U.S. GEOLOGICAL SURVEY NATIONAL COMPUTER TECHNOLOGY MEETING: PROCEEDINGS, PHOENIX, ARIZONA, NOVEMBER 14-18, 1988

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 90-4162

U.S. GEOLOGICAL SURVEY NATIONAL COMPUTER TECHNOLOGY MEETING: PROCEEDINGS, PHOENIX, ARIZONA, NOVEMBER 14-18, 1988

By Barbara H. Balthrop and John E. Terry, editors

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 90-4162

USGS
science for a changing world

Click here to return to USGS Publications

Nashville, Tennessee

1991

U.S. DEPARTMENT OF THE INTERIOR
MANUEL LUJAN, JR., Secretary

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

For additional information write to:

Coordinator, National Computer Technology Meeting Proceedings 1988
U.S. Geological Survey
300 W. Congress
Federal Building, FB-44
Tucson, Arizona 85701

Copies of this report can be purchased from:

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

FOREWORD

The U.S. Geological Survey National Computer Technology Meetings (NCTM) are sponsored by the Water Resources Division and provide a forum for the presentation of technical papers and the sharing of ideas or experiences related to computer technology. This report serves as a proceedings of the meeting held in November, 1988 at the Crescent Hotel in Phoenix, Arizona. The meeting was attended by more than 200 technical and managerial people representing all Divisions of the U.S. Geological Survey.

Scientists in every Division of the U.S. Geological Survey rely heavily upon state-of-the-art computer technology (both hardware and software). Today the goals of each Division are pursued in an environment where high speed computers, distributed communications, distributed data bases, high technology input/output devices, and very sophisticated simulation tools are used regularly. Therefore, information transfer and the sharing of advances in technology are very important issues that must be addressed regularly.

This report contains complete papers and abstracts of papers that were presented at the 1988 NCTM. The report is divided into topical sections that reflect common areas of interest and application. In each section, papers are presented first followed by abstracts. For these proceedings, the publication of a complete paper or only an abstract was at the discretion of the author, although complete papers were encouraged.

Some papers presented at the 1988 NCTM are not published in these proceedings.

John E. Terry

CONTENTS

CHAPTER E--DATA BASES AND AUTOMATED DATA HANDLING

CHAPTER F--ELECTRONIC PUBLISHING/MANUSCRIPT PREPARATION

CONVERSION FACTORS

Multiply	By	To obtain
inch (in.)	25.4	millimeter
inch per day (in/d)	25.4	millimeter per year
foot (ft)	0.3048	meter
foot per day (ft/d)	0.3048	meter per day
foot per mile (ft/mi)	0.1894	meter per kilometer
cubic foot per second (ft^3/s)	0.02832	cubic meter per second
mile (mi)	1.609	kilmometer
square mile (mi^2)	2.590	square kilometer
mile per square mile (mi/mi^2)	0.621	kilometer per square kilometer
gallons per day (gal/d)	0.003785	cubic meter per day
million gallons (Mgal)	3,785	cubic meter
million gallons per day per square mile [$(Mgal/d)mi^2$]	1,460	cubic meter per day per square kilometer

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F = 1.8 \times °C + 32$$

The use of brand, company, or trade names in this report is for indentification purposes only and does not constitute endorsement by the U.S. Geological Survey.

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 report series for the National Water Summary. 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 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.

GRAPHICS ON MICROCOMPUTERS

By

R.T. Hanson

U.S. Geological Survey

Tucson, Arizona

ABSTRACT

The use of the graphics package PC-DISSPLA[1] has extended some of the capabilities of computer-graphics programs from mainframe computers to microcomputers. PC-DISSPLA is based on the VAX version of DISSPLA and has advantages and limitations with respect to the mainframe version. This package is used in the Arizona District of the U.S. Geological Survey for preparation of x-y plots, probability plots, text slides, and contour maps. A major advantage over the mainframe time-sharing environment is the ability to have redundancy and stand-alone graphics independent of mainframe processes, which can degrade the performance of graphics programs. Another advantage of processing on a microcomputer is the ability to transfer metafiles to and from mainframe computers; this takes advantage of hardware and software capabilities on both machines for post-processing tasks. PC-DISSPLA used in combination with the Graphic Software Systems peripheral-device library provides a wide variety of graphic devices including raster devices, plotters, and the post-script interface. Future versions of PC-DISSPLA will take advantage of enhancements in DISSPLA Revision 11.0 and the new Microsoft OS-2 environment. The major limitation with PC-DISSPLA is the size of programs that can be run on microcomputers in the DOS environment.

APPLICATION OF USER-SUPPLIED TRANSFORMATIONS IN COMPUTER-GRAPHICS PROGRAMS

By

Stanley A. Leake

U.S. Geological Survey

Tucson, Arizona

ABSTRACT

User-supplied transformations of plot vectors are a feature of the CA-DISSPLA 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.

CHAPTER E--DATA BASES AND AUTOMATED DATA HANDLING

A VERTICAL SEQUENCE CORRELATOR MODEL (VERSECORR) WHICH IS USED IN THE RECOGNITION OF GEOPHYSICAL LOG SHAPES

by Merribeth Bruntz and A. Curtis Huffman, Jr.

ABSTRACT

Geophysical well log interpretation and correlation commonly are time consuming and subjective processes. Many studies involve analysis of hundreds of geophysical logs of differing types (such as, self-potential, resistivity, and gamma rays) and qualities. A pattern associator model has been developed that identifies a test set of specific log shapes. Neural computing techniques are used because of the large degree of generalization that must be performed in this identification.

Use of a neural network architecture that allows for the learning of 12 log shapes presented as part of the training set is demonstrated. Testing was performed on 20 patterns that are not part of the training set, with 15 of the test patterns succesfully identified.

INTRODUCTION

One of the most common forms of data collection and presentation is as x-y plots where intensity of a variable is displayed sequentially against time, distance, and depth. In the geological sciences, this type of plot has many applications, but perhaps the most familiar and frequently used is the geophysical well log. Since 1927, when the technique was invented in France, the well log has become an important analytical tool of the petroleum industry and is the basis of nearly any hydrocarbon exploration or development program. Although logging and analysis techniques have improved greatly, especially in the last 10 to 15 years, geologic interpretation and correlation of logs remain an art to a large extent and largely depend on the ability of the interpreter to recognize patterns.

With this in mind, the goal of this project is to develop a computer model that offers an objective interpretation of these patterns. Neural computing techniques are being used to explore the feasibility of using methods that simulate human intelligence more closely than those used in symbolic artificial intelligence. The acronym, VERSECORR, for Vertical Sequence Correlator, represents the requirement that the model recogize and interpret vertical sequences, and also correlate them with each other. In this paper we will discuss the design and implementation of a program that recognizes well log shapes, which is the first step in recognizing and interpreting vertical sequences.

NEURAL NETWORK ARCHITECTURE

Neural network models that perform specialized tasks, such as pattern recognition, have been successfully simulated on digital computers. Like the human brain, our model is composed of processing cells (neurons) and connections (synapses) between the processing cells. The processing cells function as memory units that contain pieces of information. The connections between the processing cells connect the pieces of information together, forming a network of neuron-like units , u_i and v_j (fig. E1).

Each connection in the network has a numerical weight, w_{ij}, that corresponds to the influence of unit v_j on unit u_i. Positive weights indicate reinforcement, and negative weights represent inhibition. A subset of units in the network, $v_1,....,v_n$, is considered as network inputs. These network inputs have no entering connections. The outputs of the network, $u_1,....,u_m$, form another subset of units in the network called output units.

The neural network distinguishes associations between input patterns and output patterns. In general, this type of model is called a pattern associator and is one of the simplest of neural models.

Each of the neuron-like units in the network, whether input or output, has a numerical value called an activation value. The activation value for an output unit, u_i, is based upon the activation values of the input units, v_j, directly connected to it and the corresponding weights for these connections. The activation values of the input units are discrete, taking on the values {+1, or -1}. Every output unit, u_i, computes its new activation, $\hat{u}_i$, as a function of the weighted sum of the input to unit, u_i, from directly connected units. Thus,

$$S_i = \sum_{j=1}^{n} w_{ij} v_j$$

and,

$$\hat{u}_i = f(S_i).$$

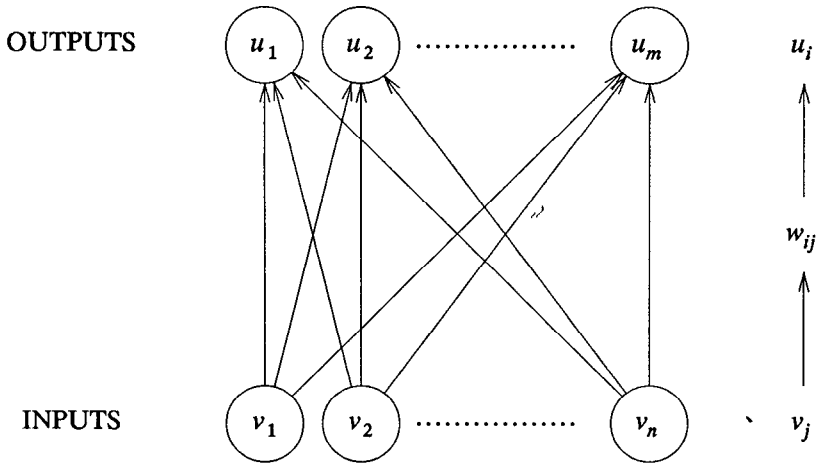


Figure E1.--Neural network architecture of a pattern associator model.

The activation of the input units are fixed by externally supplied input patterns. The activation value of a specific input unit, v_j, is determined by an interface that uses a linear least-squares approximation to determine features of the top, side, and bottom of a geophysical log shape. Some of these features include facing left or right, horizontal position, vertical position, or degree of slope (fig E2). The uniqueness of these features for a specific log shape is then used as the externally fixed input activations.

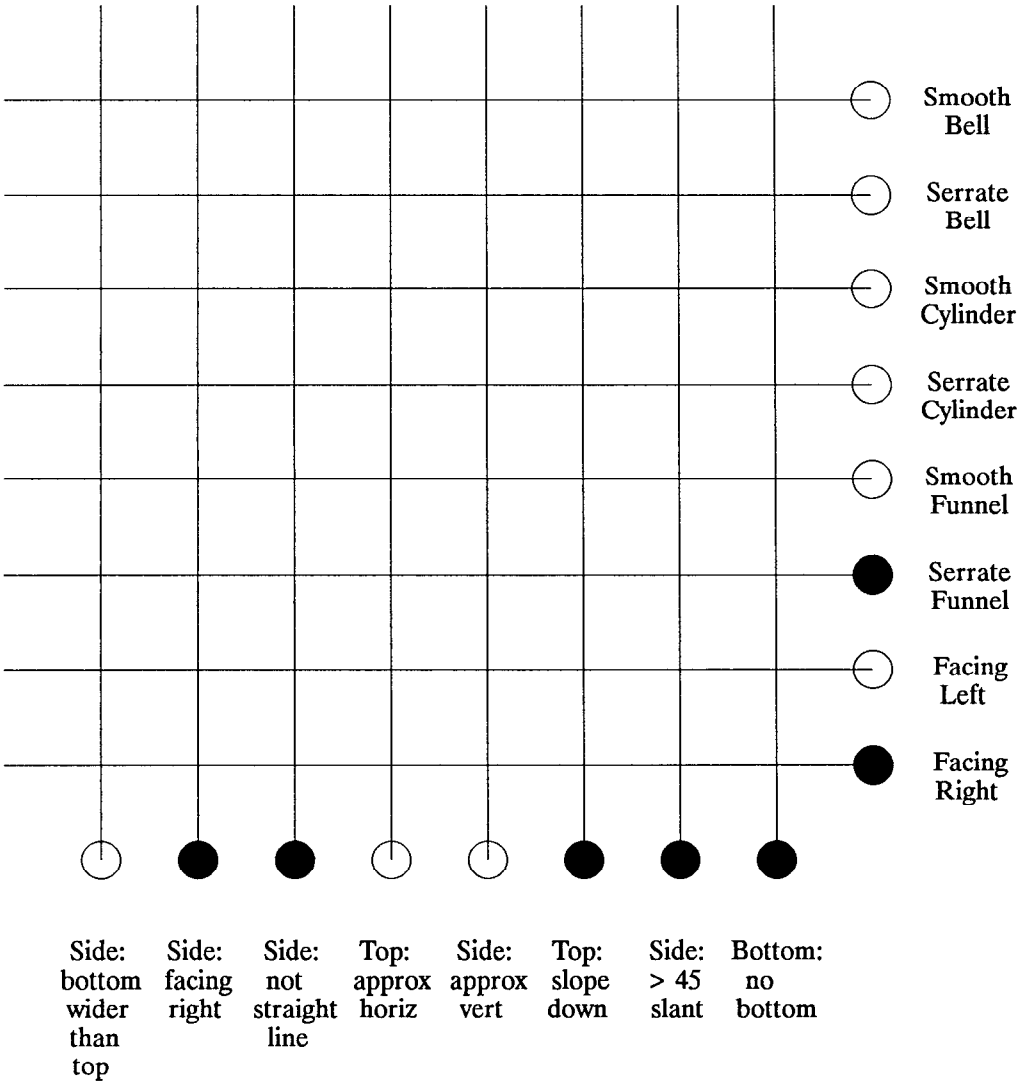


Figure E2.--Pattern associator model. The horizontal input units represent the features of the log shapes, and the vertical output units represent the shape identification. The solid units refer to a positive activation value for that unit, whereas the open units refer to a negative activation value.

LEARNING

Most neural network models use learning algorithms. In general, learning deals with finding the connection weights that will produce a desired response for the model. Usually, the learning

algorithms work from training examples of desired response to generate appropriate weights, w_{ij}. Each training example consists of the desired, or target, activations for a set of output units when the input units are fixed to specific activation values.

The process of training the network by repeatedly presenting the same input patterns and their corresponding target output patterns, causes the system to self-organize by generalizing and adapting until the correct responses are produced. In our development of VERSECORR, we tested two methods of learning: the Hebb rule, and the delta rule.

The Hebb Rule

Donald Hebb, in 1949, published the definition of a learning process that explained the modification of synapses between neurons. His definition of what is known as the "Hebb synapse" states that when two units are activated at the same time the strength of the connection between them should be increased.

The mathematical formulation of the Hebb Rule can be stated as follows (Rumelhart and McClelland, 1986, p. 53):

$$\Delta w_{ij} = \eta a_i o_j \,. \tag{1}$$

As stated earlier, the network is trained by repeatedly presenting input-output pattern pairs. So, for any learning event, k,

$$w_{ij} = \eta a_{jk} o_{ik} \,. \tag{2}$$

In this equation η refers to the value of the learning rate parameter, a_{jk} refers to the activation of input unit j with input pattern $\vec{a}_k$, and o_{ik} refers to the activation of output unit i in output pattern $\vec{o}_k$. Each weight is the product of the activation of the input unit and the activation of the output unit in learning event k.

Kohonen (Kohonen, 1988) provides an analysis of the activation of an output unit in a specific learning event; this keeps with the notation established above,

$$o_{ik} = \sum_{j=1}^{n} w_{ij} a_{jk} \,. \tag{3}$$

If we present a test input pattern, $\vec{a}_t$, to the network, then the resulting activation of output unit i when tested with input pattern $\vec{a}_t$ is

$$o_{it} = \sum_{j=1}^{n} w_{ij} a_{jt} \,. \tag{4}$$

Substituting the value of w_{ij} from equation (2) yields

$$o_{it} = \sum_{j=1}^{n} \eta a_{jk} o_{ik} a_{jk} \,. \tag{5}$$

Finally, because we are summing with respect to j in equation (5), we have

$$o_{it} = \eta o_{ik} \sum_{j=1}^{n} a_{jk} a_{jt} . \quad (6)$$

The sum of products of corresponding elements is the dot product; this represents the similarity between the two patterns $\vec{a}_k$ and $\vec{a}_t$.

Because the patterns consist of only +1 or -1 values, the dot product corresponds to the correlation between the two patterns, $\vec{a}_k$ and $\vec{a}_t$, if the dot product is normalized (that is, divided by its length to produce a unit vector). This method is standard in the application of Hebbian learning.

The normalized dot product has a value of 1 if the patterns are identical, a value of -1 if they are exactly opposite of each other, and a value of 0 if the elements of one pattern vector are completely independent of the elements of the other. When the elements of one pattern vector are completely independent of the elements of the other, the patterns are said to be orthogonal. And, when all the members of a set of patterns are orthogonal to each other, they form an orthogonal set.

In Hebbian learning, if the input patterns used in training the network do not form an orthogonal set, then it is not possible to avoid "cross-talk" between the response that is appropriate to one output pattern and the response that occurs to the others (Rumelhart and McClelland, 1986, p. 63).

The Delta Rule

The delta rule, also known as the Widrow-Hoff learning rule, allows us to overcome the orthogonality limitations of the Hebb rule. This is done by using the difference between the target activation and the obtained activation to drive learning. The goal is to adjust the strengths of the connections so that they will tend to reduce this difference measure.

The delta rule for a particular input-target pair $\vec{a}_p, \vec{t}_p$ can be stated as (Rumelhart and McClelland, 1986, p. 322):

$$\Delta w_{ij} = \eta(t_{ip} - o_{ip}) a_{jp} . \quad (7)$$

Thus, the weights that result from a group of learning pairs indexed by p can be written as:

$$w_{ij} = \eta \sum_{p} (t_{ip} - o_{ip}) a_{jp} . \quad (8)$$

If the input patterns used in training the network form an orthogonal set, then the delta rule produces the same results as the Hebb rule. If, however, the input patterns used in training the network do not form an orthogonal set, then (by equation 7) the change in the weights that occurs on a learning trial stores an association of the input pattern with the error pattern. That is, an increment is added to each weight that can be thought of as an association between the error for the output unit and the activation of the input unit.

For example, if we have an input pattern, $\vec{a}_p$, that is paired with a target pattern, $\vec{t}_p$, on the output produced by test pattern $\vec{a}_t$, then, by equation (7), the effect of the change in the weights due to a training session will change the output of some output unit i by an amount proportional to the error that occurs for that unit during the training session. This error, ε_i, is multipled by the dot product of the learned pattern with the first test pattern. This is illustrated in the following equation:

$$\Delta o_{it} = n\,\eta\varepsilon(a_p \cdot a_t)_n \tag{9}$$

where n equals the number of input units.

Thus, the change in the output pattern at test is proportional to the error vector times the normalized dot product of the input pattern that occurred during learning and the input pattern that occurred during test.

SYSTEM TRAINING and TESTING

By use of the delta rule to drive learning, it is possible to train VERSECORR to recognize the geophysical log patterns. The patterns (fig. E3) used in the training of VERSECORR do not form an orthogonal set; this makes it impossible to use Hebbian learning in the training of the network. Therefore, the network is trained using the delta rule to drive learning.

The input-output pattern pairs representing the log patterns (fig. E3) can be found in table E1. The input patterns are based on the presence or absence of a specific input feature, derived by the interface. The output patterns follow from the combination of the input features leading to the a specific output identification.

Table E1.--Training set of input-output pattern pairs

Pattern name	Input	Output
Smooth bell (left)	+ - - - - + - -	+ - - - - - + -
Smooth bell (right)	+ + - - - + - -	+ - - - - - - +
Serrate bell (left)	+ - + - - + - -	- + - - - - + -
Serrate bell (right)	+ + + - - + - -	- + - - - - - +
Smooth cylinder (left)	- - - + + - - -	- - + - - - - +
Smooth cylinder (right)	- + - + + - - -	- - + - - - - +
Serrate cylinder (left)	- - + + + - - -	- - - + - - + -
Serrate cylinder (right)	- + + + + - - -	- - - + - - - +
Smooth funnel (left)	- - - - - + + +	- - - - + - + -
Smooth funnel (right)	- + - - - + + +	- - - - + - - +
Serrate funnel (left)	- - + - - + + +	- - - - - + + -
Serrate funnel (right)	- + + - - + + +	- - - - - + - +

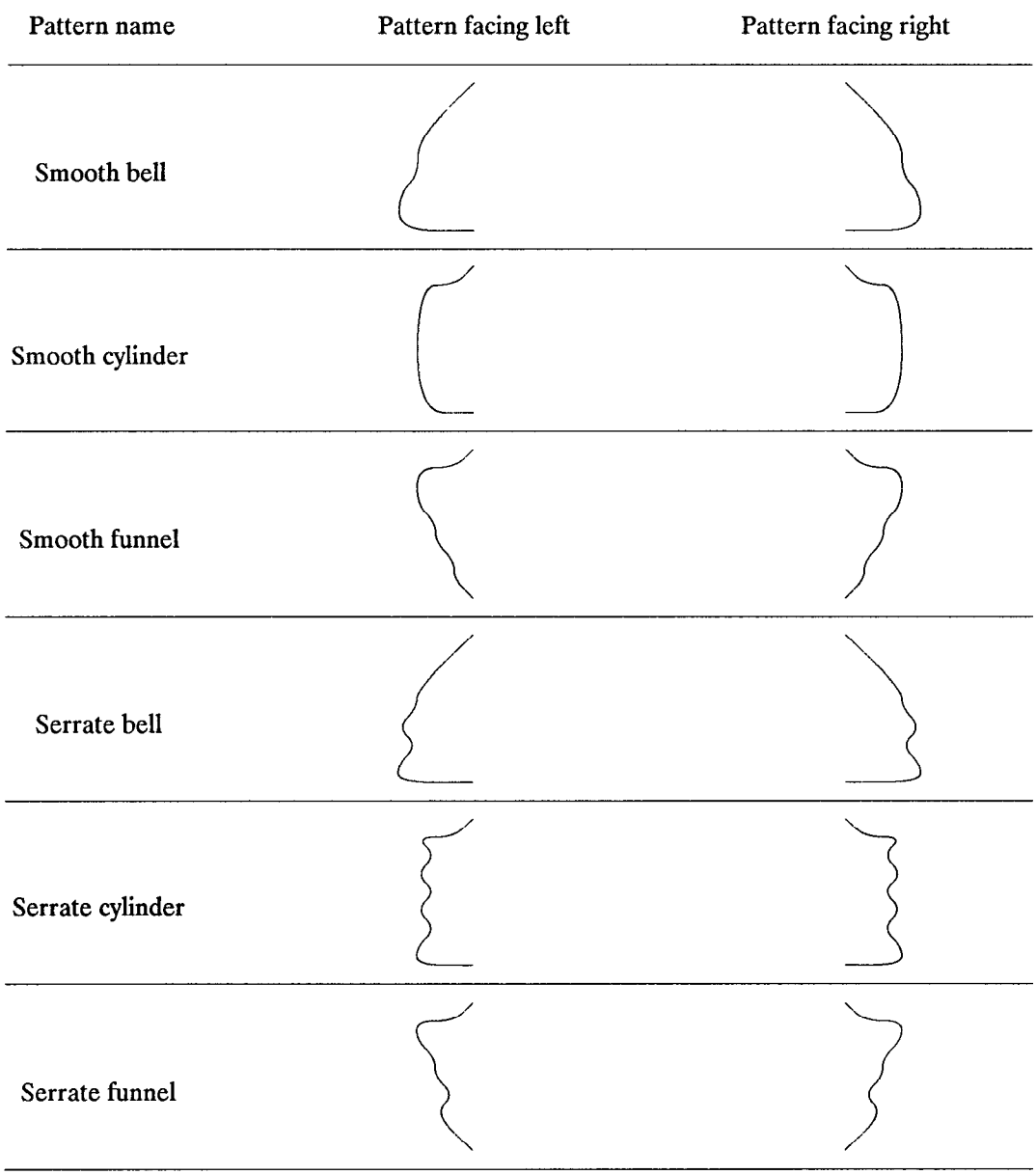


Figure E3.--Training set of log shapes

Upon completion of the network training using the training set shown in table E1, a test set of patterns were input. Figure E4 shows the test set of input patterns and the resulting output identifications.

Pattern	Identification	Pattern	Identification
A	Smooth cylinder Serrate cylinder Smooth funnel	K	Smooth bell Smooth cylinder Serrate cylinder Smooth funnel
B	All patterns	L	All patterns
C	No patterns	M	No patterns
D	No patterns	N	No patterns
E	No patterns	O	All patterns
F	Smooth cylinder Smooth funnel Serrate funnel	P	Smooth bell Smooth cylinder Serrate cylinder Smooth funnel Serrate funnel
G	Smooth cylinder	Q	Smooth cylinder Smooth funnel
H	No patterns	R	No patterns
I	All patterns	S	All patterns
J	Smooth cylinder Serrate cylinder Serrate funnel	T	Smooth cylinder Serrate cylinder Serrate funnel

Figure E4.--Test set of log shapes

RESULTS of TESTING

VERSECORR performed reasonably well in identifying correctly 15 of 20 patterns that were not part of the training set. There were, however, 5 cases (D, E, H, K, and N of fig. 4) in which VERSECORR made incorrect identifications. The system's inability to produce consistent output in these cases may have been caused by the use of a linear least-squares interface to produce the input patterns. It is possible that in some cases this method produces an input-output pattern similarity structure that is too different.

Our problems might also be solved by adding intermediate layers of hidden units that create their own internal representations of the input units. Hidden units are units whose inputs and outputs are within the neural network model. These units are not influenced by anything external to the model, such as an interface program. Moreover, for VERSECORR to recognize sequences of patterns, the addition of intermediate layers of hidden units will likely be required. This will also lead to the use of more robust learning algorithms.

CONCLUSIONS

VERSECORR is a pattern associator network that maps a set of input layer patterns directly to a set of output layer patterns. In many cases, patterns associator networks make reasonable generalizations and perform resonably on patterns that have never before been presented to the system. This is because pattern associator networks map similar input patterns to similar output patterns. The similarity of patterns is determined by their overlap. The overlap in pattern associator networks is determined by whatever produces the patterns outside the learning system.

The elementary neural computing techniques used by VERSECORR show considerable promise in pattern interpretation. As a result, continued development of these techniques is needed in future research into vertical sequence interpretation and correlation problems.

REFERENCES CITED

Kohonen, T., 1988, *Self-organization and associative memory:* New York, Springer-Verlag, 312 p.

Rumelhart, D. E., and McClelland, J. L., 1986, *Parallel distributed processing* (*v. 1 and 2*)*:* Cambridge, MA, MIT Press, 1158 p.

COMPUTER PROGRAMS FOR PROCESSING MODEL DATA AND RESULTS FOR STEADY-STATE AND TRANSIENT GROUND-WATER MODELS

By Carmen R. Baxter

U.S. Geological Survey

ABSTRACT

Several computer programs have been modified or developed by the U.S. Geological Survey to analyze the results of steady-state and transient-state data ground-water model simulations and to modify the input data for the models to develop the best possible fit between computed and observed conditions. Computer programs designed for modification of input data, such as pumpage and riverbed conductance is useful in evaluating the sensitivity of simulated aquifer properties and related parameters. Programs that redistribute data over given locations in a study area help to better address site-specific problems. Tables and hydrographs also are effective in examining time-variant output data at selected grid locations produced by transient simulation. Model output also can be used as input data for graphics computer programs for producing linear or spatial plots that may include grids enhanced with posting, contouring, or other data indicators.

INTRODUCTION

Performing model calibrations and sensitivity analyses on large amounts of data can prove to be cumbersome using text editors. A lack of consistency is a common occurrence in data modified directly by the modeler. Pre- and post-processing programs are required to analyze the results of steady-state and transient runs or to modify the input data to develop the best possible fit between observed and computed conditions. Several software techniques to interpret the results of model runs or to modify input data are currently employed in the Arkansas District of the U.S. Geological Survey. These techniques have been used to process data and to evaluate the hydrology in the East Arkansas Comprehensive Study (EARCS) project area.

Purpose and Scope

The purpose of this paper is to give examples of computer programs that have been used in the Arkansas District for the processing of model data that may be beneficial to other district personnel in need of pre- and post-processing methods on model data. The examples include overviews on global modification and data redistribution programs for reprocessing of model input. For post-processing of model output, overviews of programs using nongraphic- and graphic-oriented programs with their output are presented. Examples include tables and hydrographs produced for the purpose of calibrating the model program's input parameters as are sensitivity analyses and grid overlays of maps using the graphics packages DISSPLA and TELLAGRAF.

Background Information

The pre- and post-processing computer programs discussed in this paper were developed using Fortran 77 and the graphics packages DISSPLA and TELLAGRAF (Integrated Software Systems Corp., 1984a; 1984b). The modular three-dimensional finite-difference ground-water flow model program (McDonald and Harbaugh, 1988) was used to produce steady-state and transient-state model simulations to a ground-water flow system of a study area.

Model calibrations define aquifer parameters, boundary conditions, and other parameters in order with the aid of a model program, for known conditions to be simulated in the modeled system. The modeler may have to run the model program repeatedly, making reasonable changes to given parameters that he feels are less reliable to produce results that adequately match observed field data.

Sensitivity analyses which are performed after the final model calibration help the modeler to evaluate the reliability of the model. To make the analyses, the modeler first changes the values of selected input parameters; usually, a certain percentage change of the calibrated value is made for one parameter at a time.

The programs developed to perform model calibrations and sensitivity analyses were implemented on a ground-water flow system model of the East Arkansas Regional Comprehensive Study (EARCS) project area. The purpose of the EARCS was to digitally simulate the current conditions in the Mississippi Valley alluvial aquifer and to project the future consequences of increased water demands placed upon the aquifer. The study area is located in east-central Arkansas and encompasses a 13,400 mi^2 (square mile) area, including all or parts of 24 counties. Two grid-cell systems based on the Universal Transverse Mercator System (UTM) were developed. One grid has 86 rows and 52 columns with each grid node called a 3-mile cell that represents a 9 square mile area. The second grid overlaying the same area as the first has 258 rows and 156 columns with each grid node called a 1-mile cell that represents a 1 square mile area.

REPROCESSING OF MODEL INPUT DATA

Input parameters for the model program are processed continually to optimize the computed heads. Once a final model calibration is made, sensitivity analyses are performed to evaluate the reliability of the model and to check the sensitivity of selected location.

Global Modifications

The most common method of modification to input parameters is to increase or decrease all values within a given file by a specified percentage. The change in model-computed heads following such a change in an input parameter will give insight to the sensitivity of the model to that parameter at various locations in the study area. A program that was developed as a part of EARCS allows the user to enter a percentage to be applied to a given input file from

which a new file is created for model test runs. Another modification to the program permits different percentages of change in input parameters to be applied to each stress period.

The above program that applies an overall percentage to an input file was modified with a subroutine using the boundary array, a model input file, which allowed the user to enter a percentage for each river basin. The boundary array is a node-level map that allows the analyst to define the status of each grid node and to manipulate parameter input to the selected locations. This array contains integer codes defining the regions in the study area. Zeroes indicate inactive cells, negative values are constant heads, and positive values indicate river basins, rivers, and other geographic features important to the processing of the model. The boundary array is useful in indicating problem regions in a study area.

Data Redistribution

Data is redistributed by taking a value from a given cell location from one grid overlay and distributing the value into cell locations of another grid overlay covering the same spatial area. Two grids were used to overlay the same study area for the EARCS project: (1) a grid consisting of 86 rows and 52 columns with each cell being 3 miles on a side (9 mi^2), and (2) a grid consisting of 258 rows and 156 columns with each cell being 1 mile on a side (1 mi^2). The 3-mile grid gave an approximate regional look at the study area and provided the initial optimization data early in the project. Subsequently, a 1-mile grid was used to better address site-specific problems in the area. A data file containing river recharge values for active nodes in the 3-mile grid had been developed and a new file of river recharge for the 1-mile grid was required. A program was written that used the boundary array and the river-recharge file of the 3-mile grid and the boundary array of the 1-mile grid to produce the new values. The output river recharge to a 1-mile cell was based on the following conditions:

1. For a 3-mile basin node, the river recharge value is placed in all active 1-mile cells within 3-mile node.
2. For a 3-mile river node, its river recharge value is placed into every 1-mile river node equivalent to the 3-mile boundary code. The remaining 1-mile active cells retrieve their values from the first adjacent active cell checked that has a nonzero value. The adjacent cell may be part of the 3-mile node or an adjoining 3-mile node.
3. For a nonactive 3-mile node, any active 1-mile cell within the 3-mile node will retrieve its value from the first adjacent active cell checked with a nonzero value and the nonactive cells are set to zero. The adjacent cell may be part of the 3-mile node or an adjoining 3-mile node.

The resulting array for the 1-mile recharge was then written to a file to be used as input for the 1-mile model.

POST-PROCESSING OF MODEL OUTPUT

Model programs give the user a limited control over the amount of output and sometimes the form of the output that is generated. Depending on whether the modeler is calibrating the model or performing sensitivity analyses, he may wish to view the output in other forms that will best help him analyze the data. Some output forms would include tables, hydrographs, and grid.

Tables are commonly used in processing of model data to enable the user to view information from selected grid locations in an uncomplicated form. One or more sets of data may be displayed in a table for listing or comparative purposes. Most tables have a simple format consisting of a heading that describes the information to be listed and subheadings that indicate the value(s) and the respective row and column location.

Hydrographs and grids also are used in post-processing to analyze the results of a model run. Hydrographs may be produced with or without graphics programs to illustrate sensitivity analyses. Contouring, posting, and other forms of data indicators often are placed on graphically produced grids to give the modeler a visual aid for locating problem areas or indicating geographical features in the study area.

Tables and Hydrographs for Examining Data at Selected Grid Locations

Tables have been used to compare computed and observed water levels at selected locations in the study area. A tabling program was developed to compare the contents of a file containing selected observed water levels of a given stress period with the computed heads of the same stress period. A difference between the computed and observed heads then is calculated. The selected differences, computed heads, observed heads, and their corresponding locations are written on a single table. Further modification to the program allows a table to be produced from selected nodes containing the computed and observed water levels and their difference for all stress periods.

This tabling program was then expanded to include hydrographs. The purpose of the hydrographs is to provide a pictorial representation of the tabled data and to allow the modeler to examine the fit between the two unconnected sets of points marking the integer form of the observed and computed water levels. The symbols used to mark the water levels at a given stress period are as follows:

C -- computed water levels,
O -- observed water levels, and
* -- overlapping of computed and observed water levels.

Note that the "overlapping" of computed and observed water levels is the integer form of the values, not the decimal values. For example, if the observed value for a given stress period is 185.00 and its corresponding computed value is 185.75, then both values would be plotted to the hydrograph as 185.

The coordinates of the graph are based on the stress period and the water level in feet above sea level. The range of the water-level axis is further reduced by determining the minimum and the maximum of the levels. The minimum is lowered 10 feet and the maximum is set 110 feet above the new minimum before the two values are used as the beginning and end labels of the axis. The program also constructs the table and hydrograph for a selected location to be printed on one page if the user spools to a line printer using the spool option -FTN (fig. E3).

Input Data to Graphics Software

Model output is used regularly to produce linear plots for sensitivity analyses. The most common linear plot is the hydrograph which portrays the data results in head files produced from steady-state and transient model runs. The data compared to the results may be observed data collected for a given time period or other model output that may vary from the final calibrated results as obtained by modifying pumpage, riverbed conductance, or another input parameter. Graphics packages available such as DISSPLA and TELLAGRAF are useful in illustrating these comparisons (Integrated Software Systems Corp., 1984).

Hydrographs

The tabling program that produced line-printer hydrographs for calibration purposes was modified further to produce input data for TELLAGRAF. The user enters the run number of the head file to be processed and a separate file is created for each selected location. Each file contains a list of observed water levels and a list of computed water levels for all stress periods for a given location. After execution of the program, these files are loaded separately into a TELLAGRAF data file and saved under different file names. Each new TELLAGRAF data file is entered into the TELLAGRAF program. Enhancements to the data file were made from within TELLAGRAF to produce a hydrograph at a given row and column location. This hydrograph shows calibrated water levels as compared to observed water levels for seven stress periods (fig. E4).

Other programs were written using DISSPLA routines for sensitivity analyses. One program allows the user to indicate whether a sensitivity analysis is to be performed on files produced by a steady-state or transient model run. If a transient condition is chosen, the user enters the row and column to be analyzed. The hydrograph produced has an axis of water levels and an axis of stress periods. If a steady-state condition is selected, the user is prompted for a row and range of columns to be analyzed. The hydrograph created has an axis of water levels and an axis consisting of the range of columns entered by the user. The user then enters the final calibrated head file and up to four additional test head files that can be compared to the contents of the final head. The user also can indicate whether a title is to be printed; an explanation is to be printed; and the curves drawn are to be of different colors or all black.

Two other programs performed sensitivity analyses on steady-state models only. The programs are identical except one performs an analysis on a given row and the other program analyzes a given column. The name of the final

STP	LOC	OBSERVED	COMPUTED	DIFFERENCE
1	1537	187.00	186.52	0.48
2	1537	185.00	185.96	-0.96
3	1537	181.00	183.71	-2.71
4	1537	177.00	181.47	-4.47
5	1537	171.00	173.88	-2.88
6	1537	164.00	162.56	1.44
7	1537	161.00	150.46	10.54

EXPLANATION

O = OBSERVED VALUE
C = COMPUTED VALUE
* = OVERLAPPING OF OBSERVED AND COMPUTED VALUES

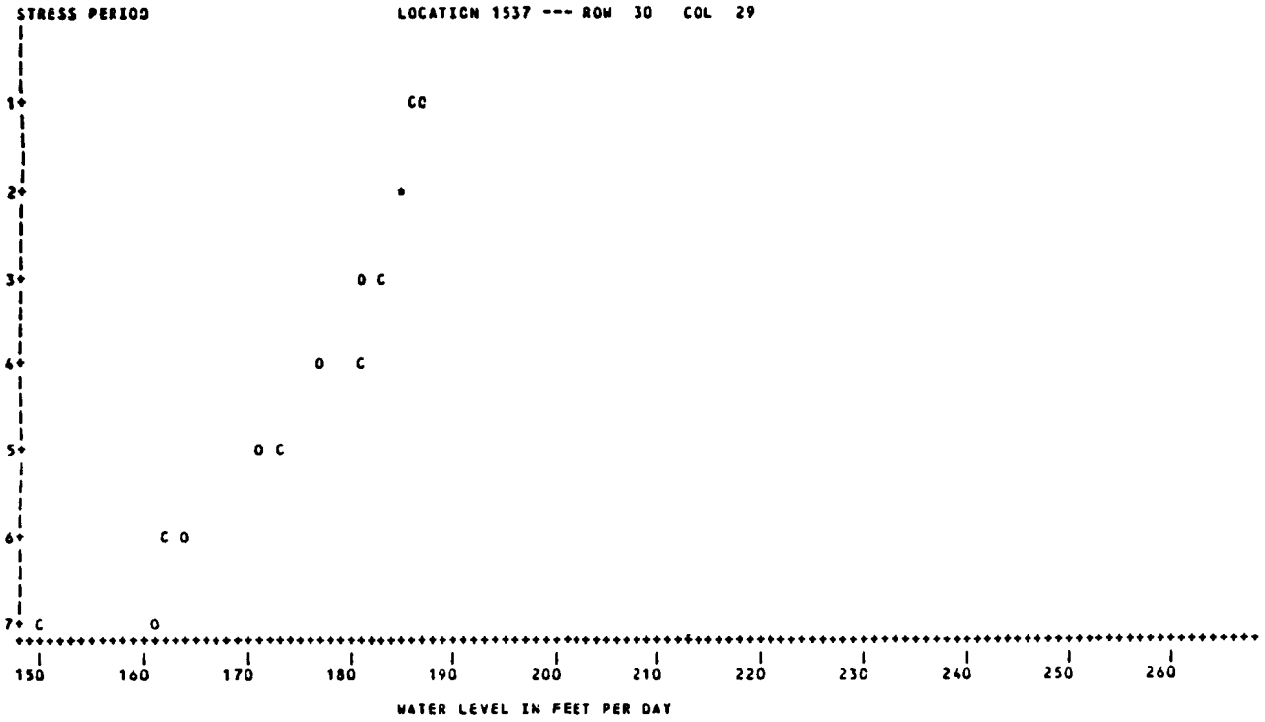


Figure E3. -- Table and hydrograph for transient calibration simulations.

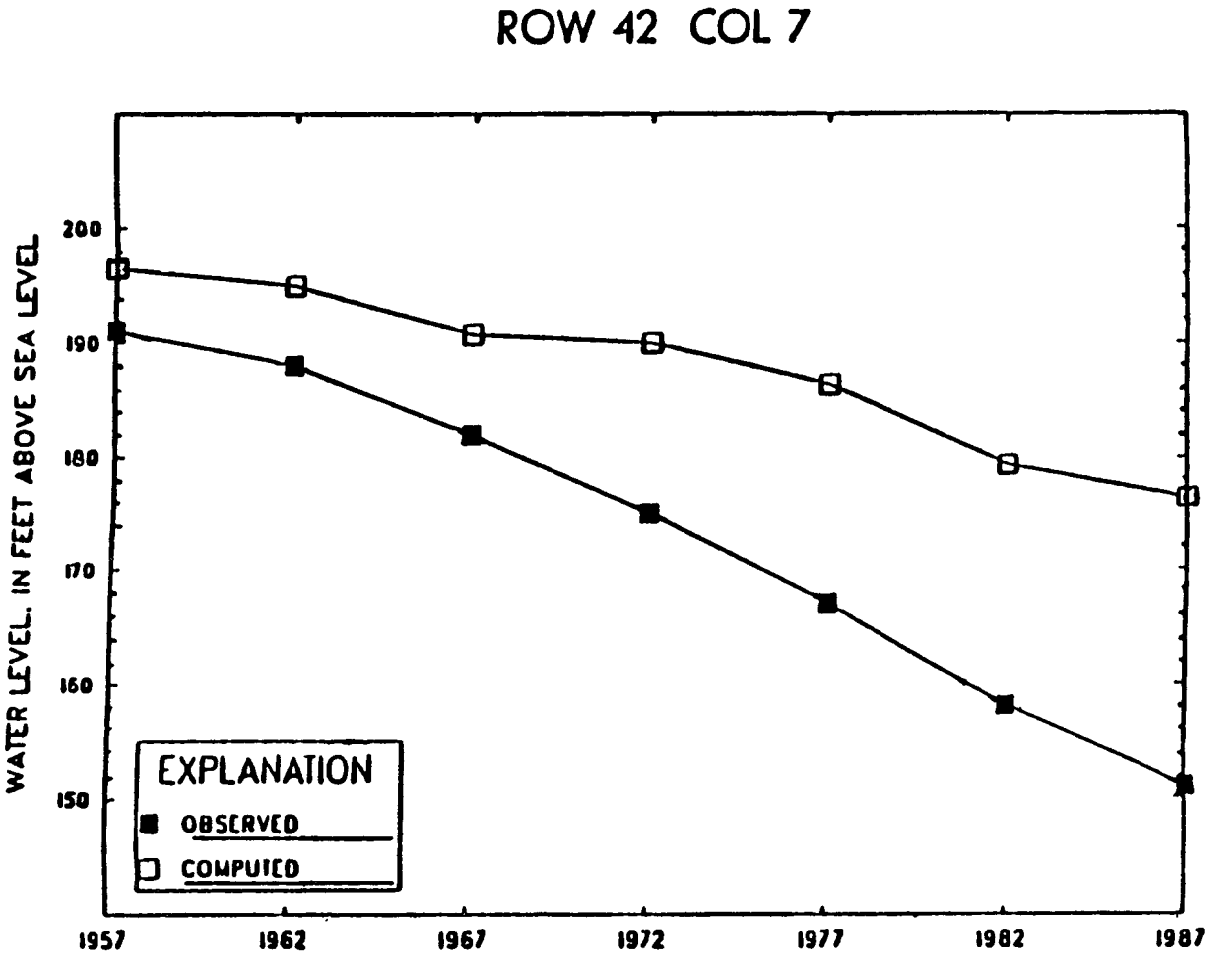


Figure E4. -- Hydrograph showing computed water levels produced from a transient model simulation as compared to observed water levels for seven stress periods.

calibrated head was coded in the programs. The user has to enter only four test files with descriptions for an explanation. The user then enters the row (or column) to analyze and the device by which the plot is to be drawn. The values marked on a plot (with a difference limit of plus or minus 40 feet) are the nodal value differences between the calibrated head and the test file. If the difference is greater than 40 feet, then the difference plotted is 40 feet. If the difference is less than -40 feet, then the difference is -40 feet. Figure E5 shows an example of such a sensitivity analysis for a given row.

Grids

A grid is a group of intersecting lines that form squares and is used as a reference for locating points in a study area. Data can be placed in the grid cells to give the modeler a visual aid for locating problem areas or indicating geographical features in the study area.

Posting

To understand the distribution of model data, one may post or print numerical values to a grid that will overlay a map of the study area. Posting data values will allow the checking of input data errors by the modeler, show the distribution of the output over various physical features in the study area, and indicate the problem regions.

A program incorporating DISSPLA routines to produce a grid overlay with 86 rows and 52 columns was developed specifically for the EARCS area. The grid was created at a 1:500,000 scale to be placed over a Lambert Conformal Conic map projection. Four latitude and longitude reference marks were drawn on the grid and a region of nonactive nodes at the bottom right was blanked for the grid title, latitude and longitude of the origin, and other information. Additional data can be added manually by the modeler or placed on the grid by a cartographer. Such data may include county codes and boundaries or the codes from the boundary array.

To permit the modeler to examine model input and output on the grid, the program was modified to query the user for a descriptive message about the data and an input file containing an array of values to be posted. An example of a posted grid is shown in figure E6. The values posted are saturated thicknesses produced from a transient run. A projected pumpage file divided into six stress periods (1990 to 2040) was used as pumpage input to the ground-water flow model and the resulting heads were manipulated to produce the saturated thickness for each cell in a given stress period.

Another available program produces grids at various scales. This program is designed specifically to be modified internally by the user to meet grid requirements. Upon execution of the run file, the user enters his boundary array file and indicates whether or not to post a data file. Other prompts include terminal type, paper size, and plot title. Therefore, the user can produce a base grid with only boundary array indicators or a grid with posted values.

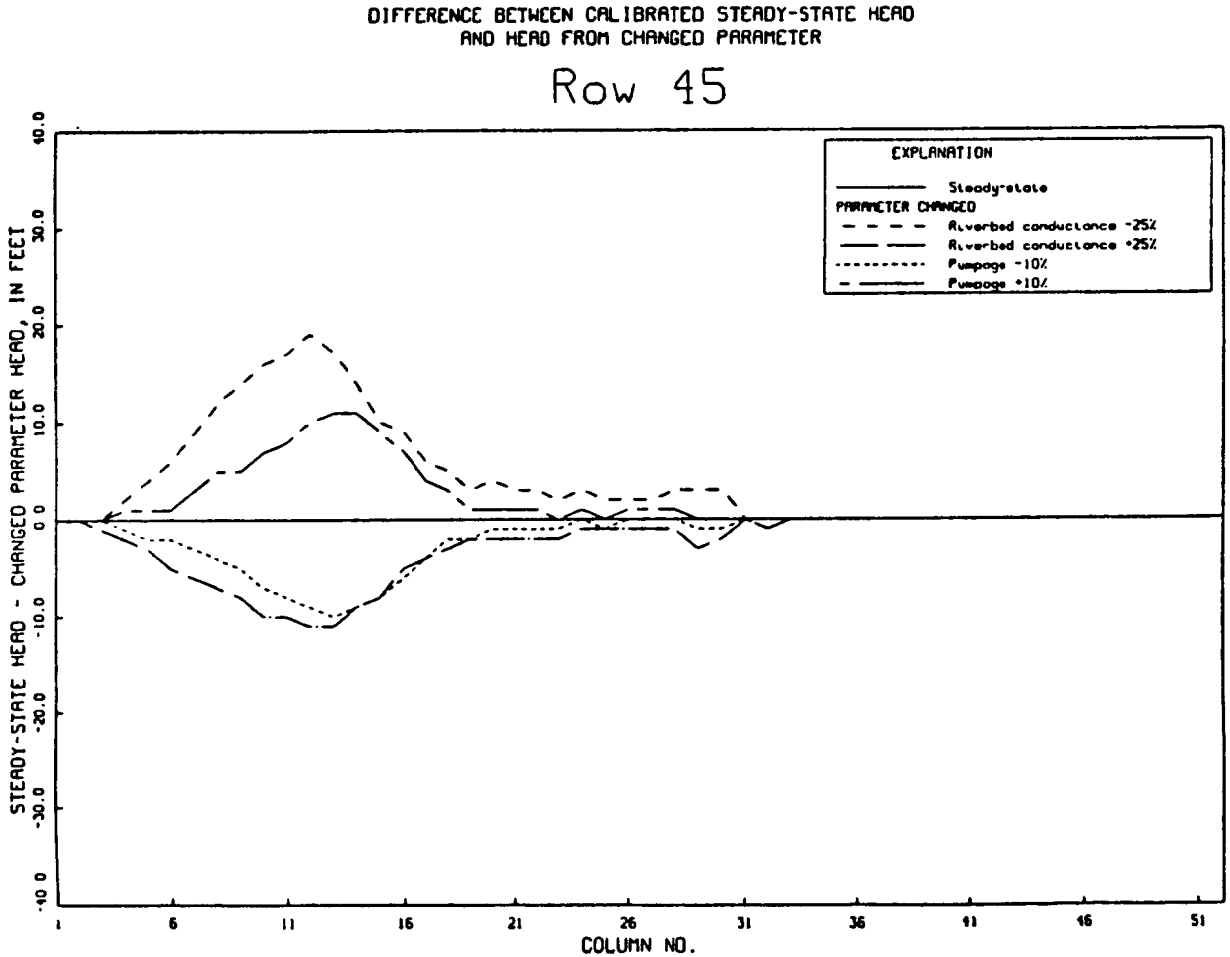


Figure E5. -- Hydrograph showing a sensitivity analysis of the difference between calibrated steady-state head and heads produced by changing other parameters.

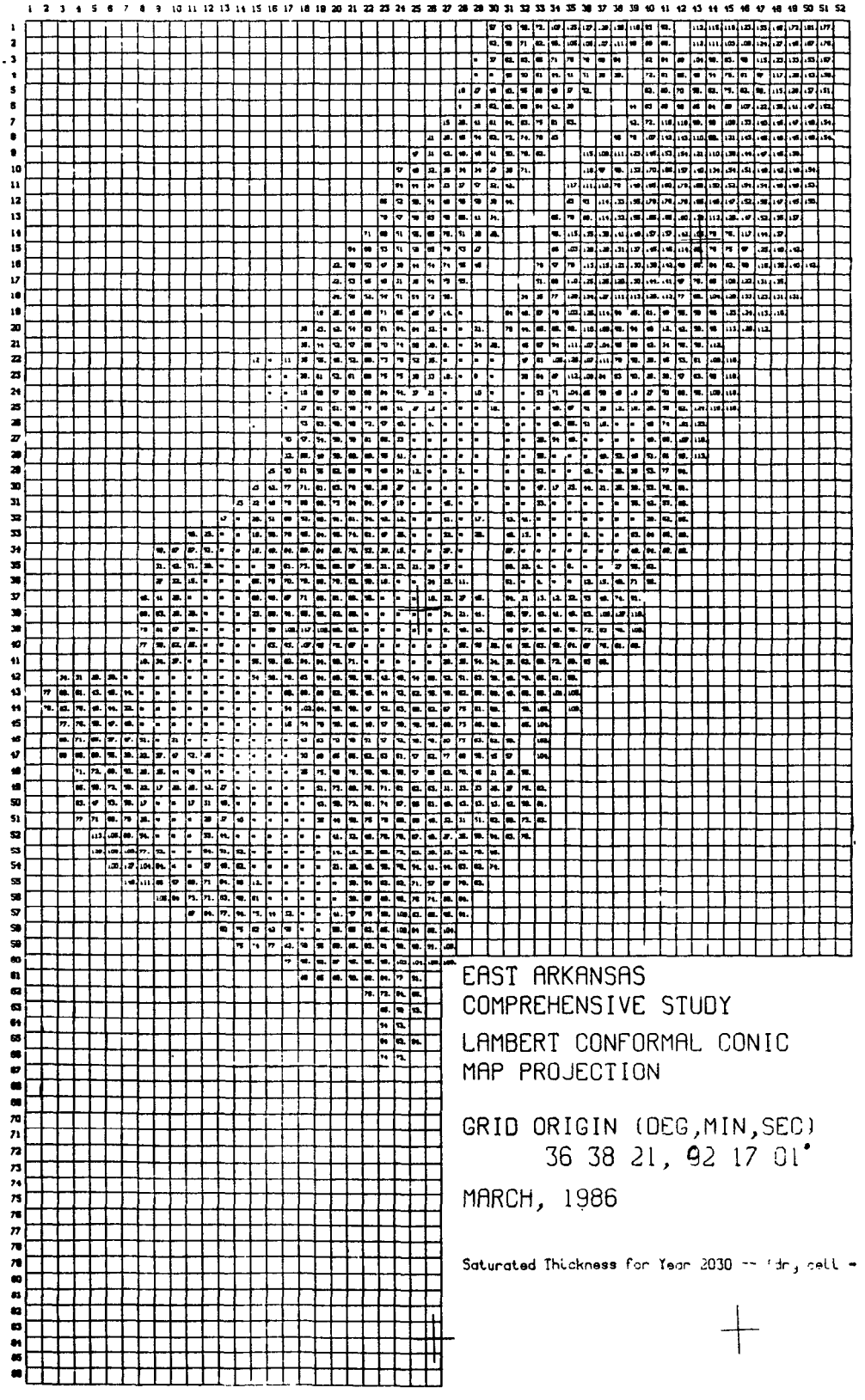


Figure E6. -- Posted values of saturated thickness projected for the year 2030.

Contouring

Contouring can be an additional feature to grids. Contour lines connect points of equal elevation and show the configuration of the surface. The posting program which produced grids at various scales, was modified with more DISSPLA routines for contouring data sets. DISSPLA generates contour lines by linear interpolation of adjacent grid cells in both the X and Y directions. With this enhancement, the modeler now has the option to post and (or) contour.

A contouring example is the potentiometric surface, which after the plotting and contouring of water-level altitudes, is the hydraulic head of an aquifer. The contour lines on the map indicate the direction of the ground-water flow that is perpendicular to the contours through the aquifer. A grid with a potentiometric map of the 1982 water levels in eastern Arkansas is shown in figure E7.

Other Data Indicators

Representative characters can be used to emphasize specific data values, ranges, or other characteristics on a grid. Additional changes to the posting routine allows certain ranges of data to be marked on the grid. The projected saturated thickness of 20 feet or less for the year 2030 is indicated with an "X" on the grid, thus allowing the modeler to note critical regions in the study area (fig. E8). One may note that this routine already emphasizes various codes from the boundary array. Asterisks (*) were inactive nodes, R indicated river cells, and C represented nodes in a hydrologic barrier called Crowleys Ridge.

SUMMARY

Several software techniques are presented that modified input or analyzed results of the McDonald-Harbaugh ground-water flow model. Two examples of reprocessing model input are given: global modification and data redistribution. For the post-processing of model output, nongraphic and graphic techniques are given. Model output for calibration purposes is presented in tables and line plotter hydrographs. Graphic routines execute sensitivity analyses and produce grid overlays with posting, contouring, and other data indicators.

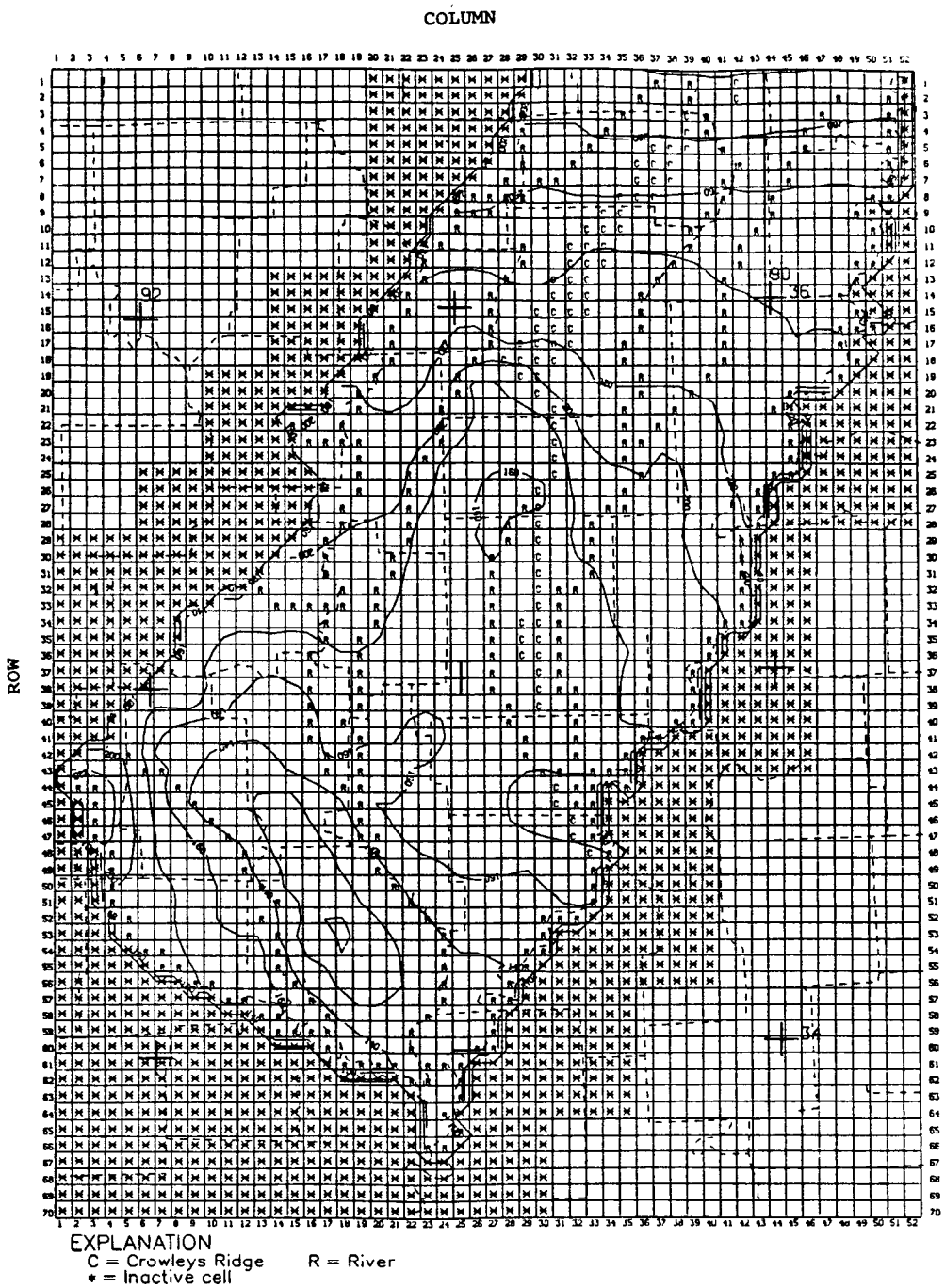


Figure E7. -- Potentiometric map of the 1982 water levels in eastern Arkansas.

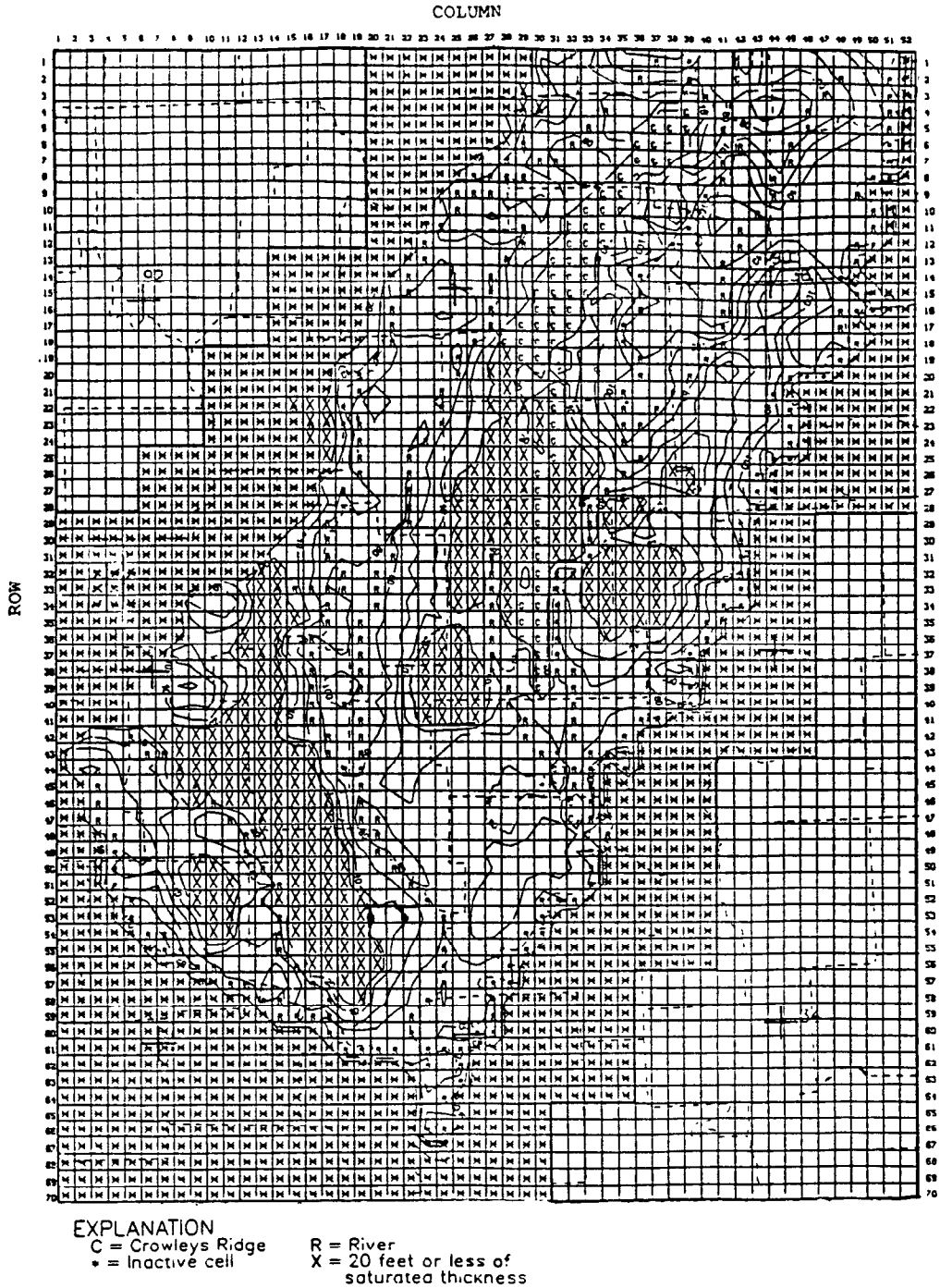


Figure E8. -- An example of data indicators by the marking of cells with 20 feet or less of saturated thickness.

SELECTED REFERENCES

Integrated Software Systems Corporation, 1984a, DISSPLA User's Manual, version 10.00.

----- 1984b, TELLAGRAF User's Manual, version 6.00.

Landy, Alice, 1982, CPL User's Guide, Revision 19.0, Prime Computer, Inc.

Lohman, S.W., Bennet, R.R., Brown, R.H., Cooper, H.H., Jr., Drescher, W.J., Ferris, J.G., Johnson, A.I., McGuiness, C.L., Piper, A.M., Rorabaugh, M.I., Stallman, R.W., and Theis, C.V., 1972, Definitions of selected ground-water terms -- revisions and conceptual refinements: U.S. Geological Survey Water-Supply Paper 1988, 21 p.

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, Chapter A1, 548 p.

A COMPUTER METHOD FOR ESTIMATING GROUND-WATER CONTRIBUTION TO STREAMFLOW USING HYDROGRAPH-SEPARATION TECHNIQUES

By Ronald A. Sloto
U.S. Geological Survey

ABSTRACT

A computer program used to estimate the ground-water component of streamflow (base flow) using hydrograph-separation techniques provides a rapid method of estimating base flow and eliminates subjectivity inherent in manual graphical methods. Three hydrograph-separation techniques are used: fixed interval, sliding interval, and local minimum. Input to the program is daily mean stream discharge in conventional daily values card-image format from U.S. Geological Survey data bases. The program provides printed statistical output, plot files, and base flow in card-image format. Printed statistical output includes monthly and annual summary information about the hydrograph separations, a summary of annual base flow and median base flow for period of record, frequency distribution of base flow, and percentage of ground-water contribution to streamflow. Plot files produce streamflow and base-flow hydrographs and frequency-distribution plots of annual base flows for period of record. Base flow estimated by computer techniques is comparable with base flow estimated by manual base-flow-recession and curve-fitting methods.

INTRODUCTION

Hydrograph separation divides streamflow into its two component parts: overland runoff and ground-water discharge (base flow). Hydrograph separation is an important technique used in a variety of water-resource studies. Hydrograph-separation techniques are commonly used to estimate ground-water contribution to streamflow and to define ground-water/surface-water relations. Hydrograph-separation techniques are also used to calculate hydrologic budgets and recharge rates.

The separation of a hydrograph into overland runoff and ground-water components is difficult and inexact, and the results are subject to different interpretations. Manual and computer techniques estimate base flow; no technique provides exact volumes of runoff and ground-water discharge.

Hydrograph separation generally is performed manually by graphical methods. Two commonly used manual methods are fitting a curve under a streamflow hydrograph (Linsley and others, 1982, p. 210) and base-flow-recession methods (Olmsted and Hely, 1962; Riggs, 1963; Rorabaugh, 1963). Quick estimates of base flow are sometimes obtained by sketching a curve below the streamflow hydrograph. Different hydrologists using the same manual hydrograph-separation method usually produce different estimates of base flow. The use of a computer program removes subjectivity inherent in manual methods. It substantially reduces the time required for hydrograph separation. For separation of a hydrograph for a long-term streamflow

station with 90 years of record, manual methods take weeks, and computer methods take seconds.

The purpose of this paper is to briefly describe the hydrograph-separation program and the savings in time and elimination of subjectivity attained by its use. A comparison of base flow estimated by manual and computer techniques is presented.

DESCRIPTION OF COMPUTER PROGRAM

The hydrograph-separation program uses the three techniques of Pettyjohn and Henning (1979) to separate the ground-water and runoff components of a streamflow hydrograph: fixed interval, sliding interval, and local minimum. All three techniques are run simultaneously so that differences in base flows can be compared.

Input Requirements

Input to the program is in a conventional format from the U.S. Geological Survey National Water Data Storage and Retrieval System (WATSTORE) and Automated Data Processing System (ADAPS) data bases. Input is daily mean stream discharge in daily values card-image format. The program will accept data in water (October 1 to September 30) or calendar (January 1 to December 31) years. The program reads as little as 1 year of record or as much as the period of record for a station. The program accounts for gaps in the period of record; however, input must be complete water or calendar years of data.

Program Output

The program produces printed statistical output, plot files, and daily values for base flow in card-image format.

The printed output provides information on the quantity of base flow and the percentage of streamflow as base flow. This information is used in hydrologic studies to calculate water budgets and characterize ground-water/surface-water relations. The user selects short or long format printed output. The short format prints program warning messages and a summary of annual base flow and base-flow frequency by water and calendar year. The annual summary lists annual base flow and median base flow in inches, cubic feet per second, and million gallons per day per square mile and gives the percentage of streamflow as base flow. In addition to warning messages and an annual summary, the long format also prints monthly values of mean and total discharge, mean and total base flow, and percentage of streamflow as base flow for each year of record. Table E4 is part of the printed output that summarizes estimated annual base flow for French Creek near Phoenixville, Pennsylvania for the period of record.

Graphical output is used to compare the three separation techniques and to present base-flow frequency data. Graphical output consists of TELAGRAF plot files. Annual water year or calendar year streamflow and base-flow hydrographs can be plotted for the three hydrograph separation methods (fig. E14). The base-flow frequency distribution for the period of record for water or calendar years can be plotted for one to three separation techniques (fig. E15).

COMPARISON OF COMPUTER-SEPARATION TECHNIQUES

Hydrographs were separated into base-flow and runoff components for 1961-85 water years

Table E4.--Base flow for French Creek near Phoenixville, Pennsylvania, 1969-85

[In., inches; ft^3/s, cubic feet per second; $(Mgal/d)/mi^2$, million gallons per day per square mile; percent, percent of streamflow as base flow]

WATER YEAR BASE FLOW SUMMARY FOR FRENCH CREEK NEAR PHOENIXVILLE, PA

Water	Fixed interval method				Sliding interval method				Local minimum method			
year	In.	ft^3/s	$(Mgal/d)/mi^2$	Percent	In.	ft^3/s	$(Mgal/d)/mi^2$	Percent	In.	ft^3/s	$(Mgal/d)/mi^2$	Percent
1969	6.86	29.86	0.327	68.69	6.85	29.82	0.326	68.59	6.61	28.78	0.315	66.20
1970	10.40	45.29	0.495	68.68	10.34	45.04	0.493	68.30	9.92	43.18	0.472	65.48
1971	12.65	55.05	0.602	55.79	12.91	56.22	0.615	56.97	11.99	52.20	0.571	52.89
1972	18.72	81.51	0.891	63.94	18.50	80.55	0.881	63.19	18.08	78.73	0.861	61.76
1973	17.97	78.25	0.856	61.30	18.25	79.47	0.869	62.25	17.98	78.29	0.856	61.33
1974	13.24	57.64	0.630	65.75	13.30	57.91	0.633	66.06	13.18	57.37	0.627	65.44
1975	15.07	65.62	0.718	62.54	14.79	64.38	0.704	61.35	14.34	62.44	0.683	59.51
1976	12.36	53.82	0.589	68.03	12.46	54.24	0.593	68.55	11.99	52.19	0.571	65.97
1977	9.74	42.39	0.464	61.98	9.88	43.00	0.470	62.87	9.40	40.92	0.447	59.82
1978	19.27	83.92	0.918	60.69	19.05	82.95	0.907	59.99	18.64	81.15	0.887	58.69
1979	16.71	72.74	0.795	56.26	16.58	72.20	0.790	55.84	15.93	69.34	0.758	53.63
1980	14.97	65.19	0.713	73.72	14.95	65.10	0.712	73.62	14.57	63.42	0.693	71.71
1981	5.42	23.62	0.258	65.33	5.47	23.81	0.260	65.87	5.29	23.04	0.252	63.73
1982	10.32	44.92	0.491	59.80	10.29	44.79	0.490	59.63	10.05	43.76	0.478	58.25
1983	14.56	63.37	0.693	68.24	14.51	63.16	0.691	68.01	13.89	60.47	0.661	65.12
1984	20.45	89.04	0.974	57.39	20.54	89.41	0.978	57.63	19.54	85.07	0.930	54.84
1985	6.92	30.12	0.329	61.07	6.96	30.30	0.331	61.43	6.66	29.01	0.317	58.82
MEDIAN												
	13.24	57.64	0.630	62.54	13.30	57.91	0.633	62.87	13.18	57.37	0.627	61.33

CALENDAR YEAR BASE FLOW SUMMARY FOR FRENCH CREEK NEAR PHOENIXVILLE, PA

Water	Fixed interval method				Sliding interval method				Local minimum method			
year	In.	ft^3/s	$(Mgal/d)/mi^2$	Percent	In.	ft^3/s	$(Mgal/d)/mi^2$	Percent	In.	ft^3/s	$(Mgal/d)/mi^2$	Percent
1969	6.40	27.88	0.305	67.40	6.43	28.01	0.306	67.71	6.18	26.88	0.294	64.99
1970	11.15	48.55	0.531	66.07	11.17	48.64	0.532	66.19	11.00	47.90	0.524	65.19
1971	14.71	64.03	0.700	59.10	14.81	64.47	0.705	59.51	13.65	59.44	0.650	54.86
1972	18.55	80.76	0.883	59.95	18.54	80.71	0.883	59.91	17.93	78.04	0.853	57.93
1973	16.23	70.65	0.773	60.60	16.41	71.43	0.781	61.27	16.37	71.26	0.779	61.12
1974	12.50	54.44	0.595	70.03	12.54	54.59	0.597	70.22	12.38	53.91	0.589	69.34
1975	16.92	73.66	0.805	64.93	16.58	72.18	0.789	63.63	16.11	70.14	0.767	61.83
1976	11.44	49.79	0.544	67.07	11.53	50.19	0.549	67.60	11.08	48.25	0.528	64.99
1977	10.72	46.68	0.511	56.17	11.01	47.93	0.524	57.67	10.80	47.01	0.514	56.56
1978	18.56	80.82	0.884	63.98	18.22	79.34	0.868	62.81	17.58	76.56	0.837	60.61
1979	18.55	80.78	0.883	58.85	18.42	80.21	0.877	58.44	17.59	76.59	0.837	55.80
1980	11.43	49.75	0.544	73.42	11.41	49.67	0.543	73.31	11.26	49.03	0.536	72.37
1981	5.75	25.02	0.274	66.01	5.80	25.24	0.276	66.57	5.53	24.07	0.263	63.50
1982	10.64	46.34	0.507	59.79	10.61	46.18	0.505	59.58	10.33	44.96	0.492	58.02
1983	16.21	70.58	0.772	60.16	16.24	70.71	0.773	60.27	15.48	67.38	0.737	57.43
1984	19.08	83.05	0.908	62.90	19.10	83.14	0.909	62.97	18.32	79.78	0.872	60.42
MEDIAN												
	13.61	59.24	0.648	63.44	13.67	59.53	0.651	62.89	13.02	56.67	0.620	60.86

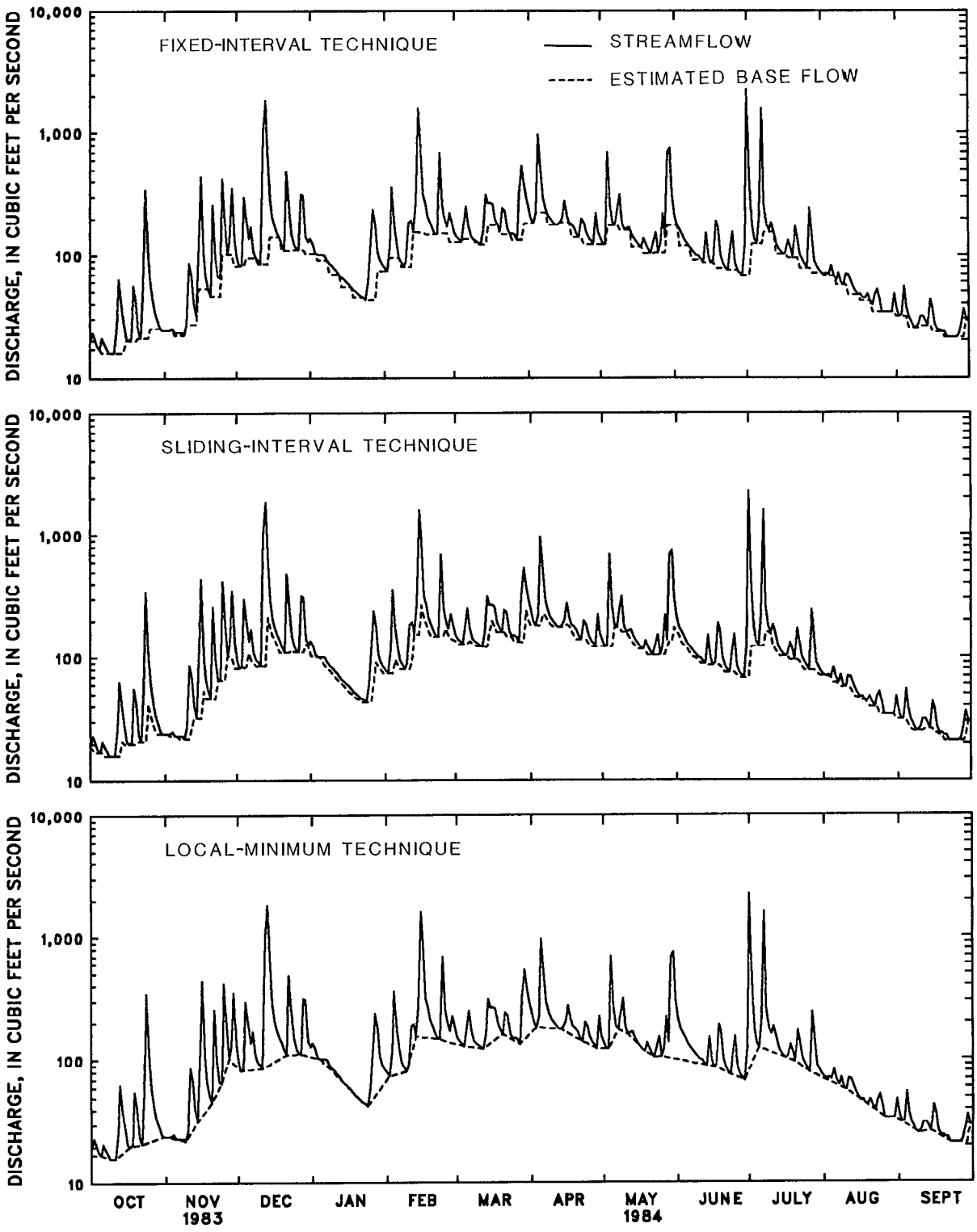


Figure E14.--Daily mean streamflow and base flow for French Creek near Phoenixville, Pennsylvania, 1984 water year.

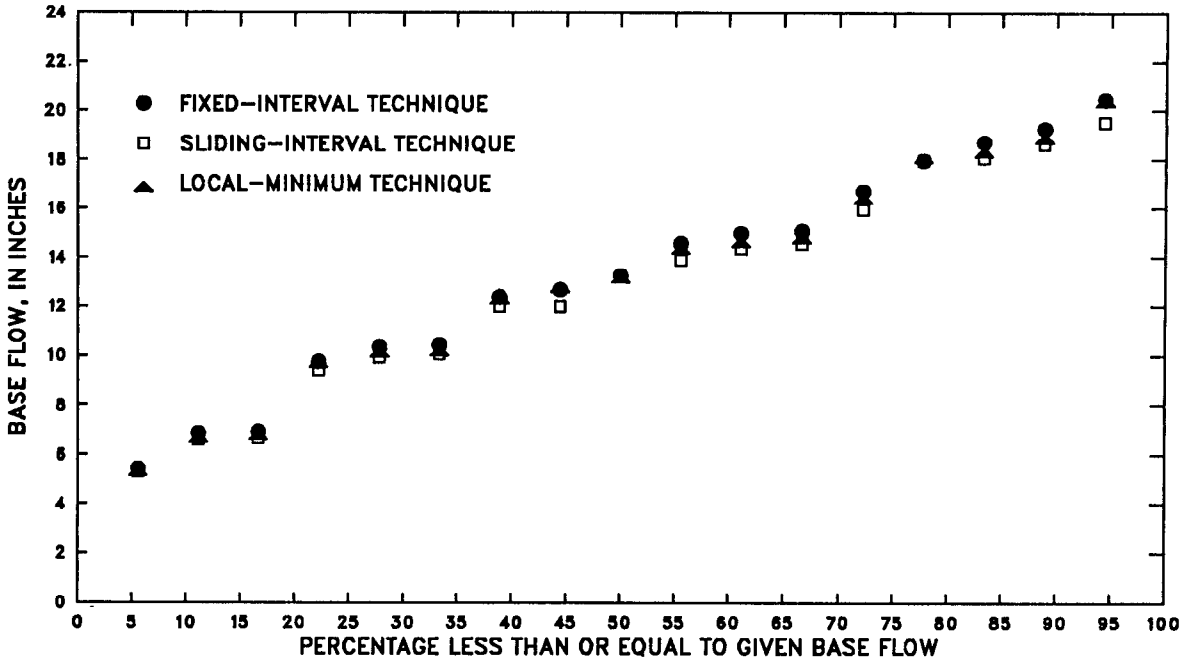


Figure E15.--Frequency distribution of base flow for French Creek near Phoenixville, Pennsylvania, 1969-85.

for three streamflow stations in southeastern Pennsylvania: West Branch Brandywine Creek near Honey Brook (drainage area of 18.7 mi^2), Little Lehigh Creek near Allentown (drainage area of 80.8 mi^2), and Neshaminy Creek near Langhorne (drainage area of 210 mi^2). Base flows are compared in table E5. The sliding-interval technique gave the highest median base flow for two of the three stations, and the fixed-interval technique gave the highest median base flow for one station. The local-minimum technique gave the lowest median base flow for all three stations.

COMPARISON OF COMPUTER- AND MANUAL-SEPARATION TECHNIQUES

Computer hydrograph-separation techniques were compared with those performed manually by base-flow recession and graphical methods for streams in southeastern Pennsylvania. Base flows estimated with computer techniques closely agree with those estimated by manual methods.

A study by Olmsted and Hely (1962) examined the relation between ground water and surface water in the 287 mi^2 Brandywine Creek basin in southeastern Pennsylvania. Olmsted and Hely (1962, p. 4-9) used base-flow recession curves to estimate ground-water contribution to streamflow for 1928-31 and 1952-53. Base flow estimated by Olmsted and Hely (1962) and by computer techiques (fixed interval, sliding interval, and local minimum) is given in table E6. All three computer techniques produced base flows comparable to the manual base-flow recession method. The average base flow for the 6-year period estimated by the sliding-interval technique was 1 percent higher, the estimate by the fixed-interval technique was less than 1 percent higher, and the estimate by the the local-mimimum technique was less than 1 percent lower than the base flow estimated by the manual base-flow recession method. Base flow estimated by computer techniques for individual years ranged from 5 percent lower to 5 percent higher than base flow estimated by the manual base-flow recession method. Base-flow hydrographs for 1952-53 published by Olmsted and Hely (1962, p. 5) are shown for comparison with those estimated by the local-minimum technique in figure E16. The base-flow hydrographs are very similar.

McGreevy and Sloto (1977, p. 38-39) used a graphical hydrograph-separation method to estimate basin ground-water yield in Chester County, Pennsylvania for a dry (1966), near-average (1968), and wet (1973) water year. The average base flow estimated at four streamflow stations by the fixed-interval technique was 6 percent higher, the estimate by the sliding-interval technique was 7 percent higher, and the estimate by the local-minimum technique was 4 percent higher than the base flow estimated by the manual graphical method. Base flow estimated by computer techniques for individual years ranged from 7 percent lower to 19 percent higher than base flow estimated by the manual graphical method.

Biesecker and others (1968, p. 27-29) determined the contribution of ground water to total streamflow for seven stations in the Schuylkill River basin in southeastern Pennsylvania and for Brandywine Creek at Chadds Ford, Pennsylvania for the 1952 and 1964 water years. They used a combination of base-flow recession and curve-fitting methods. The average base flow estimated at the seven streamflow stations by the fixed-interval and sliding-interval techniques was 9 percent lower and the estimate by the local-minimum technique was 14 percent lower than the base flow estimated by the manual method. Base flow estimated by computer techniques for individual years ranged from 7 percent higher to 31 percent lower than base flow estimated by the manual method.

Table E5.--Summary of base flow estimated by computer hydrograph-separation techniques for three streamflow stations in southeastern Pennsylvania, water years 1961-85

[mi^2, square miles]

Water year	Base flow (inches)								
	West Branch Brandywing Creek near Honey Brook (18.7 mi)			Little Lehigh Creek near Allentown (80.8 mi)			Neshaminy Creek near Langhorn (210 mi)		
	Fixed interval	Sliding interval	Local minimum	Fixed interval	Sliding interval	Local minimum	Fixed interval	Sliding interval	Local minimum
1961	12.23	12.58	12.37	11.92	11.93	11.79	12.47	12.54	12.34
1962	9.63	9.89	8.75	8.98	9.21	8.82	6.83	7.29	6.54
1963	7.27	7.18	6.21	9.57	9.40	9.02	5.07	5.09	4.90
1964	9.27	9.07	7.93	7.98	7.99	7.82	7.28	7.27	6.59
1965	6.66	6.57	6.25	5.52	5.50	5.44	5.57	5.22	4.56
1966	7.17	7.19	6.31	5.27	5.23	5.24	4.43	4.24	3.95
1967	8.96	8.77	8.11	7.65	7.68	7.67	11.01	11.11	10.56
1968	9.07	9.05	8.40	10.33	10.31	9.99	7.98	7.95	7.25
1969	7.70	7.76	7.55	7.62	7.59	7.50	6.88	6.59	6.16
1970	9.75	9.77	9.42	10.55	10.54	10.31	9.30	9.17	8.89
1971	10.67	10.83	10.23	15.62	15.69	15.16	10.32	10.18	8.53
1972	14.75	15.11	14.39	21.34	21.41	21.18	14.75	14.30	13.34
1973	14.43	14.70	14.15	23.11	23.29	23.20	14.53	15.03	14.74
1974	11.71	11.68	11.28	17.83	17.94	17.98	11.27	11.45	11.11
1975	15.13	15.23	13.79	19.66	19.60	19.07	13.75	13.52	12.53
1976	14.33	13.76	12.99	18.45	18.61	18.10	8.84	9.21	8.57
1977	11.67	11.85	11.51	12.89	12.93	12.73	5.02	5.15	4.20
1978	16.36	16.77	15.30	20.59	20.64	20.40	13.27	13.08	12.72
1979	15.59	16.20	13.97	16.53	16.64	16.41	13.63	13.74	12.97
1980	14.06	14.16	13.58	14.80	14.84	14.68	11.59	11.49	10.95
1981	6.47	6.52	6.29	6.97	7.00	7.01	6.23	6.21	5.99
1982	8.98	8.67	8.25	10.55	10.48	10.22	8.50	8.44	7.53
1983	14.14	13.92	12.96	14.79	14.71	14.46	11.91	11.91	11.74
1984	17.52	17.40	15.51	23.03	23.02	22.26	14.95	14.54	13.24
1985	9.56	9.64	9.18	9.18	9.21	9.06	4.05	4.07	3.64
Median	10.67	10.83	10.23	11.92	11.93	11.79	9.30	9.21	8.57

Table E6.--Summary of base-flow estimates by manual base-flow recession and computer hydrograph-separation techniques, Brandywine Creek at Chadds Ford, Pennsylvania, 1928-31 and 1952-53

	Base flow (inches)			
Year	Base-flow recession technique[1]	Fixed-interval technique	Sliding-interval technique	Local-minimum technique
1928	18.23	17.63	17.57	17.24
1929	1.73	1.53	1.93	1.99
1930	8.67	8.36	8.40	8.24
1931	6.15	6.31	6.42	6.44
1952	18.68	19.21	19.23	19.01
1953	16.61	17.35	17.46	16.90

[1]From Olmsted and Hely (1962)

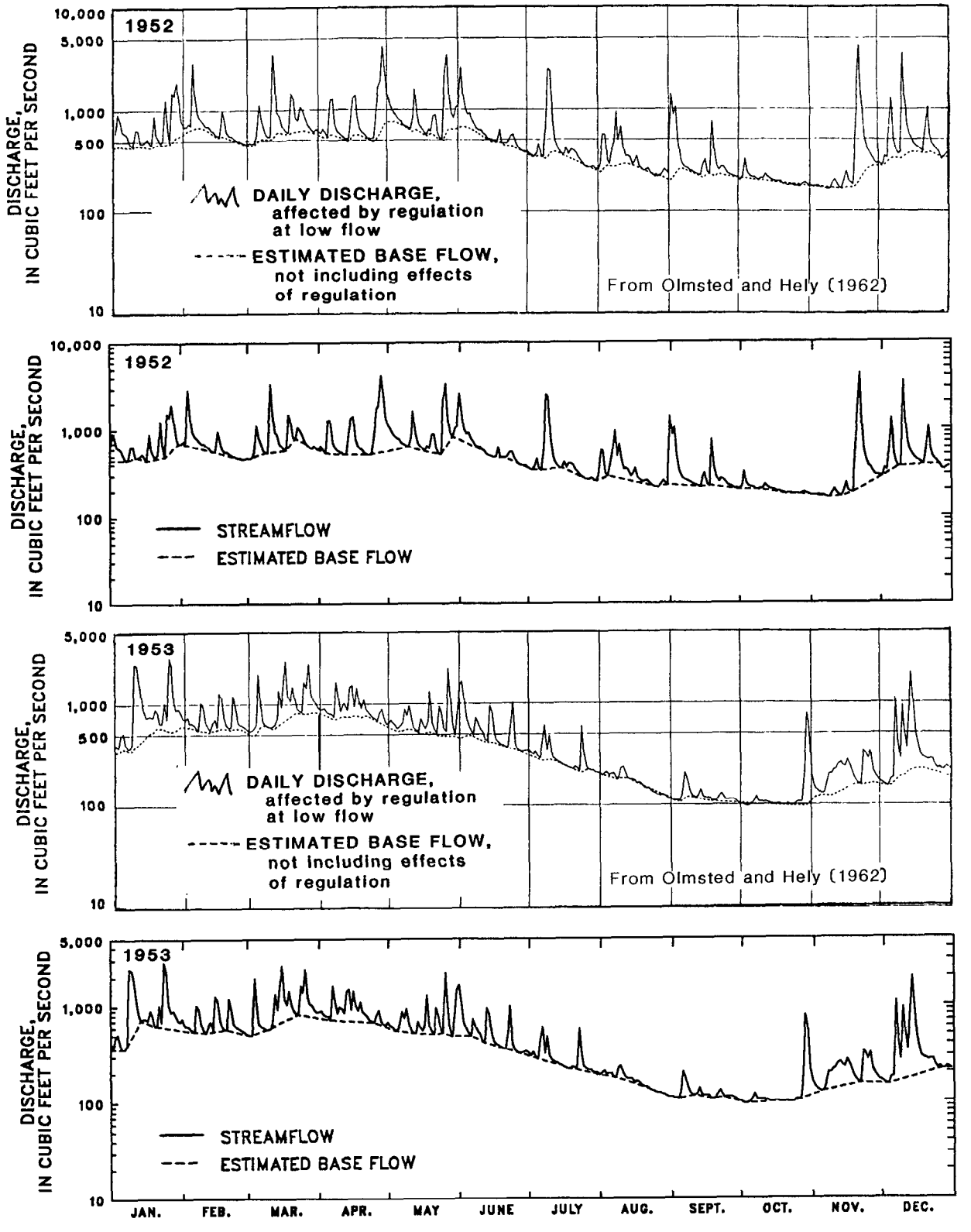


Figure E16.--Daily mean streamflow and base flow for Brandywine Creek at Chadds Ford, Pennsylvania, for 1952-53. Hydrographs from Olmsted and Hely (1962) are based on manual techniques; the other hydrographs are based on computer techniques.

AVAILABILITY OF COMPUTER PROGRAM

The computer code for the hydrograph-separation program and user documentation are available from the Malvern Subdistrict Office of the U. S. Geological Survey, 111 Great Valley Parkway, Malvern, PA 19355. The program length is 2,300 lines of code.

CONCLUSIONS

Computer hydrograph-separation techniques provide a rapid method for estimating ground-water contribution to streamflow. They substantially reduce the time required for hydrograph separation from weeks or days to seconds. Computer techniques give results comparable to manual methods based on base-flow-recession or curve-fitting techniques. They remove subjectivity inherent in manual methods.

REFERENCES CITED

Biesecker, J. E., Lescinsky, J. B., and Wood, C. R., 1968, Water resources of the Schuylkill River basin: Pennsylvania Department of Forests and Waters Bulletin Number 3, 198 p.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L., 1982, Hydrology for engineers, 3rd edition: McGraw-Hill, New York, 508 p.

Olmsted, F. H., and Hely, A. G., 1962, Relation between ground water and surface water in Brandywine Creek basin, Pennsylvania: U.S. Geological Survey Professional Paper 417-A, 21 p.

McGreevy, L. J., and Sloto, R. A., 1977, Ground-water resources of Chester County, Pennsylvania: U.S. Geological Survey Water Resources Investigations Report 77-76, 76 p.

Pettyjohn, W. A., and Henning, Roger, 1979, Preliminary estimate of ground-water recharge rates, related streamflow and water quality in Ohio: Ohio State University Water Resources Center Project Completion Report Number 552, 323 p.

Riggs, H. C., 1963, The base-flow recession curve as an indicator of ground water: International Association of Scientific Hydrology Publication 63, p. 352-363.

Rorabaugh, M. I., 1963, Estimating changes in bank storage and ground-water contribution to streamflow: International Association of Scientific Hydrology Publication 63, p. 432-441.

STATISTICAL AND GRAPHICAL METHODS USED TO DESCRIBE GROUND-WATER QUALITY IN ILLINOIS

R.H. Coupe and K.L. Warner
U.S. Geological Survey, Urbana, Ill.

ABSTRACT

The U.S. Geological Survey uses various statistical and graphical techniques to characterize ground-water quality in Illinois. The statistics and graphics are based on public-supply-well information collected from 1984-87 by the Illinois Environmental Protection Agency and the U.S. Geological Survey. The statistical packages associated with the National Water Information System were used for tabular and graphical representation of ground-water quality in Illinois.

A Geographic Information System was used to display the spatial distribution of wells where selected inorganic constituents in ground water exceeded U.S. Environmental Protection Agency water-quality maximum contaminant levels. For those constituents for which data were sparse but whose presence is important (such as volatile organic carbon, volatile organic aromatic compounds, and pesticides), a Geographic Information System also was used to represent graphically the spatial distribution of the data.

INTRODUCTION

The Illinois Environmental Protection Agency's (IEPA) Division of Public Water Supplies, in cooperation with the U.S. Geological Survey (Survey), sampled approximately 2,080 public-supply wells throughout Illinois during the period 1984-87. The water-quality data collected during this program were useful to IEPA in reviewing current and in proposing new State ground-water-quality standards for health-related chemical compounds.

Prior to proposing the standards, the IEPA first had to analyze incidence concentrations in combination with health effects for a wide range of organic and inorganic constituents. The Survey, using the National Water Information System (NWIS) data bases and a Geographic Information System (GIS), assisted the IEPA in this data analysis.

The need to complete a statistical analysis of these data is immediate but has long-range implications. This is the first analysis of ground-water-quality data from public-supply wells in Illinois. The statistical analysis of that data is useful to IEPA for determining legally enforceable ground-water-quality standards for the State.

This paper documents the steps taken in developing a statistical and graphical description of the quality of ground water in Illinois. The statistical techniques and some of the graphical techniques used are available in NWIS software. Other graphical methods use a GIS available throughout the Survey.

This study was done in cooperation with the Illinois Environmental Protection Agency.

DATA BASE

The data used in this analysis are from one-time-only water samples from Illinois public-supply wells and were collected during March 1984 through July 1987 (Voelker, 1986; Voelker and others, 1988). Approximately 100 wells were selected to be sampled quarterly for 2 years. The first sample only, from each of these sites, was included in the data set. The IEPA plans to sample every public-supply well in Illinois. About 2,080 of 5,000 public-supply wells in Illinois have been sampled as of July 1987. The 2,080 wells that were sampled contain water that is considered to be representative of the five major aquifers in the State.

STATISTICAL METHODS OF ANALYSIS

The IEPA's ground-water-quality data are stored in the Survey's NWIS data base as part of the cooperative efforts between the Survey and the IEPA. The statistical programs used on the NWIS data base were chosen because they provided the needed statistical summaries without having to manipulate the data into some other form for use with a different statistical program.

The simplest of these new statistical programs in NWIS is called QWDETLIMS.F77 (T.L. Schertz, U.S. Geological Survey, written commun., 1988). The output from this program is a listing of all the constituents for selected water-quality records, and the total number of times that each constituent was analyzed (total non-missing values). Also included is a listing of the different detection limits (all values encoded with a '<') for each constituent and a count of the number of times the value from individual analysis equaled or exceeded that detection limit for each year included in the study. An example of the output from the program is shown in figure E17.

This program not only provided a listing of the constituents in the data set, but also indicated those constituents that are present in concentrations above the detection limit and that are suitable for basic statistical analysis. Statistical analysis of a constituent with few values above the detection limit could lead to a misrepresentation of the quality of ground water on a statewide basis. The output from the QWDELTIMS.F77 program indicates if the detection limit for a constituent has changed and when, and gives an indication of when the method of analysis might have changed or a different laboratory was used. The table created by this program allows the user to easily detect inconsistencies in the detection limits and to verify the data before a statistical analysis is done.

To obtain the basic statistics for those constituents that were suitable for basic analysis, a statistical program called QWPRCNTL.F77 was used. The output from QWPRCNTL.F77 (fig. E18) gives the sample size; the maximum, minimum, and mean values; and the percentile ranking of the data, including the median.

PARAMETER: LEAD DISSOLVED 01049 TOTAL NON-MISSING VALUES = 25

DETLIM	1984	1985	1986	1987	SUM	% OF TOTAL
5.00	0	0	0	18	18	72.00
10.00	0	0	5	0	5	20.00

PARAMETER: LEAD TOTAL 01051 TOTAL NON-MISSING VALUES = 2514

DETLIM	1984	1985	1986	1987	SUM	% OF TOTAL
5.00	159	410	1013	783	2365	94.07
1.00	0	0	1	1	2	0.08
7.00	0	1	0	0	1	0.04

PARAMETER: MANGANESE TOTAL 01055 TOTAL NON-MISSING VALUES = 2557

DETLIM	1984	1985	1986	1987	SUM	% OF TOTAL
5.00	19	77	155	219	470	18.38
6.00	0	1	0	0	1	0.04
10.00	0	0	1	0	1	0.04
70.00	0	1	0	0	1	0.04

PARAMETER: MANGANESE DISSOLVED 01056 TOTAL NON-MISSING VALUES = 25

DETLIM	1984	1985	1986	1987	SUM	% OF TOTAL
5.00	0	0	0	2	2	8.00

Figure E17.--Example output from program QWDETLIMS.F77.

STATISTICAL SUMMARY OF SELECTED WATER QUALITY DATA COLLECTED FROM APR 1984 TO JUL 1987

WATER-QUALITY CONSTITUENT	SAMPLE SIZE	DESCRIPTIVE STATISTICS MAXIMUM	MINIMUM	MEAN	PERCENT OF SAMPLES IN WHICH VALUES WERE LESS THAN OR EQUAL TO THOSE SHOWN 95 %	75 %	(MEDIAN) 50 %	25 %	5 %
00610 NITROGEN NH4 ASN TOT	1995	27.000	<0.010	1.042*	3.950	1.050	0.400	<0.100	<0.100
00630 NO2 + NO3 AS N TOT	1998	29.000	<0.100	0.822*	5.350	0.160	<0.100	<0.100	<0.100
00665 PHOSPHORUS TOT AS P	1996	879.000	<0.010	1.262*	0.510	0.100	0.020	<0.010	<0.010
00916 CALCIUM TOTAL EPA	1991	280.000	0.510	80.138	140.000	95.000	75.000	61.000	33.000
00927 MAGNESIUM TOTAL EPA	1991	120.000	0.300	36.955	66.000	45.000	35.000	27.000	15.000
00929 SODIUM TOTAL EPA	1991	1100.000	0.600	58.247	230.000	58.500	28.000	12.000	3.700
00937 POTASSIUM TOTAL EPA	1991	150.000	0.300	3.531	12.000	3.800	2.100	1.300	0.600
00940 CHLORIDE	2006	1700.000	<1.000	41.541*	170.000	38.000	13.000	3.300	<1.000
00945 SULFATE DISS	2005	1500.000	<7.500	103.157*	410.000	120.000	46.000	12.000	<10.000
00951 FLUORIDE TOTAL	1998	4.500	0.100	0.486	1.300	0.600	0.300	0.200	0.100
00956 SILICA, TOTAL	1266	43.000	1.000	14.942	25.000	19.000	15.000	9.800	7.200
01002 ARSENIC TOTAL	2000	96.000	<1.000	4.261*	25.000	2.000	<1.000	<1.000	<1.000
01007 BARIUM TOTAL	1991	23000.000	5.000	228.843	500.000	200.000	90.000	40.000	10.000
01022 BORON TOTAL	1988	2300.000	<8.000	282.523*	1000.000	380.000	150.000	<50.000	<50.000
01037 COBALT TOTAL	1990	30.000	<5.000	3.472*	9.000	<5.000	<5.000	<5.000	<5.000
01042 COPPER TOTAL RECOV.	1990	1700.000	<5.000	9.681*	27.000	<5.000	<5.000	<5.000	<5.000
01045 IRON TOTAL	1990	45000.000	<50.000	1402.073*	4800.000	1800.000	620.000	130.000	<50.000
01055 MANGANESE TOTAL	1990	2100.000	<5.000	89.573*	460.000	74.000	21.000	7.000	<5.000
01067 NICKEL TOTAL	1990	190.000	<3.000	3.362*	12.000	<5.000	<5.000	<5.000	<5.000
01082 STRONTIUM TOTAL	1989	10000.000	5.000	803.072	2900.000	1000.000	390.000	150.000	70.000
01147 SELENIUM TOTAL	1993	39.000	<1.000	0.466*	1.500	<1.000	<1.000	<1.000	<1.000
70300 RESIDUE DIS 180C	1968	3700.000	67.000	571.884	1180.000	648.000	471.000	382.000	300.500

* - VALUE IS ESTIMATED BY USING A LOG-PROBABILITY REGRESSION TO PREDICT
THE VALUES OF DATA BELOW THE DETECTION LIMIT

NOTE: MULTIPLE DETECTION LIMITS DURING THE PERIOD OF RECORD MAY RESULT IN VARYING VALUES
FLAGGED WITH A "<"

Figure E18.--Example output from program QWPRCNTL.F77.

The rankings created by this program clearly indicated where each individual sample fell in relation to the entire data set. The output was also useful in determining which values needed to be verified before continuing the analysis.

One of the most difficult problems in trying to analyze this type of data statistically is determining what should be done with values that are reported below the detection limit. Values flagged with a less-than sign (<) indicate that the laboratory was not able to quantify the constituent below the reported detection limit; it cannot be assumed, however, that the constituent is not present in quantities below that limit. If these less-than values are used in the analysis, then the risk of bias in the data set can occur. In this situation, the calculated mean would be higher than the true mean; if these less-than values are set to zero (a common solution), the calculated mean will be lower than the true mean. The QWPRCNTL.F77 program offers an intermediate solution. If more than 5 percent of the data are reported below the detection limit, the option to estimate the values below the detection limit is available. The mean and percentile rankings are estimated with a log-probability regression procedure. This method estimates the value below a detection limit and uses those values and the detected values of a constituent to estimate the mean of the data and the percentile rankings. The estimating method was chosen as the best way to handle the problems presented by detection limits in water-quality data. An example of the output from program QWPRCNTL.F77 is shown in figure E19.

GRAPHICAL METHODS OF ANALYSIS

The display of data by graphical methods is a useful technique for presenting information to a general audience. Previously, this information was manually plotted on topographic maps or displayed in tabular format by the IEPA.

Exploratory data analysis (Velleman and Hoaglin, 1981) is a practical method of data analysis that minimizes the use of assumptions. The NWIS software has a very useful procedure for exploratory data analysis--the boxplot.

Boxplots are useful for graphically displaying the statistical distribution of the concentrations of a water-quality constituent. The boxplot (fig. E20) graphically shows where the median of the data lies and of how it relates to the rest of the distribution of the data.

The development of a boxplot starts with the median value. Next, the medians of the data above and below the principle median are found (these are the hinges) and a box is drawn enclosing these hinges (fig. E20). The H-spread is the difference between the hinges and gives the range covered by the middle half of the data. The inner fence is defined as the distance 1.5 times the H-spread from either hinge. A line is drawn to the last data value that is not beyond the inner fence. The outer fence is defined as being three times the H-spread from either hinge. Symbols, *'s, are used to show where data values fall within the outer fence. Values beyond the outer fence are shown with o's and are considered far outside values.

STATISTICAL SUMMARY OF SELECTED WATER QUALITY DATA COLLECTED FROM APR 1984 TO JUL 1987

WATER-QUALITY CONSTITUENT	SAMPLE SIZE	DESCRIPTIVE STATISTICS MAXIMUM	MINIMUM	MEAN	PERCENT OF SAMPLES IN WHICH VALUES WERE LESS THAN OR EQUAL TO THOSE SHOWN 95 %	75 %	(MEDIAN) 50 %	25 %	5 %
00610 NITROGEN NH4 ASN TOT	1995	27.000	0.002	1.042*	3.950*	1.050*	0.400*	0.119*	0.029*
00630 NO2 + NO3 AS N TOT	1998	29.000	0.000	0.822*	5.350*	0.171*	0.045*	0.007*	0.000*
00665 PHOSPHORUS TOT AS P	1996	879.000	0.000	1.262*	0.510*	0.100*	0.020*	0.005*	0.001*
00916 CALCIUM TOTAL EPA	1991	280.000	0.510	80.138	140.000	95.000	75.000	61.000	33.000
00927 MAGNESIUM TOTAL EPA	1991	120.000	0.300	36.955	66.000	45.000	35.000	27.000	15.000
00929 SODIUM TOTAL EPA	1991	1100.000	0.600	58.247	230.000	58.500	28.000	12.000	3.700
00937 POTASSIUM TOTAL EPA	1991	150.000	0.300	3.531	12.000	3.800	2.100	1.300	0.600
00940 CHLORIDE	2006	1700.000	0.040	41.541*	170.000*	38.000*	13.000*	3.300*	0.671*
00945 SULFATE DISS	2005	1500.000	0.377	103.157*	410.000*	120.000*	46.000*	13.677*	3.956*
00951 FLUORIDE TOTAL	1998	4.500	0.100	0.486	1.300	0.600	0.300	0.200	0.100
00956 SILICA, TOTAL	1266	43.000	1.000	14.942	25.000	19.000	15.000	9.800	7.200
01002 ARSENIC TOTAL	2000	96.000	0.000	4.261*	25.000*	2.000*	0.467*	0.100*	0.011*
01007 BARIUM TOTAL	1991	23000.000	5.000	228.843	500.000	200.000	90.000	40.000	10.000
01022 BORON TOTAL	1988	2300.000	2.430	282.523*	1000.000*	380.000*	150.000*	57.674*	18.651*
01037 COBALT TOTAL	1990	30.000	0.214	3.472*	9.000*	4.331*	2.599*	1.560*	0.746*
01042 COPPER TOTAL RECOV.	1990	1700.000	0.001	9.681*	27.000*	3.618*	0.818*	0.185*	0.022*
01045 IRON TOTAL	1990	45000.000	2.708	1402.073*	4800.000*	1800.000*	620.000*	130.000*	36.759*
01055 MANGANESE TOTAL	1990	2100.000	0.062	89.573*	460.000*	74.000*	21.000*	7.000*	1.171*
01067 NICKEL TOTAL	1990	190.000	0.027	3.362*	12.000*	3.619*	1.574*	0.687*	0.207*
01082 STRONTIUM TOTAL	1989	10000.000	5.000	803.072	2900.000	1000.000	390.000	150.000	70.000
01147 SELENIUM TOTAL	1993	39.000	0.004	0.466*	1.500*	0.485*	0.219*	0.099*	0.031*
70300 RESIDUE DIS 180C	1968	3700.000	67.000	571.884	1180.000	648.000	471.000	382.000	300.500

* - VALUE IS ESTIMATED BY USING A LOG-PROBABILITY REGRESSION TO PREDICT
THE VALUES OF DATA BELOW THE DETECTION LIMIT

NOTE: MULTIPLE DETECTION LIMITS DURING THE PERIOD OF RECORD MAY RESULT IN VARYING VALUES FLAGGED WITH A "<"

Figure E19.--Example output from program QWPRCNTL.F77 using the option to estimate data points below the detection limit.

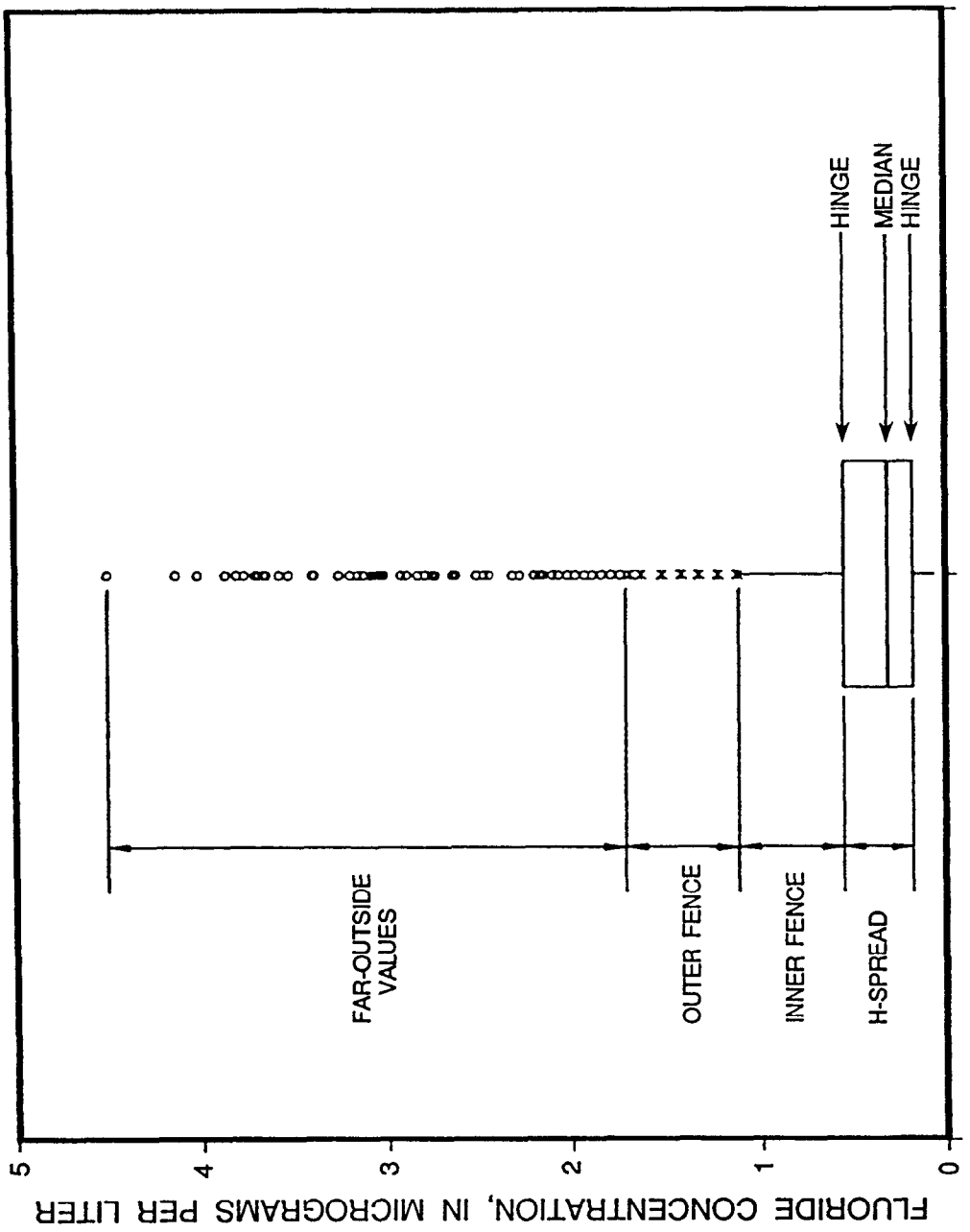


Figure E20.--Example of a boxplot.

One of the investigative uses of boxplots is to organize the data on the basis of hydrogeologic characteristics and to compare the boxplots with one another and with the original data set (fig. E21).

Figure E21 shows that most of the water samples containing arsenic are from wells that have depths less than or equal to 250 feet and that are located in hydrologic units 0512 and 0514 (fig. E21B). The opposite is true for hydrologic unit 0713 where most of the water samples containing arsenic are from wells having depths greater than 250 feet (fig. E21C). These plots by themselves are not conclusive, but their usefulness is in indicating areas where further study is needed.

An ARC/INFO GIS was used to show spatial distribution of wells and detected constituents. This system was chosen because of its data-management capabilities and availability. Most major scientific agencies in Illinois use ARC/INFO; this facilitates the exchange of compatible information.

The chemical constituents were divided into three coverages including (1) inorganic constituents, (2) VOC/VOA (volatile organic carbon and volatile organic aromatic) constituents, and (3) pesticides as suggested by the IEPA. A coverage is a digital analog of a single map sheet stored in a suite of files. The coverages do not include all the constituents sampled, but do include IEPA prioritized inorganic constituents, all VOC/VOA, and detected pesticides (table E7).

Table E7.--Chemical constituents included in each of three coverages: inorganic constituents, volatile organic carbon (VOC) and volatile organic aromatic compounds (VOA), and pesticides

[VOC/VOA includes volatile organic carbon and volatile organic aromatic compounds. This division of constituents was suggested by the Illinois Environmental Protection Agency]

Inorganic constituents	VOC/VOA	Pesticides
Residue upon evaporation	1,1 Dichloroethylene	PCB
	1,1 Dichloroethane	Atrazine
Chlorine	1,2 Trichloroethylene	Alachlor
Sulfate	1,2 Dichloroethane	Metolachlor
Nitrogen	1,1,1 Trichloroethane	
Phosphorus	Carbontetrachloride	
Lead	Trichloroethylene	
Copper	Tetrachloroethylene	
Iron	Chlorobenzene	
	Benzene	
	Toluene	
	Ethylbenzene	

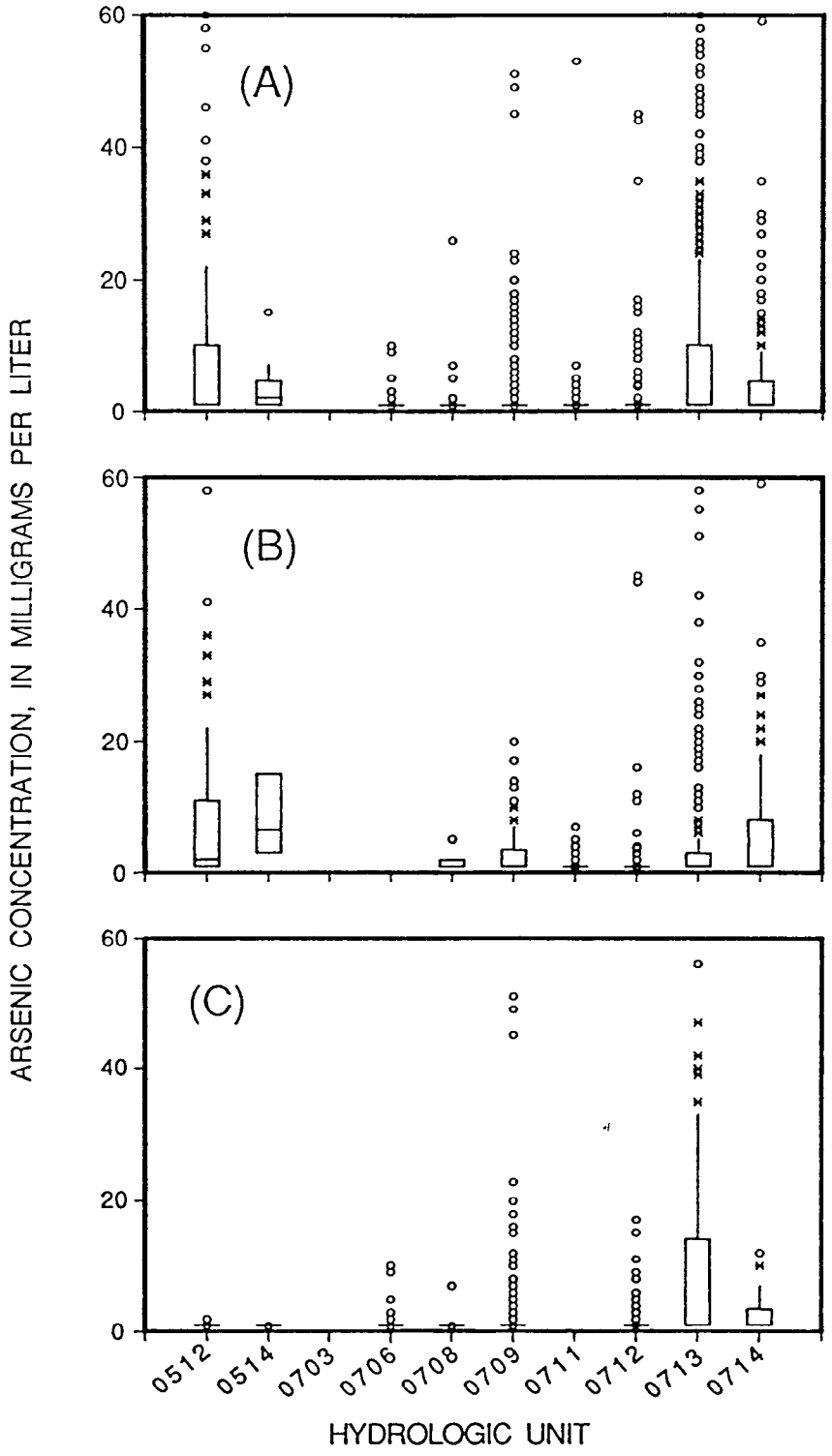


Figure E21.--Arsenic concentration in ground water in Illinois by hydrologic unit for (A) all public wells sampled, (B) wells sampled that were less than or equal to 250 feet deep, and (C) wells sampled that were greater than 250 feet deep.

A GIS was used to display the spatial distribution of wells in which selected inorganic constituents exceeded U.S. Environmental Protection Agency water-quality maximum contaminant levels. For those constituents, for which data are sparse, but whose presence is important (such as volatile organic carbon, volatile organic aromatic compounds, and pesticides), a GIS was used to graphically represent the spatial distribution of the data (fig. E22).

Ground-water sites from NWIS were stored in the data base structure of a GIS. This involved formatting the data to be put into INFO and then into a coverage. A retrieval was run on the NWIS that pulled the first sample for each site as output to a separate file that could be read into INFO. A CPL (command procedure language) formats the file, builds a data base using INFO, and creates a coverage (fig. E22).

A problem arose with encoding concentrations below the detection limit. These values had a preceding less-than sign, which is a character and not an integer. This prevented the values from being defined as integers in INFO. The detection limit value could not be used without the less-than sign because there were some legitimate values at the detection limit. To temporarily solve the problem, so that wells that were sampled but had concentrations below the detection limit could be recorded, all values below the detection limit were defined as '888888'.

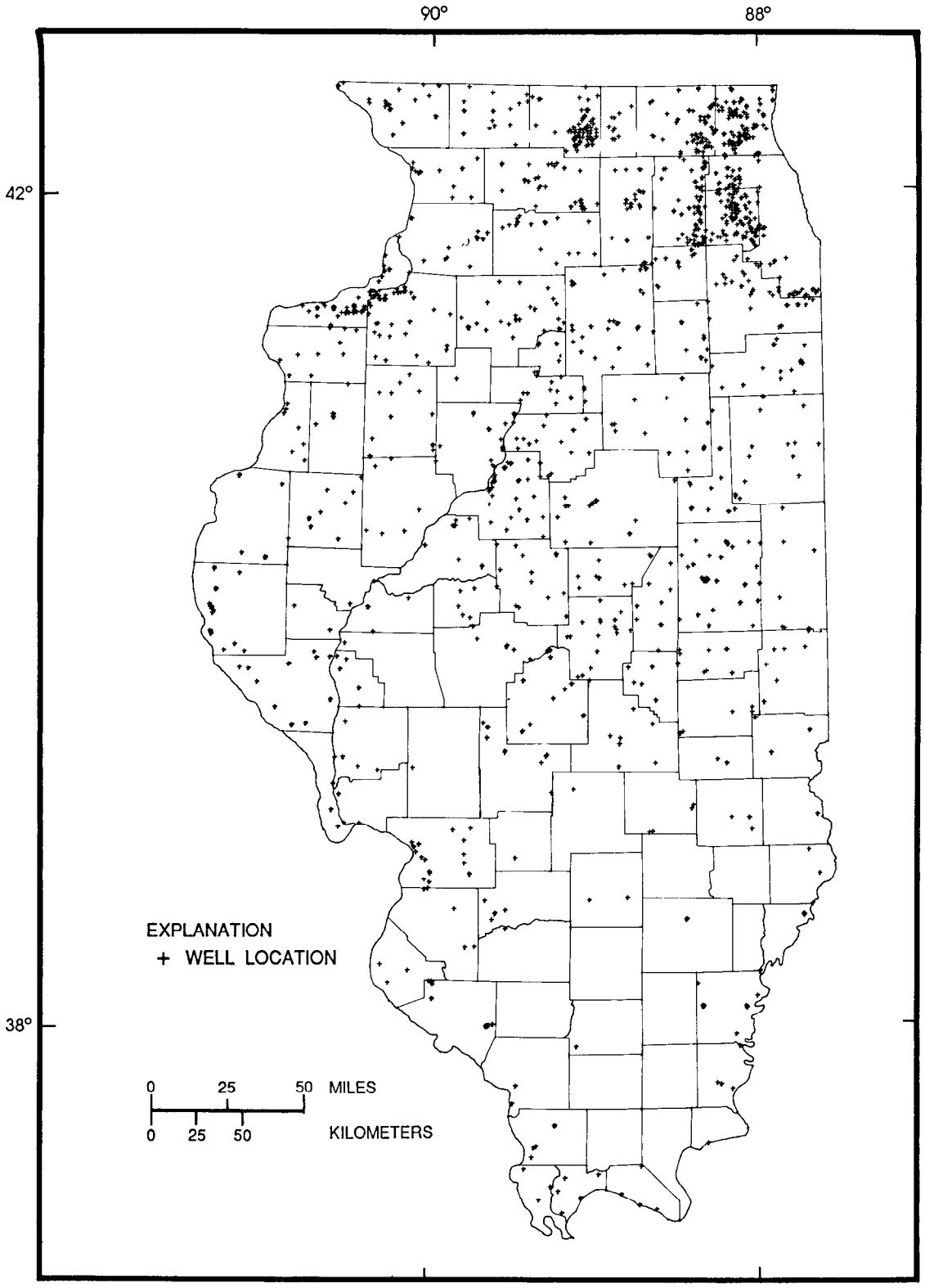
A GIS that uses this type of data has many potential applications. The immediate application was to locate abnormalities in the data. Discrepancies can be detected by overlaying sites for a certain area on a polygon outlining that area. A number of coverages can be used for this purpose including coverages by county, river basin, major aquifer, lake, and river.

By use of the NWIS software in combination with maps developed using the GIS, the IEPA was supplied with the analytical tools they needed to aid in establishing appropriate ground-water-quality standards and to aid in determining the costs of implementing those ground-water standards.

SUMMARY

Illinois has developed water-quality standards for protection of its ground water. The Illinois Environmental Protection Agency and the U.S. Geological Survey sampled 2,080 public-supply wells during 1984-87 for water-quality analysis in Illinois. The U.S. Geological Survey statistically and graphically analyzed this data base.

Chemical constituents for which there were sufficient data to represent ground-water quality in Illinois were identified, and appropriate basic summary statistics were calculated using programs in the NWIS software package. The distributions of these constituents, for selected hydrogeologic characteristics, were represented graphically and compared by use of boxplots. The data were selected and grouped by other hydrogeologic characteristics; these groups were then used to develop boxplots of the data for comparison.



Base from U.S. Geological Survey, 1:2,000,000 digital line graph.

Figure E22.--Location of public-supply wells in Illinois that were sampled as of July 1987.

A Geographic Information System was used to display the spatial distribution of wells in which selected inorganic constituents exceeded U.S. Environmental Protection Agency water-quality maximum contaminant levels. For those constituents for which data were sparse, but whose presence is important (such as volatile organic carbon, volatile organic aromatic compounds and pesticides), the Geographic Information System was used to represent graphically the spatial distribution of the data.

The NWIS software, in combination with maps developed using a Geographic Information System, provided the Illinois Environmental Protection Agency the analytical tools useful for selecting appropriate ground-water-quality standards and for determining the costs of implementing the standards.

REFERENCES CITED

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

Voelker, D.C., 1986, Observation-well network in Illinois, 1984: U.S. Geological Survey Open-File Report 86-416(W), 108 p.

Voelker, D.C., Oberg, D.J., and Grober, M.J., 1988, Water-quality data from the observation-well network in Illinois, 1985-87: U.S. Geological Survey Open-File Report 87-538, 725 p.