mode and use a pen, pencil, aperture, cathode-ray-tube beam, or scribing stylus to trace a string of sequential positioning commands that are based on a local x, y coordinate system. Vector plotters require digital data to be in a graphics-based vector format consisting of points or line strings described by coordinate and symbol type. Sophisticated vector plotters are equipped with symbol or text apertures for printing on photosensitive film. Otherwise, symbols and text must be mapped out by coordinates or manually added after plotting. Vector plotters allow linear features to be accurately plotted with consistent line weight and cartographic quality. The time needed to produce a vector map is directly related to the complexity of lines and symbols. High-precision plotters (positional accuracy within 0.1 mm, or 0.004 in.) must operate slower than high-speed plotters that have lower accuracy. Plot times vary from several minutes, for verification plots of simple maps, to many hours, for complex cartographic-quality maps.

Raster Plotters

Raster plotters print in a scanning mode, which requires that the digital data be in a raster format before plotting. Therefore, all map detail, including symbolized points and lines and text, are represented by a matrix of pixels. Commonly, a blocky or stairstep appearance is evident on the output of most raster plotters that have a raster size of 20 lines per millimeter (about 500 lines per inch) or less. Raster plotters are fast and produce plots in as few as 30 seconds for electrostatic verification drum plotters and in less than 1 hour for high-resolution, scale-stable-film recorders. Raster plotters are especially useful for complex graphics because the plotting time is dependent on the pixel resolution and not on the number of lines or symbols on the map. High-resolution plotters that have a raster size of 40 lines per millimeter (about 1,000 lines per inch) or less are available, and probably most future high-volume and high-accuracy mapping will be accomplished by using scanning technology. These resolutions allow mapping of narrow line weights and complex symbols that meet cartographic standards and esthetics. Raster plotters are also highly suited to film recording of image maps that have an array of gray scales or colors.

CARTOGRAPHIC STANDARDS

Cartographic standards establish the quality criteria of map products. These requirements are important to systematize production processes, to establish contract specifications, and to provide a description and level of reliability of map products to the user community. The many aspects of standardized mapping can be grouped into three basic components: map content, map design, and geometric accuracy. Because mapmaking companies and agencies establish their own standards, their maps are unique, especially in content and design. Computer-assisted mapping offers new possibilities for interactive design and custom maps. The development of standards will preserve cartographic integrity without inhibiting creativity in content selection and design.

Map Content

Map content criteria include specifications for features to be mapped, the projection and graticule to be used, and a statement of source and processing history. For cartographic base maps, standard features include transportation networks, hydrography, boundaries, contours and elevations, cultural features, survey monuments, and vegetation cover. Agency-specific thematic maps may include some of the above features and a specific theme overlay, such as soils, land ownership, geology, and demographic zones. Image maps are rectified photoimagery, sometimes enhanced with cartographic or thematic features. Large-scale maps often include engineering design criteria, utilities, and cadastral parcels. In automated cartography, selected map features are chosen by rule bases derived from the specifications and applied to feature identifiers or attribute codes in a digital data base.

Specification of projection and graticule provides for the spatial coordinate reference system for mapped features. Scale and shape anomalies are minimized by selecting a map projection such as the Universal Transverse Mercator, the Lambert Conformal Conic, the Transverse Mercator, or the Albers Equal Area. The graticule is based on the selected projection and includes either lines of latitude and longitude or Universal Transverse Mercator or State Plane Coordinate System grids. For convenience, an arbitrary quadrangle area, such as a county boundary, can be selected to limit the map, and internal grids or crosses show the established reference system.

Source and processing history are contained in a brief credit legend that has information on source data, date of aerial photography, mapping date, and reference system. Often a brief note is included to describe the production process used to create the map. This information is especially important to computer-produced maps because they are often created with little regard to map accuracy standards and established cartographic practice.

Map Design

Map design specifications include criteria for depicting specific features, text, colors or shading, and feature generalization that are appropriate for a particular map scale. Feature symbols are specified by size, line weight, and the displacement criteria to be used when several features are coincident or immediately adjacent to each other. Symbols often have been complex, including double-line portrayal and extensive use of dots, dashes, and intricate patterns. Automated cartography often uses simpler symbols, such as single lines, which aid the digitizing and plotting process but often compromise the diversity of symbols available. Text is specified as a font and type size related to features or sets of features. Colors and shading can vary considerably, depending on the system and technology used to create the map. Screens are often employed to minimize the number of distinct printing inks required and to create several shades of a particular color or combinations of colors. Screen specifications include density and angle of rotation of the screen.

Geometric Accuracy

Standards for geometric accuracy include the graticule and feature content within a map. Because of the importance of referencing a map to an established coordinate system, the graticule accuracy should be consistent with the accuracy and precision of the technology used to plot a map. The accuracy threshold is about 0.1 mm (0.004 in.), although some standards allow up to two times this threshold. The National Map Accuracy Standards, established by the Office of Management and Budget for use by Federal agencies, is the most commonly used U.S. standard for location of mapped features. This standard requires 90 percent of well-defined features portrayed on the map to be within 1/50 in. (1/30 in. for scales larger than 1:20,000) and contours to be within one-half of the contour interval of their true position. The American Society for Photogrammetry and Remote Sensing (ASPRS) has proposed a modified standard for large-scale maps. The standard is based on a root-mean-square-error threshold rather than a 90 percent criterion and has a consistent accuracy criterion across all scales. This standard also employs ground unit thresholds for a range of metric and U.S. customary unit scales of maps. For thematic overlays, standards are often relaxed due to the ambiguous definition of many thematic boundaries. Geometric standards are especially important in automated cartography because of the dependence on the accuracy of the source data base and the flexibility to plot such data at any desirable scale. Maps are often enlarged to scales too large to meet accuracy standards. Labeling is vitally important on such maps to alert a user to limits of the accuracy of spatial coordinates or areas extracted from these maps.

GEOGRAPHIC INFORMATION SYSTEMS

A geographic information system (GIS) is a computerized hardware and software system designed to collect, manage, manipulate, analyze, and display spatially referenced digital data. In the simplest sense, a GIS automates the process of gathering and analyzing spatial data as needed by the user. For example, the layering of categories depicting certain types of information, such as vegetation, soils, and roads, has been used for decades to identify areas that have a specific set of characteristics. A GIS can perform this task quickly and automatically calculate statistics about the areas. The capabilities to combine many more categories than is feasible through a manual process and to relate the categories mathematically are advantages of GIS's.

GIS Hardware

A GIS is an integrated collection of hardware and software components. Although many vendors market turn-key systems consisting of matched hardware and software components, many GIS users choose to build systems that have individual hardware and software components and the requisite interfaces. Hardware components include a host computer, an interactive workstation or terminal, and a graphics display. Input and output channels are linked, directly or indirectly, to digitizing tables, existing data bases or files, and graphics plotting devices. Host computers can range in size from personal to mainframe, and 32-bit microprocessors are increasingly common.

GIS Software

Software for GIS's includes operating system software and application software. The system software is configured as part of the host computer and provides for component interfaces, utilities, compilers, and an internal data management system. The application software provides the direct spatial data management and processing required by the user. GIS's are extremely diverse in data base structure and functionality and have many methods for storing and manipulating either vector or raster data sets and related attribute, coordinate, and topological information. A model for a particular GIS can be constructed by specifying the functional components that it must possess to meet requirements. Although there are many ways to group functional components of applications software, the Federal Interagency Coordinating Committee on Digital Cartography (Guptill, 1988) has identified the following five classes:

1. The user interface functions simplify and organize the interaction between the user and the applications software.
2. The data management functions access and archive data and include storage, retrieval, updating, and security protection.
3. The data base creation functions include input of data in a format that can be acted upon (by digitizing from a map or photograph or loading an existing data file), attribute tagging, and format conversion.
4. The data manipulation and analysis functions include redefinition or selective extraction of a subset of data, restructuring (such as vector-to-raster conversion), geometric coordinate transformations, and the capability to analyze resulting data. Examples of analysis functions include compositing disparate data sets, impedance, flow and shed computations, and derivation of descriptive statistics.
5. The display functions generate and annotate graphic images and include the mapping of features by using color and symbology.

GIS Applications

Most GIS applications involve mapping, modeling, or monitoring. Mapping applications include the graphic display of results of data merging or manipulation. Within the content and resolution limits of the data, a map can be generated at any scale and contain selected feature groups, customized colors, labeling, and symbols. When elevation data are available, three-dimensional perspective views can be generated. Particular areas of interest can be highlighted, eliminating map clutter. Modeling applications include the analysis of spatial relations of disparate data categories. The output may be statistical, a graphic plot, or an answer to a what-if question. A simple example of a constraint modeling application is a project to identify potential locations for a particular development, adhering to rules on where such development sites can be placed, such as on slopes less than 5 percent, more than 1,000 feet from a road, and on sandy soil outside of a flood plain. The analysis combines four data sets: roads, flood plains, elevation, and soils. Solids modeling includes the analysis and portrayal of subsurface features. Monitoring applications analyze or display the change over time of time series data.

GIS User Community

Increasingly, GIS applications are being developed for low-cost personal computers. Initially, GIS users included government resource agencies, companies involved with natural resources, and businesses involved in management strategies relating to the locations of assets and facilities. More recently, users in the fields of economics, demography, emergency management for natural disasters, and environmental planning are exploiting GIS technology.

So many Federal agencies are recognizing the benefits of GIS's as management tools that many budget initiatives specifically refer to a GIS. GIS's represent a technology appropriate for any agency that has used maps to solve problems. When digital data sets become available on large-capacity, mass-distribution media, such as read-only memory disks, GIS's will become available for personal and home use.

GIS Cost-Benefit Issues

A GIS can perform agency or company mission activities more effectively, efficiently, and accurately than traditional manual map-analysis approaches. A cost-benefit analysis may be appropriate for those seeking to incorporate a GIS into their programs. Costs include hardware and software purchase, personnel, digitizing, training, maintenance and support, management, and product distribution. Benefits, which are more difficult to determine, include cost avoidance due to error avoidance, information efficiency, analysis capability, and ability for exact repetition.

Continuing development of GIS's is necessary to provide the more sophisticated analytical capabilities needed for scientific analysis and resource assessment. A principal limitation to GIS use is the availability and quality of data in digital form. Digitizing can be one of the most costly and time-consuming elements in the use of a GIS. The development of the large, structured digital data bases necessary for the effective use of GIS's, particularly in the Federal Government, is an expensive and long-term process. However, this cost can be mitigated by establishing interfaces between existing digital data bases and by using industrywide standards for new digital data capture. Interfaces and standards will increase multipurpose applications and reduce duplication of effort.

Land Information Systems

A land information system (LIS) is a type of GIS that relates to land-focused data categories. The hardware and software tools used are similar to those described for a GIS. LIS's can be applied to local or small regional areas such as cities and counties and require data sets of higher resolution and accuracy than most GIS data sets. Natural LIS data categories include geology, soils, vegetation, and hydrography. Cultural LIS categories are those that divide the Earth into manmade features, political jurisdictions, or land parcels relating to ownership and use. LIS development on local levels has often had a single application; data standards, formats, and geodetic reference systems are unique to a local agency and are not easily integrated with other data categories covering the same area or adjoining regional areas.

Multipurpose Cadastre

The multipurpose cadastre is a geodetically based framework that supports graphic and digital land-related information at the cadastral parcel level. The cadastral parcel is a proprietary land unit that has defined boundaries and an index to ownership, title, tax, and other local administrative records. There are four basic components of a multipurpose cadastre: (1) the geodetic framework, densified high-order control surveys, usually referenced to the local State Plane Coordinate System, (2) current and accurate large-scale base maps and (or) digital base category data referenced to the geodetic framework, (3) the cadastral parcel overlay in graphic or digital form that delineates the cadastral boundaries with unique identifying numbers, also referenced to the geodetic framework, and (4) a series of administrative and property-based records keyed to the parcel identifier numbers. A multipurpose cadastre data base allows other thematic or natural data categories to be referenced to the same geodetic framework, thus permitting multiple applications at the local level. The cadastre can also be compatible with other data bases and GIS's for broader regional, State, or National applications.

In its publication, Need for a Multipurpose Cadastre, the National Research Council (NRC, 1980) Committee on Geodesy, Council Panel on a Multipurpose Cadastre, identified "a critical need for a better LIS in the United States to improve land-conveyance procedures, furnish a basis for equitable taxation, and provide much-needed information for resource management and environmental planning."

The ACSM–ASPRS Joint Cadastre Task Force published recommendations for implementing a multipurpose cadastre, including areas of research, information exchange, education, legislation, and development of standards and specifications (ASPRS, 1985). Many of these recommendations are now being followed by the Geographic Information Management Systems Committee, jointly supported by both societies. At the Federal level, the Federal Geodetic Control Committee, chaired by members of the National Oceanic and Atmospheric Administration and comprising representatives from at least 11 different agencies, has established a Multipurpose Cadastre Working Group to coordinate demonstration pilot projects and to develop guidelines for a multipurpose cadastre. The ACSM–ASPRS Government Affairs Committee has urged the passage of Federal legislation to create a multipurpose cadastre coordinated among appropriate Federal agencies.

CONCLUSION

Future Trends

Trends in the practical application of digital technology are driven by the continued advances in computing power and availability, accompanied by significant decreases in cost. These advances have exceeded a tenfold order of magnitude each decade since the mid-1950's. The awareness in the user community of the importance of spatial data is increasing rapidly, as exemplified by the proliferation of computer-aided drafting and mapping systems and by the advanced graphics capability on personal computers.

While requirements for digital spatial data far exceed the available resources for providing the data, there are some exciting development trends that can bridge the data availability gap. For geodetic data acquisition, the Global Positioning System will become increasingly portable and have instantaneous output of positions. This economical and accurate georeferencing capability will provide a foundation for all other data categories and is especially critical for large-scale, land-based applications. Digital sensors in aircraft will complement current satellite data collection methods to increase instantaneous digital imagery collection with multispectral sensors to aid in feature classification. For digitizing from existing source graphics, automated symbol and color recognition and feature tracking will greatly enhance raster digitizing technology and conversion to vector-data form.

The current development of exchange standards and transformation software will allow easier integration of disparate data sets. New telecommunications technology will allow rapid transfer of data bases to distributed processing sites. Workstations that can superimpose imagery on existing data are becoming more cost effective and provide a means to update and maintain a data base in a timely manner. As data bases become more available, the current data integration and analysis capability must increase further in sophistication.

The Challenges

The challenge to mapping agencies is to help guide the obvious future trends. Mapping professionals will continue to be responsible for providing cadastral and geodetic, photogrammetric, remote sensor, and digitized map data. They will also continue to be responsible for providing products and services to the user community, and digital data bases will be an increasing source of such products and services.

The principal challenges are to foster development of an integrated scheme that addresses source data integrity and standards; the development of Federal, State, local, and private sector roles in building, maintaining, and disseminating multipurpose data bases; the definition of map accuracy standards for derivative graphic products; and the demonstration of spatial data applications to new segments of society that have not employed surveying, mapping, and remote sensing.

The not so obvious and possibly more important challenge is to educate the user community on the appropriate technology, data, and limitations of given applications. Not all users of spatial data are trained in the geodetic and mapping sciences, and these users outnumber the traditional users of engineering surveys, aerial photographs, remote sensor imagery, topographic maps, and other graphic forms of spatial data. Successful implementation of digital technology is vital.

REFERENCES CITED

American Congress on Surveying and Mapping, 1988, The proposed standard for digital cartographic data: The American Cartographer, January 1988, v. 15, no. 1, 144 p.

American Society for Photogrammetry and Remote Sensing, 1985, Implementing a National multipurpose cadastre: Photogrammetric Engineering and Remote Sensing, August 1985, v. 51, no. 8, p. 1175–1180.

FIPSPUB79, 1980, Magnetic tape labels and file structure for information interchange: October 17, 1980, Federal Information Processing Standard, 64 p.

FIPSPUB123, 1986, Specification for a data descriptive file for information interchange: September 19, 1986, Federal Information Processing Standard, 32 p.

Guptill, S.C., 1988, A process for evaluating geographic information systems: U.S. Geological Survey Open-File Report 88–105, 55 p.

National Research Council Panel on a Multipurpose Cadastre, 1980, Need for a multipurpose cadastre: Washington, D.C., National Academy Press, 112 p.

CHAPTER E

Radon Potential Defined by Exploratory Data Analysis and Geographic Information Systems

By KATHERINE FITZPATRICK-LINS, THOMAS L. JOHNSON, and JAMES K. OTTON

U.S. GEOLOGICAL SURVEY BULLETIN 1908

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES 1990

CONTENTS

FIGURES

TABLES

Radon Potential Defined by Exploratory Data Analysis and Geographic Information Systems

By Katherine Fitzpatrick-Lins, Thomas L. Johnson, *and* James K. Otton

Abstract

Through the use of exploratory data analysis and a geographic information system, indoor radon values located by zip code were analyzed statistically and related spatially to geology and soil data for Fairfax County, Virginia. A geographic information system overlay of geology on zip code areas that have high upper quartile levels (25 percent or more of the samples exceed the U.S. Environmental Protection Agency guideline of 4 picocuries of radon per liter) revealed that the majority of such areas are on Paleozoic metamorphic rock. In areas where the upper 25 percent of the radon values from Paleozoic metamorphic rock are lower than 4 picocuries per liter, soil seems to provide a mitigating effect. Four levels of radon potential, low, moderate, moderate to high, and high, are defined for the study area.

INTRODUCTION

Radon, widely recognized as a major health problem (National Research Council, 1988), has the highest risk coefficients of any naturally occurring contaminant. The U.S. Environmental Protection Agency (EPA) and the U.S. Public Health Service Centers for Disease Control, estimate that as many as 20,000 people die of radon-induced lung cancer each year (U.S. Environmental Protection Agency and U.S. Public Health Service Centers for Disease Control, 1986). As concern increases about how the geology and soil in an area affect radon levels in homes, and thus public health, meaningful ways of displaying, comparing, and analyzing multiple geologic and indoor radon data sets have become more important. Analysis of such data and derivation of radon potential maps based on this data provide important tools for public officials who must estimate the extent and the degree of the problem, plan surveys of homes to directly measure indoor levels, and study means of mitigating the problem in their areas of jurisdiction.

One of the most important features of geographic information systems (GIS's) is the ability to search for and display relations among mapped attributes. In this study, a GIS and an exploratory data analysis (EDA) were used to examine the relations among indoor radon data, geologic mapping, and soil mapping. Radon is a colorless, radioactive, inert, gaseous element formed by the disintegration of radium. The association of geology, soil, and indoor radon has been documented in several studies. Tanner (1964, 1980) studied the mechanisms by which radon is generated and migrates in rock and soil and established the framework for understanding how radon filters into structures through their foundations. Sachs and others (1982) first noted a correlation between elevated indoor radon values and certain rock types in Pennsylvania. Akerblom and others (1984), Nero and Nazaroff (1984), and Eaton and Scott (1984) all concluded that soil and underlying rock are the primary source for indoor radon. Otton and others (1988) and Gundersen and others (1988) generated radon potential maps for two counties in suburban Washington, D.C., based on the geologic attributes of the rock and soil and site-specific indoor radon data.

Hundreds of thousands of indoor radon measurements have been made nationwide over the past 3 to 4 years by private companies. Because of the importance of preserving confidentiality of the homeowner, location information for these data are usually listed by postal zip code only. These measurements represent a valuable source of information on the variation in radon levels across substantial areas of the country. With the advent of digital soil and geologic mapping, these large zip-code-based indoor radon data sets can be studied in a GIS environment. Application of EDA in conjunction with GIS's allows for testing the hypothesis that radon values are associated with the underlying geology.

Fairfax County, Va., was selected for study because indoor radon problems were recognized within the county (Fairfax County, 1987) and a large indoor radon data set and digitized geologic and soil maps were available. In addition, previous investigations of the geologic origins of the radon problems (Fairfax County, 1987; Otton and others, 1988; Douglas Mose, George Mason University,

Manuscript approved for publication April 21, 1989.

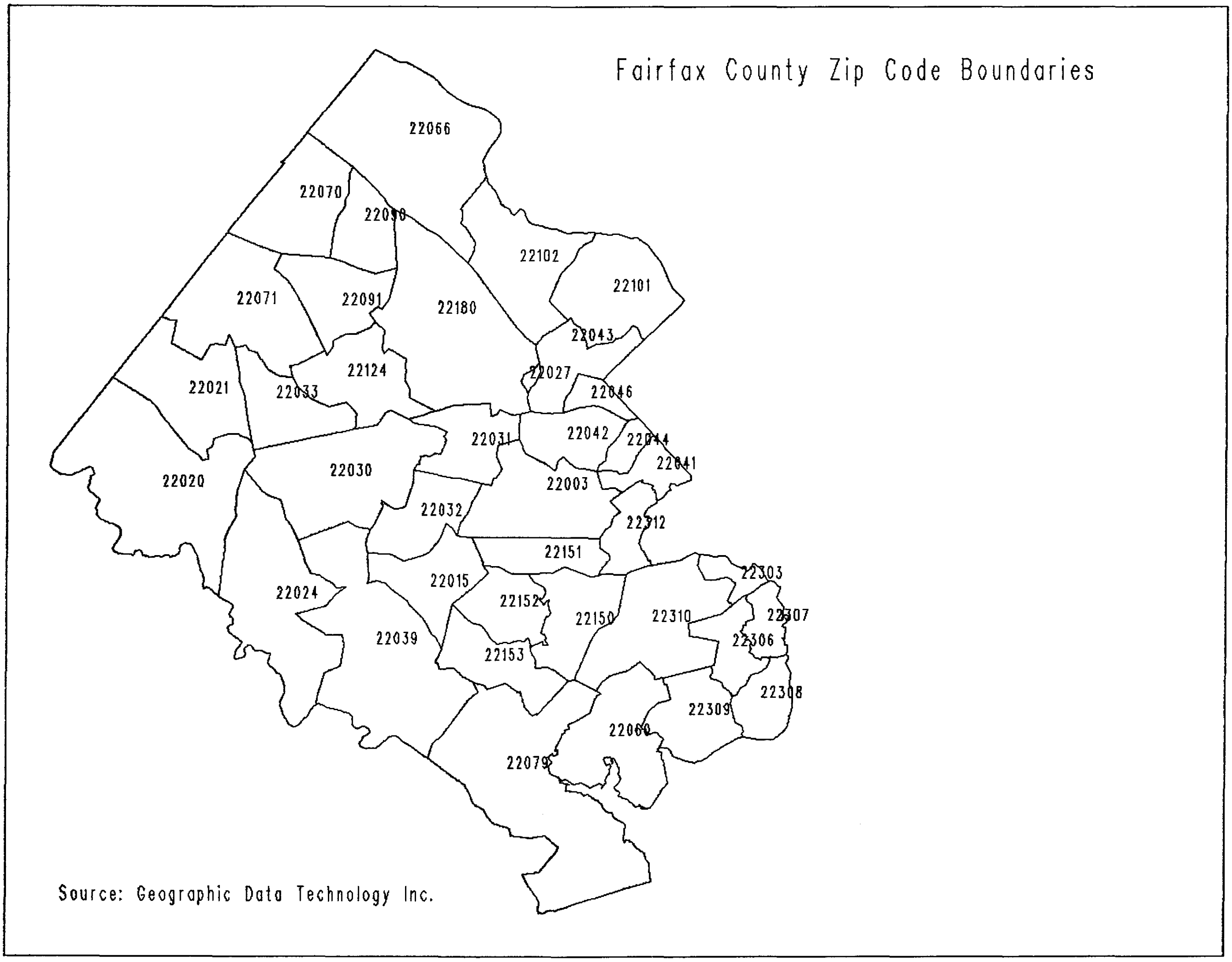


Figure 1. Zip code boundaries for Fairfax County, Va.

written commun., 1987, 1988) provided a check of the reasonableness of our results.

Portions of Fairfax County are underlain by rock types similar to those of the Reading Prong in Pennsylvania and are believed to be associated with high levels of radon. The geologic formations most likely to produce radon are those in the western portion of the county, west of the Fall Line (Otton and others, 1988). The Fall Line is the contact between the metamorphic and intrusive rock of the Piedmont and the unconsolidated sediment of the Coastal Plain. Higher radon readings are expected to be found in homes in those areas of the county underlain by metamorphic and intrusive rock.

BASIC STATISTICS

The data used in this study were obtained from the EPA and represent vendor data for Fairfax County collected from 1985 to the end of 1986. The data are the results of charcoal canister tests taken in homes during the winter. The data for the county comprise 1,408 measurements across 36 zip codes from participants who volunteered their homes for sampling. In this study, measurements reported for the independent cities of Falls Church and Fairfax were included with the county data. The radon values for Fairfax County range from zero to 347 picocuries of radon per liter (pCi/L) and have a mean of 3.28 pCi/L and a median of 1.90 pCi/L.

GIS Analysis

Portraying summary data in map form gives areal and spatial perspective to the results of radon sampling. A GIS such as ARC/INFO can produce plots of a zip code boundary file (fig. 1; Geographic Data Technology, Inc., 1986) that show the maximum and the mean indoor radon

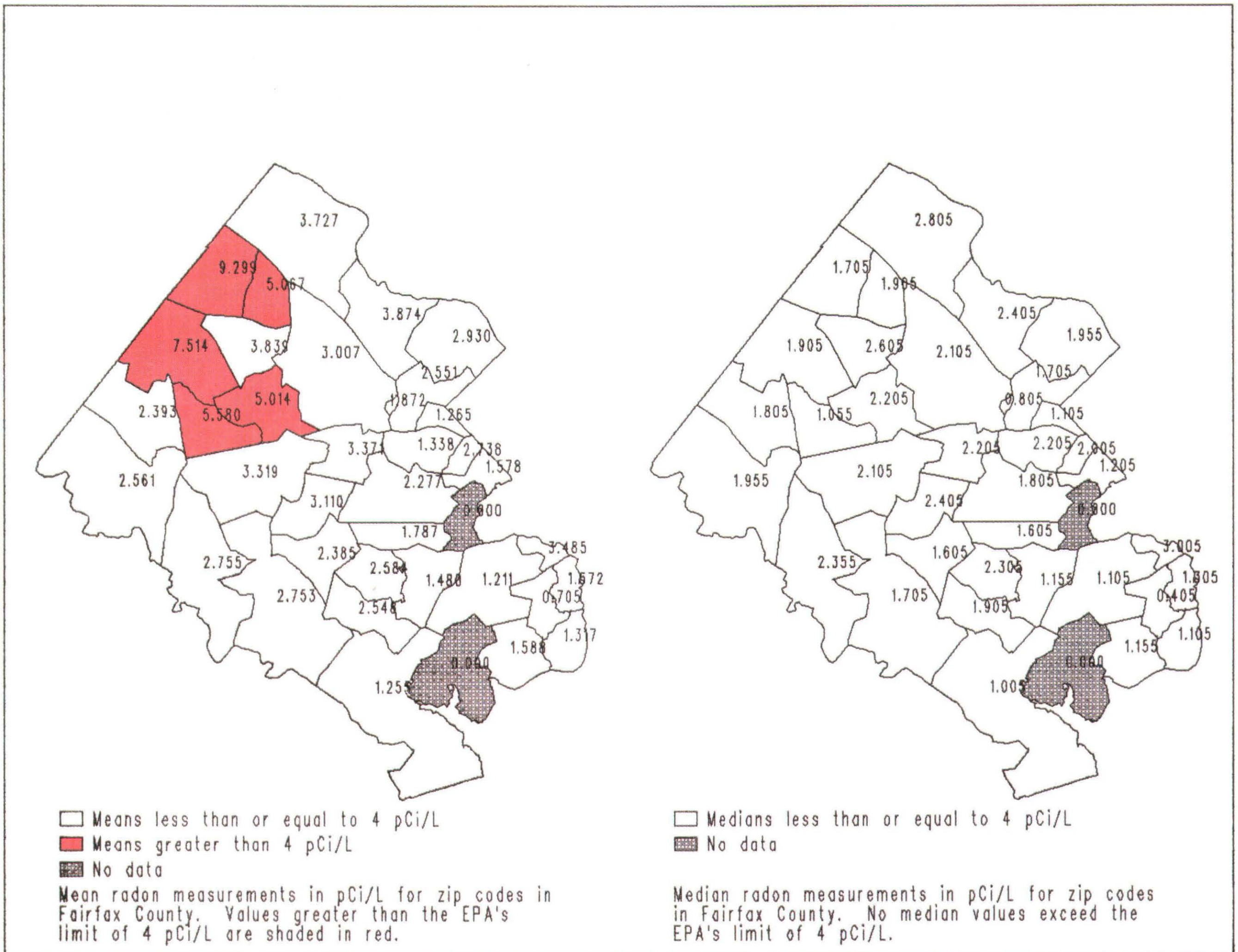


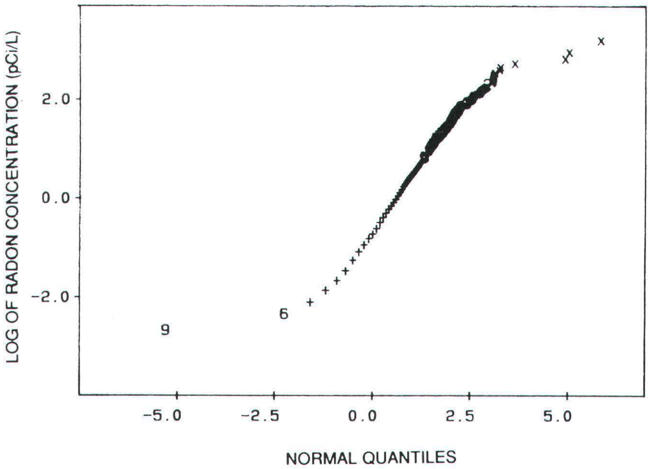
Figure 2. Mean and median radon values in pCi/L for Fairfax County, Va.

values. Then, from these plots, shaded maps of any specified threshold value can be constructed. The type of summary statistic used is important because it influences the conclusions drawn from the map. Mean values are affected by extreme values in a distribution. Thus the use of maximum or mean values alone can be misleading. For example, the radon map of the mean values appears different from that of the median values. Five zip codes are identified that have mean values exceeding the 4 pCi/L EPA action level for radon (fig. 2). No median values for the same zip codes exceed the EPA action level.

The median values reflect the more frequently found radon levels and are a better measure of overall risk. However, use of median values may conceal individual measurements that are of concern. An additional GIS plot of maximum values may be used to identify the presence of extreme values in a zip code area. Summary statistics, from EDA, coupled with a GIS provide a better method of examining the data.

Exploratory Data Analysis

EDA (Tukey, 1977; Chambers and others, 1983; Velleman and Hoaglin, 1981) techniques are useful in graphically portraying and describing the entire radon data set. Histograms of these data show a distinctly positive skew that has several extreme values. These extreme values, an order of magnitude greater than the majority of radon readings in the county, significantly increase the mean value of the data set. A more normally distributed data set to use for analysis is obtained by taking the logarithms of the original data set. Lognormal distributions are common for trace constituent data in the natural environment. Expressing the values as logarithms gives a more normal distribution as depicted in the quantile-quantile plot of the log-transformed data in figure 3. For this expression, quantiles, Q, or fractional proportions of the data (as defined by Chambers and others, 1983), are plotted against normal quantiles. A normal distribution results in a straight line.



NOTE: Multiple observations are shown by numerals in the plot

Figure 3. Quantile-quantile plot of the logarithms of radon in pCi/L for Fairfax County, Va.

Another useful EDA graphic is the box plot. Box plots of the log transformed data (fig. 4) summarize the radon readings for the 36 zip codes and allow the data distribution for the zip codes to be compared among each other. The center bar of each box plot is the median value for its data. Clearly the majority of the radon values in this data set are below the EPA action level of 4 pCi/L. The box plot is also used to identify areas of highest radon values. The box is drawn around the central 50 percent of the data, the interquartile range or spread, from the first quartile (Q1) to the third quartile (Q3). Those zip codes for which the upper quartile, or 25 percent of the sites, exceed the EPA action level of 4 pCi/L are of particular concern for radon risk. Thus the upper quartile values can be more descriptive than the median values for risk assessment.

Other useful graphics of the box plot are the whiskers, drawn from the quartiles to the highest and lowest value within 1.5 times the interquartile range. These values are descriptively defined as the adjacent values. Data points

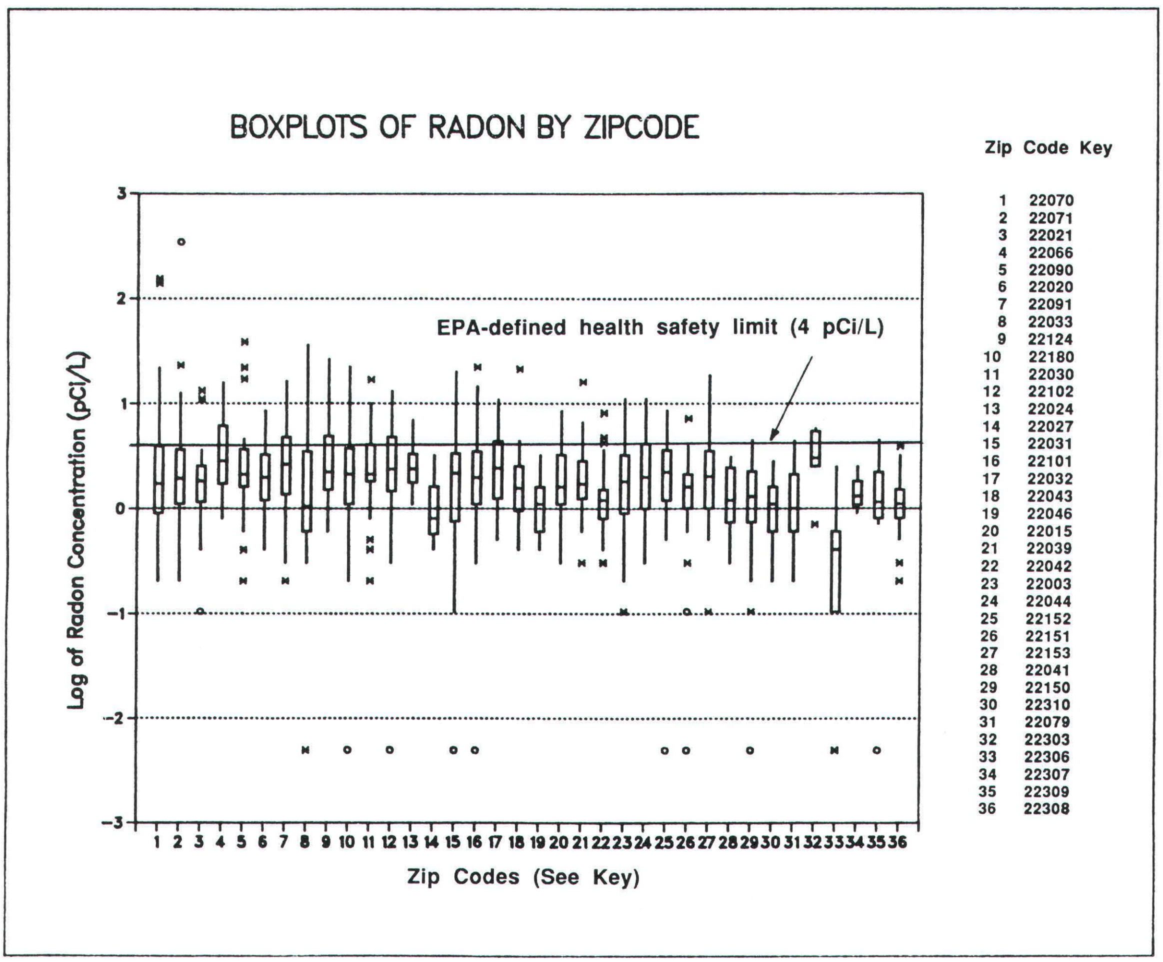


Figure 4. Box plots of log transformed radon values aligned along a cross-county transect, by zip code, for Fairfax County, Va.

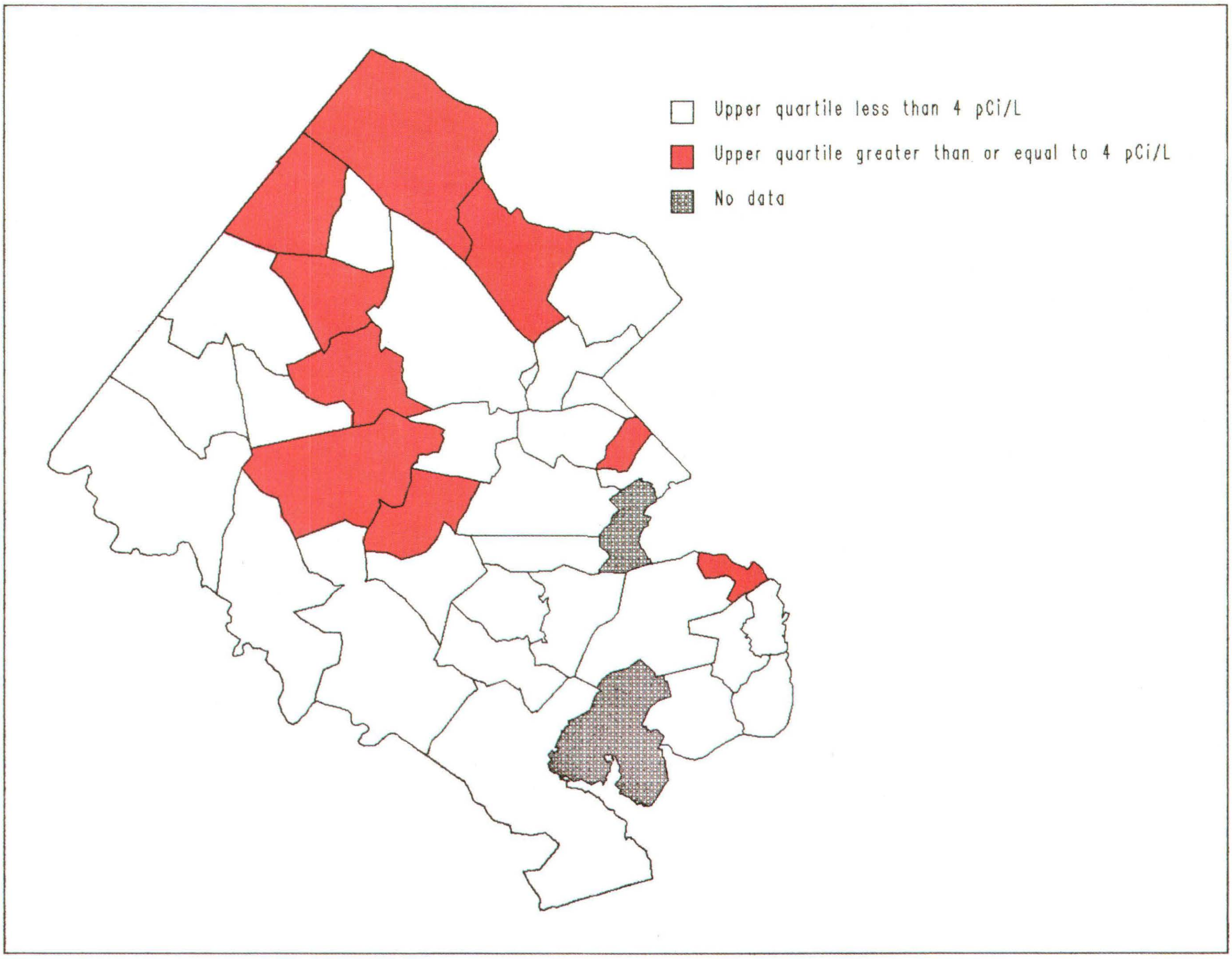


Figure 5. Zip code zones that have upper quartiles of radon measurements greater than 4 pCi/L for Fairfax County, Va.

beyond the adjacent values are identified as outside, and those more than 3 times the interquartile range are known as the far outside values. The outside and far outside values are the extreme values in the data set. The adjacent high value for the entire Fairfax County data set (1,408 values) is 6.675 pCi/L. Those readings beyond the adjacent high value may be indications of individual homes or possibly clusters of individual homes (hot spots or clusters) that have serious radon contamination.

Combining GIS with EDA

A GIS allows the analyst to portray these summary statistics in map form and to overlay them with other mapped attributes such as geology and soil. Figure 5 is an ARC/INFO plot of those zip codes for which the upper quartile is at or exceeds the EPA action level of 4 pCi/L. Figure 6 is a plot of those zip codes that have outside or far outside radon readings beyond the upper adjacent value for Fairfax County of 6.675 pCi/L. These are the zip codes that have extreme high values for Fairfax County as a whole. Such extreme high values are found only in the western two-thirds of the county.

To test the hypothesis, suggested by figure 6, that the zip codes in the western portion of the county yield significantly higher indoor radon values than those in the eastern portion, box plots of radon values for each zip code were compared. From ARC/INFO, a centroid for each zip code was identified from the zip code file and ordered from west to east along a transect drawn orthogonally to the Fall Line. Box plots of the data organized by zip code were then aligned along this transect for comparison. A natural break occurs between the 27th and the 28th box plot of figure 4. The Mann-Whitney test for difference of medians for these two sets (plots 1–27 vs. plots 28–36) shows a statistically significant difference between the western and eastern set at the 0.01 probability level. The nonparametric Mann-

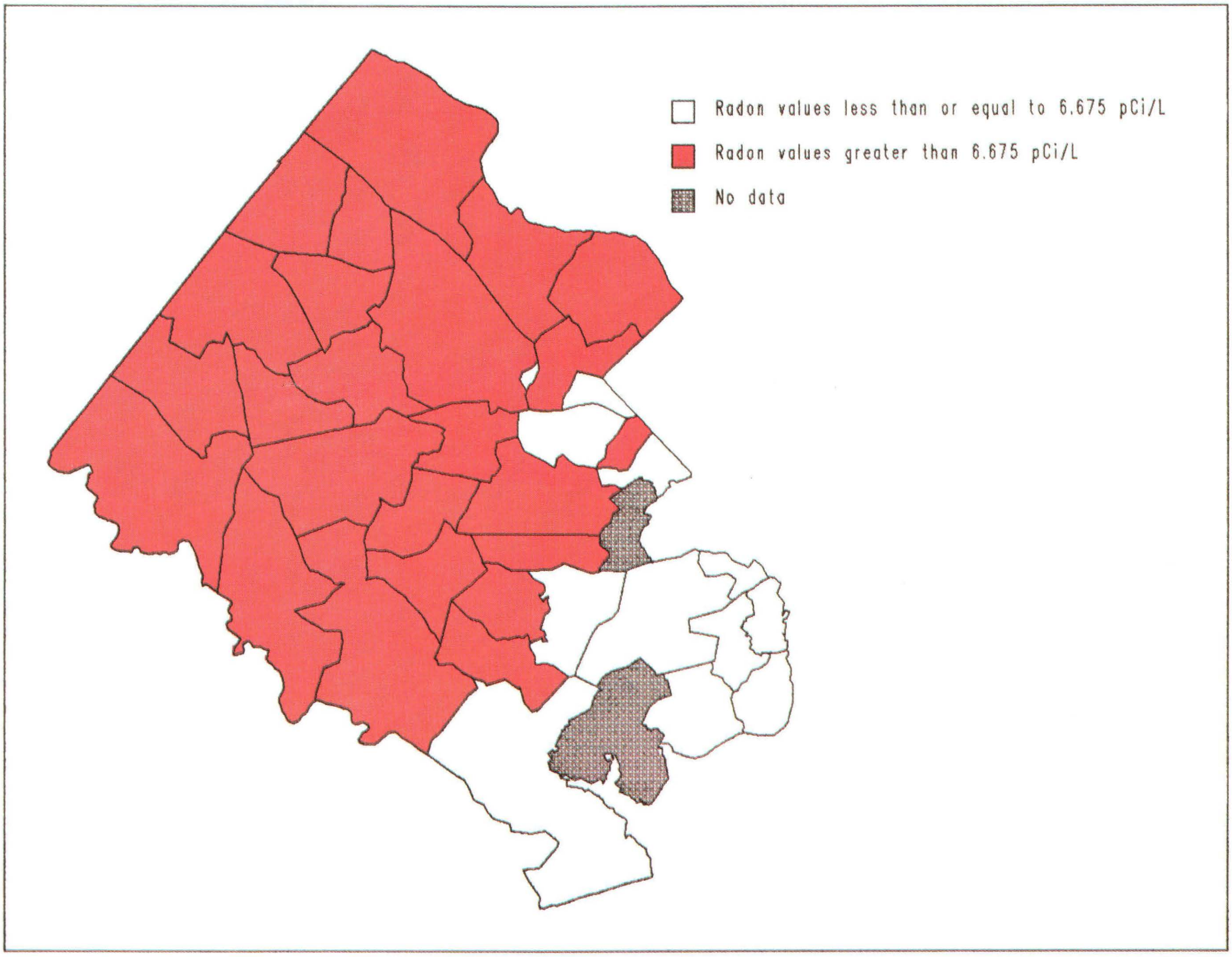


Figure 6. Zip code zones that have radon values above the adjacent high value of 6.675 pCi/L for Fairfax County, Va.

Whitney test was employed because the data could not be assumed to be normally distributed.

The geologic coverage in ARC/INFO overlain on the zip code boundaries enables a direct comparison of the geology with the statistical findings. At the time of this study, the U.S. Geological Survey was in the process of digitizing the Geologic Map of the United States (King and Beikman, 1974). The eastern half of the map was digitized and converted into an ARC/INFO cover. For this study, the geology of Fairfax County was extracted in ARC/INFO and geometrically intersected with the zip code file (fig. 7). The resulting output data set contained polygons that have two identifiers for each polygon, the geologic name and the zip code number. The geology in each zip code was calculated as a percent of the total area of the zip code polygon, and the percentages were generated by computer. Zip codes were identified with the geology underlying the major portion of the zip code polygon, and a comparison of the geology with the radon data set was possible. A visual comparison of figures 6 and 7 shows a distinct correspondence between the natural geologic boundary and the change in median radon values. The higher values occur in the Piedmont metamorphic rock, and the lower values occur in the Coastal Plain sediment.

Radon values for the western two-thirds of the county show additional differences. Three zip codes, located along the western edge of the county on Triassic sandstone and shale, appear from the box plots to have lower radon values than those on the metamorphic rock of the Piedmont. The Piedmont itself shows a notable bias towards higher radon values in the western zip codes. For example, zip codes that have upper quartiles greater than 4 pCi/L mainly occur in the central and northern part of the Piedmont (fig. 5).

An additional GIS data layer was obtained in ARC/INFO format from the U.S. Department of Agriculture, Soil Conservation Service (1987) State Soil Geographic Data Base. The soil data illustrates additional reasons for the differences within the Piedmont. Soil data

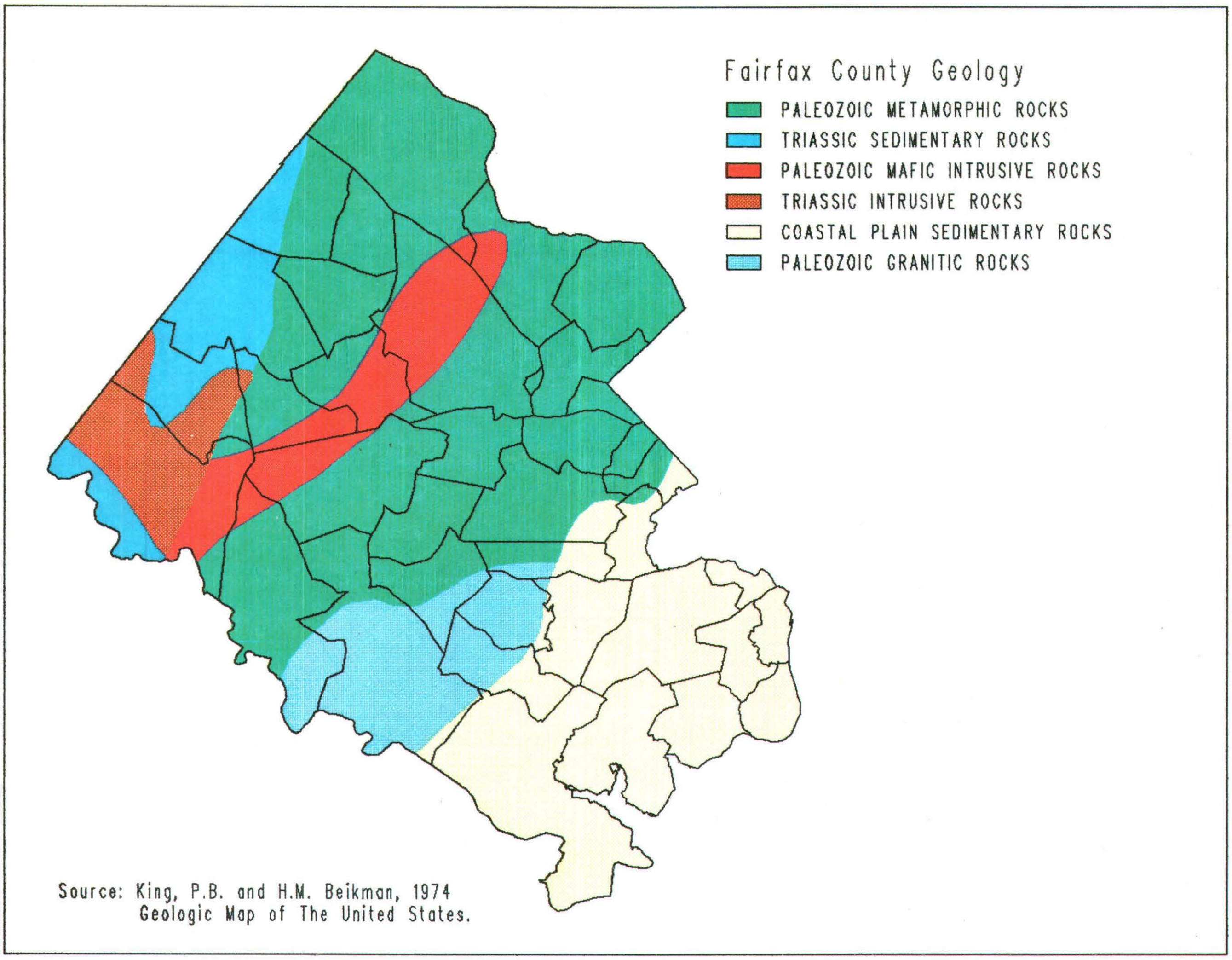


Figure 7. Geology and zip code boundaries of Fairfax County, Va.

for the Piedmont in Fairfax County can be divided into two soil provinces—soil formed on crystalline rock of the Piedmont Upland and soil formed on mixed crystalline rock and older Coastal Plain sediment (fig. 8).

ASSESSMENT OF RADON POTENTIAL

A Mann-Whitney test of differences between radon values for zip codes of the Piedmont, separated into the two soil provinces, gives a probability (p) value of 0.001. Radon values for zip codes located on the Piedmont Upland soil in the west are significantly higher than those on the mixed crystalline rock and older Coastal Plain sediment in the east.

Evaluating the data by zip code and comparing the zip code statistics with the underlying geology and soil data identifies four regions—Triassic basin, Coastal Plain, eastern Piedmont, and Piedmont Upland—that have varying radon values controlled by geology and soil. Mann-Whitney test results for differences among the four regions are seen in table 1. The Triassic basin is significantly different from the eastern Piedmont ($p=0.048$) and also significantly different from the Piedmont Upland ($p=0.092$) as compared to the other significance levels of $p=0.001$. For this reason the Triassic basin is assigned to a separate category of moderately high. Data summaries for these four radon potential units are shown in table 2. The resultant map is portrayed in figure 9. These results are similar to an assessment of radon potential for Fairfax County by Otton and others (1988).

SUMMARY AND CONCLUSIONS

EDA and GIS techniques can be combined to help evaluate the statistical characteristics of a large data set (indoor radon data), to look for relations within the data set

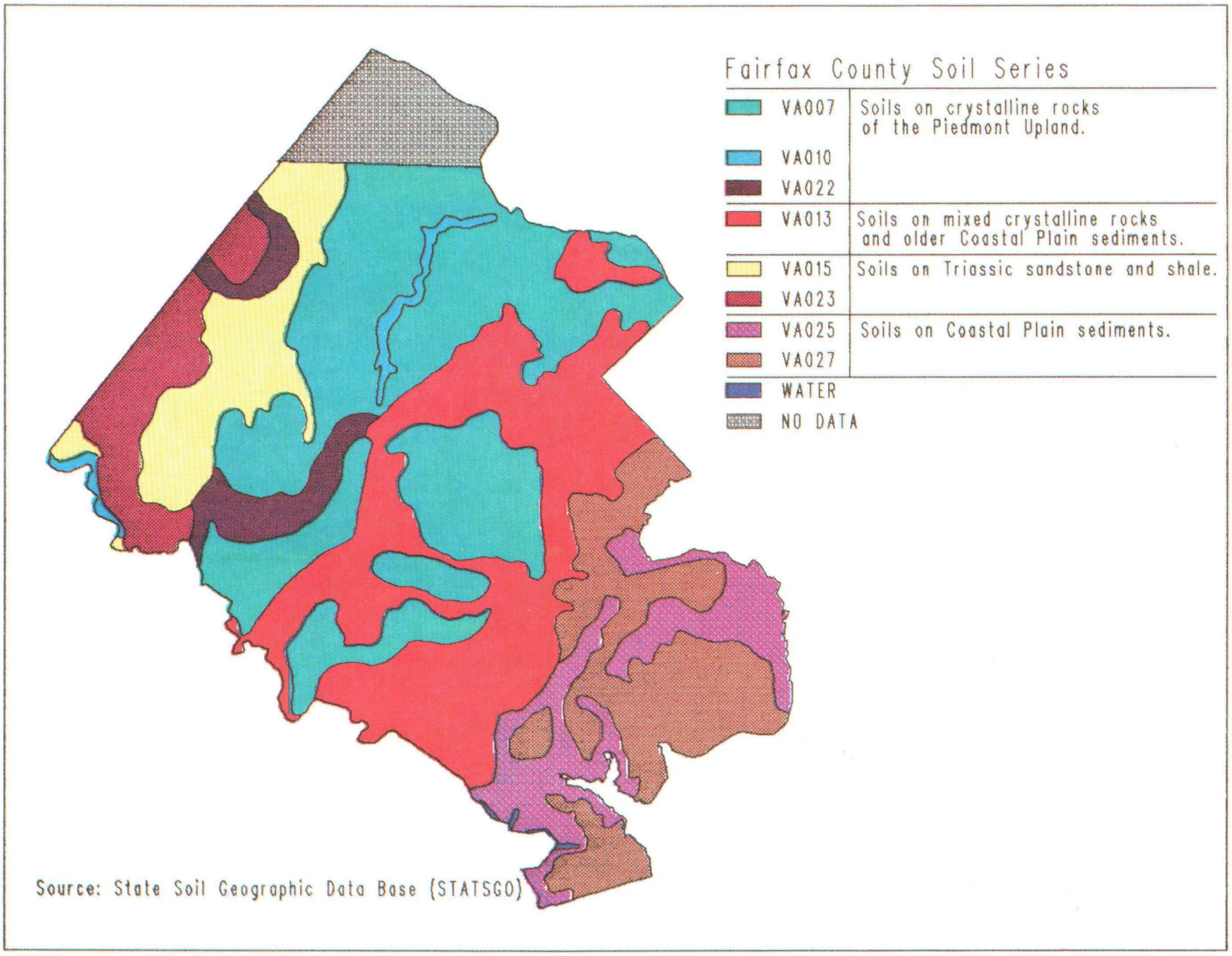


Figure 8. Soils of Fairfax County, Va.

Table 1. Results of Mann-Whitney tests for differences among regions of Fairfax County, Va.

Region	Sample size	Median	*p*-level
Western county[1] vs.	1290	1.90	
Coastal Plain	118	1.10	0.001
Piedmont Upland vs.	732	2.10	
eastern Piedmont	397	1.60	.001
Triassic basin vs.	161	1.90	
Piedmont Upland	732	2.10	.092
Triassic basin vs.	161	1.90	
eastern Piedmont	397	1.60	.048

[1]Western county refers to that part of the county west of the Fall Line and includes the Piedmont Upland, the Triassic basin, and the eastern Piedmont.

(spatial variations within the data), to compare the data set to information in other GIS layers (geology and soil maps), and to create derivative layers (a radon potential map). This study shows that elevated radon levels in homes in Fairfax County are associated with the presence of metamorphic rock. The greatest radon values are associated with the metamorphic rock of the Piedmont Upland. The metamorphic rock to the west probably contains more radium in the soil or is more permeable to the flow of soil gas or both. In the eastern Piedmont, the soil may be more deeply weathered, contain less radium, and be less permeable; thus, less radon is generated, and the flow of radon gas is restricted. The Triassic basin also appears to have elevated radon potential, but the soil may also slow the migration of radon gas indoors. In the east, the rock and soil characteristics of the Coastal Plain yield lower radon readings and lower radon potential overall.

Although the most accurate way of establishing correlations between geologic parameters and indoor radon

Table 2. Summary data for the four regions of Fairfax County, Va., in pCi/L.

Region	Samples	Mean	Median	Maximum	Q1[1]	Q3[1]	Radon potential
Triassic basin................	161	4.83	1.90	347.00	1.16	3.11	Moderately high.
Piedmont Upland.............	732	3.81	2.10	155.00	1.20	3.98	High.
Eastern Piedmont.............	397	2.23	1.60	21.40	.90	2.80	Moderate.
Coastal Plain	118	1.39	1.10	5.80	.60	2.00	Low.

[1]Q1 and Q3 refer to the first and third quartiles and represent the radon values at the 25th and 75th quantiles, respectively.

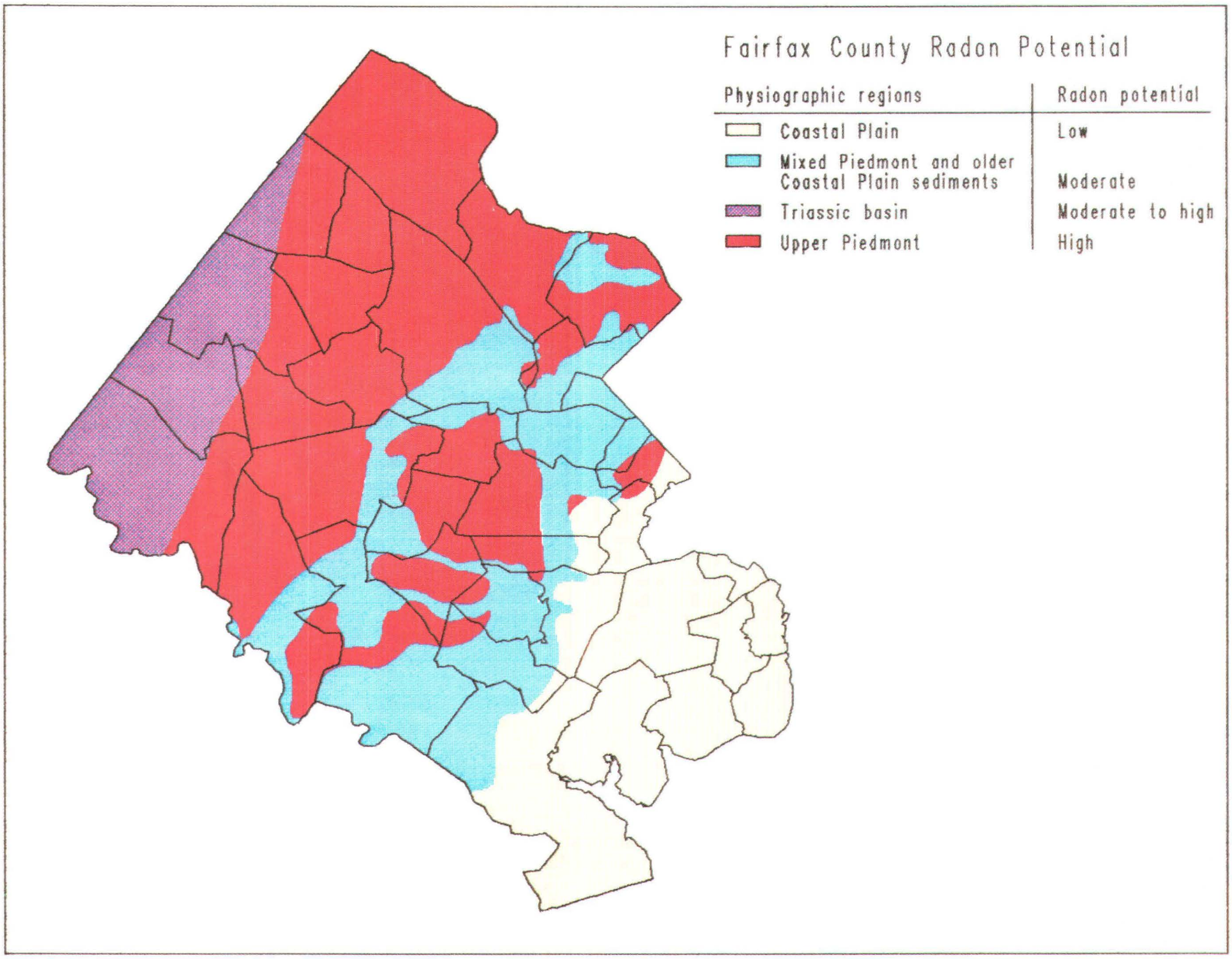


Figure 9. Radon potential of Fairfax County, Va.

data is to examine site-specific data, the large zip-code-based indoor radon data sets can be studied effectively in a GIS environment. Through the use of exploratory data analysis and robust statistical techniques, inferences about the relative radon levels may then be related to other GIS layers such as geology and soil. Comparison of derivative radon potential maps to other possible GIS layers, such as land-use maps, may facilitate land-use planning. For instance, school administrators trying to estimate construction costs on new school sites could compare a school site GIS layer to a radon potential map to see which school sites would require additional testing or preventive measures and thus extra costs.

The methods presented in this paper have broader application to analysis and presentation of environmental data. Complementary use of a GIS, providing perspective on the spatial distribution of a geographic data set, with EDA, providing perspective on the statistical characteristics of the data set, supplies meaningful results when correlating multiple layers in a GIS. In this study, the statistical analysis was compiled outside the GIS, and then the GIS was utilized to display and manipulate the data layers.

REFERENCES CITED

Akerblom, Gustav, Andersson, Per, and Clavensjo, B., 1984, Soil gas radon—A source for indoor radon daughters: Radiation Protection Dosimetry, v. 7, p. 49–54.

Chambers, J.M., Cleveland, W.S., Kleiner, Beat, and Tukey, P.A., 1983, Graphical methods for data analysis: Boston, Mass., Duxbury Press, 395 p.

Eaton, R.S., and Scott, A.G., 1984, Understanding radon transport into houses: Radiation Protection Dosimetry, v. 7, p. 251–253.

Fairfax County, 1987, Fairfax County radon project—Interim report: Fairfax County Staff Memo to the Board of Supervisors, July 27, 1988, 20 p.

Geographic Data Technology, Inc., 1986, Five-digit zip code boundary file user documentation: Lyme, N.H., Geographic Data Technology, Inc., 12 p. [revised, October 1986].

Gundersen, L.C.S., Reimer, G.M., Wiggs, C.R., and Rice, C.A., 1988, Map showing radon potential of rocks and soils in Montgomery County, Maryland: U.S. Geological Survey Miscellaneous Field Studies Map MF–2043, scale 1:62,500.

King, P.B., and Beikman, H.M., 1974, Geologic map of the United States: U.S. Geological Survey, scale 1:2,500,000.

National Research Council, 1988, Health risks of radon and other internal deposits of alpha emitters: BIER IV, Washington, D.C., Committee on the Biological Effects of Ionizing Radiations, National Academy Press, 602 p.

Nero, A.V., Jr., and Nazaroff, W.W., 1984, Characterizing the source of radon indoors: Radiation Protection Dosimetry, v. 7, p. 23–39.

Otton, J.K., Schumann, R.R., Owen, D.E., Thurman, Nelson, and Duval, J.S., 1988, Map showing radon potential of rocks and soils in Fairfax County, Virginia: U.S. Geological Survey Miscellaneous Field Studies Map MF–2047, scale 1:48,000.

Sachs, H.M., Hernandez, T.L., and Ring, J.W., 1982, Regional geology and radon variability in buildings: Environmental International, v. 8, p. 59–66.

Tanner, A.B., 1964, Radon migration in the ground, *in* Adams, J.A.S., and Lowder, W.M., eds., The natural radiation environment: Chicago, Ill., University of Chicago Press, p. 161–190.

———1980, Radon migration in the ground, *in* Gesell, T.G., and Lowder, W.M., eds., Natural radiation environment III, Conference-780442: Springfield, Va., National Technical Information Service, p. 5–56.

Tukey, J.W., 1977, Exploratory Data Analysis: Reading, Mass., Addison-Wesley, 688 p.

U.S. Environmental Protection Agency and U.S. Public Health Service Centers for Disease Control, 1986, A citizen's guide to radon: What it is and what to do about it: U.S. Environmental Protection Agency and U.S. Public Health Service Centers for Disease Control, OPA–86–004, 14 p.

U.S. Soil Conservation Service, 1987, State soil geographic data base (STATSGO): Fort Worth, Tex., National Cartographic Center, U.S. Soil Conservation Service.

Velleman, P.F., and Hoaglin, D.C., 1981, Applications, basics, and computing of exploratory data analysis: Boston, Mass., Duxbury Press, 354 p.

CHAPTER F

Automation of Data Systems—Minnesota's Approach for Water-Use Data

By LEE C. TROTTA

Prepared in cooperation with the
Minnesota Department of Natural Resources

U.S. GEOLOGICAL SURVEY BULLETIN 1908

SELECTED PAPERS IN THE APPLIED COMPUTER SCIENCES 1990

CONTENTS

TABLE

Automation of Data Systems—Minnesota's Approach For Water-Use Data

By Lee C. Trotta

Abstract

A computer-based relational data-base-management system was chosen and developed jointly by the Minnesota Department of Natural Resources and the U.S. Geological Survey to automate data-base management of water for the State and national programs. The system can identify (1) users in a specific location, (2) sources of water, (3) types of water use, and (4) rates and volumes of water used. Data were collected for 10 different categories of water use for the U.S. Geological Survey National Water-Use Information Program. Summaries of water use within a county or watershed were required to transfer these data to the U.S. Geological Survey Aggregated Water-Use Data System. Data aggregation programs to provide these summaries, a simulation model based on Standard Industrial Classification Codes to determine the economic value of water, and various utility programs to speed processing and analysis have been developed.

INTRODUCTION

Users of large amounts of water may affect the quantity and quality of water available to other users. Many States monitor the water use of their larger users. In Minnesota, water-use information is collected by various agencies to satisfy legislative mandates. Because of the complexity and scope of water use, management of water resources requires classification of data and automation of data entry, processing, and retrieval. The Minnesota Water-Use Data System (MWUDS), an example of computer methods and philosophy, has been used successfully to automate a data base and data system. The MWUDS was developed independently from the State Water-Use Data System distributed by the U.S. Geological Survey (USGS).

Permitting systems are used to obtain site-specific water-use information in Minnesota. The State of Minnesota has established one permit system for water appropriations greater than 10,000 gallons per day (or 1 million gallons per year, whichever is less) and another system to monitor the quality of water discharged. Water appropriations (withdrawals) are reported to the Minnesota Department of Natural Resources (MDNR). Compliance with the reporting provision of permits is about 85 percent by irrigation users and 95 percent by all other users (Young, 1987, p. 6). A water-quality permit system regulated by the Minnesota Pollution Control Agency provides limited data on water returned to the environment at sewage-treatment plants.

Manuscript approved for publication February 10, 1989.

COMPUTER METHODS

Thirty-nine years passed between issuance of the first Minnesota water-use permit and development of a rudimentary automated data-storage system in 1976. During that time, information needed by State planners was retrieved by leafing through thousands of pages of data in notebooks (Gil Young, MDNR, written commun., 1987). In 1982, the MDNR and the USGS began a joint project to implement a water-use data-base-management system (Horn, 1986). After a long search for applicable computer-based systems, a relational data-base-management system was chosen that allows staff that has minimal training or computer experience to input, update, and retrieve data. Members of the Land Management Information Center (LMIC), part of the Minnesota State Planning Agency, helped design the data-base structure and programs needed. Menu-driven programs provide users access to water-use data for approximately 11,500 withdrawal sites currently (1988) in the system.

Data-management philosophy distinguishes MWUDS from data-storage systems in other states. Data storage is not the only purpose for the Minnesota water-use program. Specific retrievals are often needed. Water-use managers need to identify users in a specific location in the State, describe the source of water (such as well, lake, or stream) and the type and rate of use. Summaries of water use within a county, a watershed, or statewide were needed for input into the USGS Aggregated Water-Use Data System (AWUDS). A philosophy of gearing data-system design and development toward more efficient multi-purpose retrievals lent management support to continuing

Table 1. U.S. Geological Survey water-use categories for which permit data are collected by the Minnesota Department of Natural Resources

Code(s)	Water-use category
20, 22–26	Thermoelectric power generation
21	Hydroelectric power generation
54, 80–86, 90, 96	Irrigation
30–35, 73	Commercial
40–42, 46–48, 50–52	Industrial
43, 44, 62–64	Mining
10–16, 53	Public supply
45	Sewage treatment

automation. When the work force is well trained in the data-base management system and the data-handling processes are well documented, the data processors are motivated to do programming, which further automates slow and tedious retrievals or manual operations.

One of the first accomplishments of MWUDS automation was the development of data-aggregation programs (Minnesota LMIC, 1984; Trotta, 1988a), which are useful for calculating county and watershed water-use summaries. After verification, this aggregate information can be automatically transferred to AWUDS. These data summaries are useful for analyzing regional water-use trends.

The MDNR stores information from permit reports in 8 (table 1) of 10 water-use categories defined by the USGS. The ten categories are thermoelectric power generation, hydroelectric power generation, irrigation, livestock, commercial, domestic, industrial, mining, public supply, and sewage treatment. However, the MDNR has defined about 50 subcategories to allow more detailed analysis (table 2). The MDNR also collects data on aquaculture (code 72), real-estate quality improvement (codes 60, 61, and 71), and preservation (codes 60, 61, and 74). These categories are not part of the current focus of the USGS. Relations among water-use categories specified by both the MDNR and the USGS were defined by Horn (1984, p. 4) and are revised and expanded upon in table 1. Some USGS-specified categories in table 1 contain only limited data (hydroelectric power generation and sewage treatment) or are absent (livestock and domestic) because of the nature of the MDNR withdrawal-permit reporting requirements. Withdrawals for three of these USGS-specified categories (hydroelectric power generation, livestock, and domestic) have been estimated by the USGS annually since 1984 and on a 5-year schedule before that (Trotta, 1988b).

Each of the water-use categories corresponds to a group of related Standard Industrial Classification (SIC) codes (Snavely, 1986, p. 152–160). SIC codes define and classify the entire field of economic activities (Office of Management and Budget, 1972). These SIC codes are included on tax forms used by businesses and on surveys made by the U.S. Department of Commerce, Bureau of the Census. Individuals or facilities thus are categorized on the basis of the purpose of their water use in relation to their economic activity. Published use-ratio surveys, such as that by Snavely (1986, p. 182–187), when combined with economic reports, such as those published by the Minnesota Department of Jobs and Training (1986), allow the use of economic-policy-analysis simulation models to estimate water-use amounts for each SIC code (Young, 1987, p. 5–9). These estimates are used when actual data are not available.

Combining information in this way from several data bases requires a linkage or relation between them. In some cases, stand-alone data bases, such as the Well Log Listing System (WELLS) or the Permit Information Exchange (PIX), are purposely linked to MWUDS. WELLS contains a brief description (location, aquifer, and well-construction data) for approximately 100,000 well logs. Locational search routines are available to verify or transfer aquifer data about a particular user's source of water. Menu-driven systems recently have been developed. WELLS also has a routine to load data to a geographic information system. PIX contains detailed information from well permits on fees and legal actions for the regional water-data network. The permit-processing software in PIX does the name and address maintenance on permit applications. WELLS and PIX are linked either with a common geographic-location code or with reference numbers that help identify the same water user in different data bases (Horn, 1986, p. 11). One PIX program allows easy access to both historic and active data. Another menu-driven system uses information from PIX and MWUDS to produce written communications and forms for reporting pumpage data (Rick Gelbmann, MDNR, written commun., 1988). Computer linkage to other data bases is described by Horn (1986, p. 18–25).

Utility programs have been developed by MDNR to provide data for many specific applications. Some of the applications are to (1) aid in tracking permits once they have been issued, (2) ensure compliance with regulations or restrictions, and (3) assist State education functions by producing illustrative reports. The MDNR program INDEX, for example, is useful for producing reports that summarize 3 years of site-specific data for any MDNR administrative region (Rick Gelbmann, MDNR, oral commun., 1988). For example, search routines in MWUDS can outline quickly areas where water-level declines may be related to municipal lawn sprinkling or irrigation. The MDNR also has developed input programs for an economic model that determines the economic value of Minnesota water for various water-use scenarios. The economic value of conservation initiatives and changing priorities of use are discussed in a report that summarizes results of this model (Adelsman and Bloomgren, 1987, p. 19–22).

USGS utility programs provide information for trend analysis. The AWUDS software, distributed by the

Table 2. Water-use codes and categories used by the Minnesota Water-Use Data System

Code	Category used by the Minnesota Department of Natural Resources	Code	Category used by the Minnesota Department of Natural Resources
3	Well abandoned	45	Sewage treatment
4	Terminated	46	Petroleum (chemical processing)
5	Active not under permit	47	Metal processing
10	Waterworks	48	Nonmetallic products (rubber/plastic/glass)
11	Municipal	50	Temporary
12	Private waterworks (trailer courts/small housing units)	51	Construction (nondewatering)
13	Commercial and institutional	52	Construction (dewatering)
14	Cooperative waterworks	53	Pipeline and tank testing
15	Fire protection	54	Landscape watering
16	State parks/waysides/highway rest areas	60	Water level maintenance
20	Power generation	61	Basin (lake) level
21	Hydropower	62	Mine dewatering
22	Steam power cooling (once through)	63	Quarry dewatering
23	Steam power cooling (wet tower)	64	Sand/gravel pit dewatering
24	Steam power cooling (ponds)	65	Tile drainage and pumped sumps
25	Steam power (other than cooling)	71	Pollution confinement
26	Nuclear power	72	Hatcheries and fisheries
30	Air conditioning	73	Snow making
31	Commercial building	74	Peat fire control
32	Institutions (schools, hospitals)	80	Noncrop irrigation
33	Heat pumps	81	Golf course
34	Coolant pumps	82	Cemetery
35	District heating	83	Landscaping
40	Industrial processing	84	Sod
41	Food and livestock (agricultural processing)	85	Nursery
42	Paper/pulp	86	Orchard
43	Mine processing (other than sand and gravel washing)	90	Major crop irrigation
44	Sand and gravel washing	96	Wild rice irrigation

National Water-Use Information Program, accepts aggregated data from MWUDS and produces graphs; pie diagrams and choropleth maps of the State, depicting county or watershed data, can be produced. Another part of the AWUDS creates a table useful for the comparison of water-use data for 1980 (or another reliable base year) with any other year for which data are available (Howard Perlman, USGS, written commun., 1987). Additional programs were developed in the Minnesota District of the USGS to extract data needed by the national program from specialized data files associated with MWUDS. These programs calculate municipal population by watershed, sewage-treatment facilities by watershed, water withdrawals for livestock by watershed, and diversion of water across watershed boundaries (Trotta, 1988c). Timely predictions of the effect of diversions or changes in pumpage can prevent undue stress on the water resource.

SUMMARY

When water-use managers in Minnesota needed a system to store and retrieve data, the MDNR and the USGS jointly developed a computer-based relational data-base-management system. The relation to associated data bases allows the combination of geographic, aquifer, fee, and permit information. Through a philosophy of management support for documentation and training, processes are continually being automated so that retrievals and resultant decisions can be streamlined.

REFERENCES CITED

Adelsman, H.R., and Bloomgren, P.A., 1987, Summary—The economic value of water: St. Paul, Minn., Department of Natural Resources, Division of Waters Report to the Legislative Commission on Minnesota Resources, 22 p.

Horn, M.A., 1984, Annual ground-water use in the Twin Cities metropolitan area, Minnesota, 1970–79: U.S. Geological Survey Open-File Report 84–577, 130 p.

———1986, Development of a water-use data system in Minnesota: U.S. Geological Survey Water-Resources Investigations Report 85–4306, 59 p.

Minnesota Department of Jobs and Training, 1986, Average covered employment and wages by economic region for calendar year 1985 by two digits of industry: St. Paul, Minn., Minnesota Department of Jobs and Training, 45 p.

Minnesota Land Management Information Center, 1984, State water use data system (SWUDS) system documentation: St. Paul, Minn., State Planning Agency, 42 p.

Office of Management and Budget, 1972, Standard industrial classification manual: Washington, D.C., Executive Office of the President, Office of Management and Budget, Statistical Policy Division, 649 p.

Snavely, D.S., 1986, Water-use data-collection programs and regional data base of the Great Lakes-St. Lawrence River Basin States and Provinces: U.S. Geological Survey Open-File Report 86–546, 204 p.

Trotta, L.C., 1988a, Aggregation of Minnesota water-use data and transfer of data to the National Water-Use Data System—Procedures and programs: U.S. Geological Survey Open-File Report 87–40, 16 p.

———1988b, Sources of water-use data in Minnesota: U.S. Geological Survey Open-File Report 87–544, 2 p.

———1988c, Inventory of interbasin water transfers in Minnesota: U.S. Geological Survey Open-File Report 88–466, 15 p.

Young, P.G., 1987, An analysis of water use in Minnesota: St. Paul, Minn., Department of Natural Resources, Division of Waters Report to the Legislative Commission on Minnesota Resources, 49 p.

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 7.5- 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 7.5-minute quadrangle photogeologic maps on planimetric bases that 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" are 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.