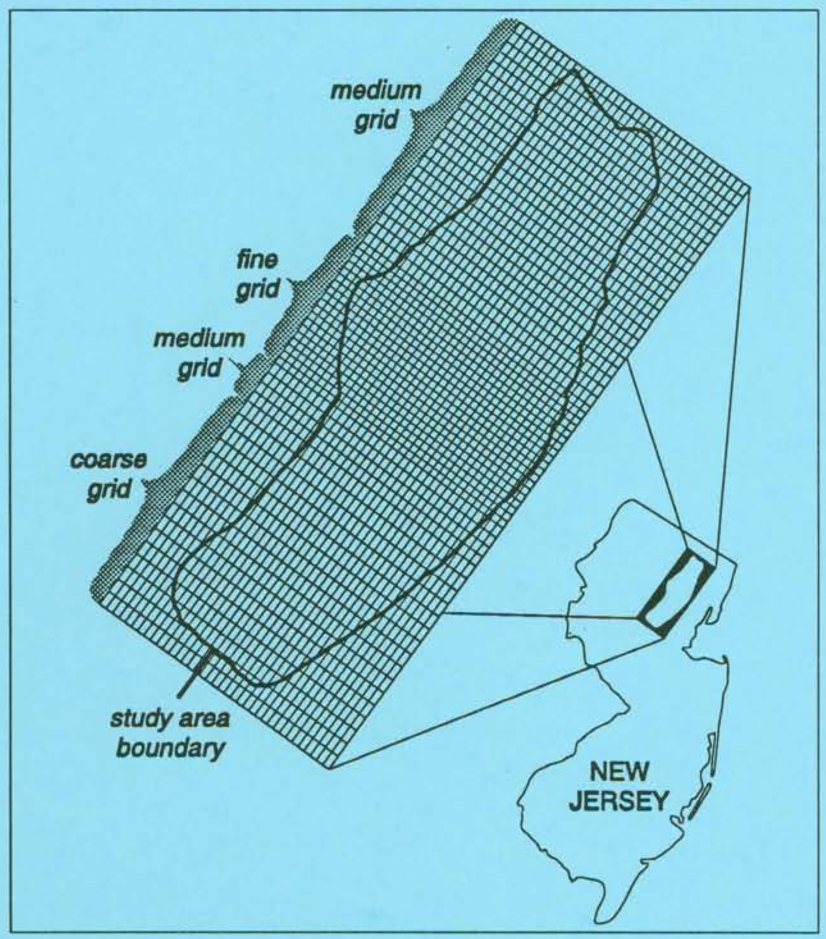
New Jersey Geological Survey
Open-File Report OFR 95-2

GEOGRAPHIC INFORMATION SYSTEM (GIS) DATABASE DEVELOPMENT FOR THE CENTRAL PASSAIC RIVER BASIN HYDROGEOLOGIC INVESTIGATION



N.J. Department of Environmental Protection - Division of Science and Research

NEW JERSEY DEPARTMENT OF ENVIRONMENTAL PROTECTION

The mission of the New Jersey Department of Environmental Protection is to assist the residents of New Jersey in preserving, sustaining, protecting and enhancing the environment to ensure the integration of high environmental quality, public health and economic vitality.

NEW JERSEY GEOLOGICAL SURVEY

The mission of the New Jersey Geological Survey is to map, research, interpret and provide scientific information regarding the state's geology and ground-water resources. This information supports the regulatory and planning functions of DEP and other governmental agencies and provides the business community and public with the information necessary to address environmental concerns and make economic decisions.

Cover Illustration: Model grid for the Passaic study area (see figure 6).

New Jersey Geological Survey
Open-File Report OFR 95-2

Geographic Information System (GIS) Database Development for the Central Passaic River Basin Hydrogeologic Investigation

by

Mark A. French

New Jersey Department of Environmental Protection
Division of Science and Research
Geological Survey
CN 427
Trenton, NJ 08625
1995

Printed on recycled paper

New Jersey Geological Survey Open-File Reports are published by the New Jersey Geological Survey, CN 427, Trenton, NJ 08625. This report may be reproduced in whole or part provided that suitable reference to the source of the copied material is provided.

Additional copies of this and other reports may be obtained from:

DEP Maps and Publications Sales Office
Bureau of Revenue
CN 417
Trenton, NJ 08625

A price list is available on request.

CONTENTS

FIGURES

TABLES

ABSTRACT

The development of a Geographic Information System (GIS) database accompanied the Central Passaic River Basin Hydrogeologic Investigation (Passaic Study). This database provided data for use in the three-dimensional, finite-difference ground-water model, MODFLOW. A short discussion of ground-water flow, the ground-water-flow model and the ARC/INFO GIS is included. The content of each GIS coverage in the database is listed and described. The development of the model grid, the loading of the grid, its integration with the finite difference ground-water model, and the treatment of the model's results are outlined. This report discusses some of the problems encountered in creating the database and recommends possible solutions and improvements to the GIS finite-difference-model integration process.

INTRODUCTION

1.1. Context

The ability of the GIS to handle geographic and spatial data made it particularly well suited for use in ground water modeling for the Central Passaic River Basin hydrogeologic investigation (the Passaic Study) by the New Jersey Geological Survey (NJGS). The inherent ability of the GIS to handle each data set as a discrete layer occupying a geographically concurrent location was particularly important. The development of the GIS database and use of unique and powerful GIS abilities helped meet the overall goals of the Passaic Study. GIS is commonly used to prepare site-specific and regional-scale data for use in ground-water models.

The following statements, from "Plan of study for the Central Passaic River Basin hydrogeologic investigation" (Hoffman, 1989), best define the overall goals and scope of the Passaic Study:

> "This study will define the geology, bedrock topography, ground-water pumpage, water levels, and hydrogeochemistry of the [Passaic Study] area .... The overall goal of this study is to enhance efficient management of the ground-water resources of the Central Passaic River Basin. To do this it is necessary to first quantify those factors which govern ground-water-flow, quality and quantity. Flow paths will be delineated. An investigation of ground-water chemistry will indicate natural quality limitations to ground water use. A computer model will be developed to predict the effects of additional pumping. The model will also be used, if feasible, to determine optimal well locations."

This study covers 167,000 acres (260 square miles) in northeastern New Jersey (fig. 1). It includes all or part of 40 municipalities in 5 counties (fig. 2). The study examines two types of aquifers: buried-valley and bedrock. Funding was from the New Jersey Water Bond Issue of 1981.

It was decided to utilize the ARC/INFO GIS of the New Jersey Department of Environmental Protection (NJDEP) because of the size and complexity of the Passaic Study area. The GIS was utilized to prepare data for use in the finite difference model and also to help analyze model results. It provided a method by which a large part of the manual preparation and analysis of data could be avoided. It also provided an inherently more accurate method of data preparation and analysis.

Manual preparation would have entailed drafting of the pertinent data maps and the model grid, and hand

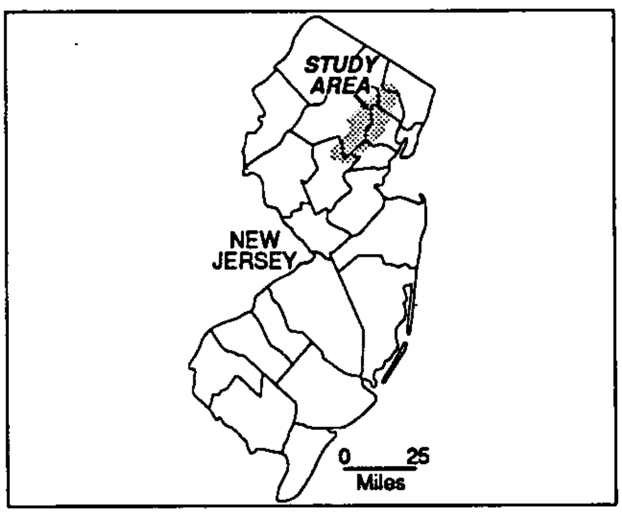


Figure 1. Location of the Passaic study area in New Jersey.

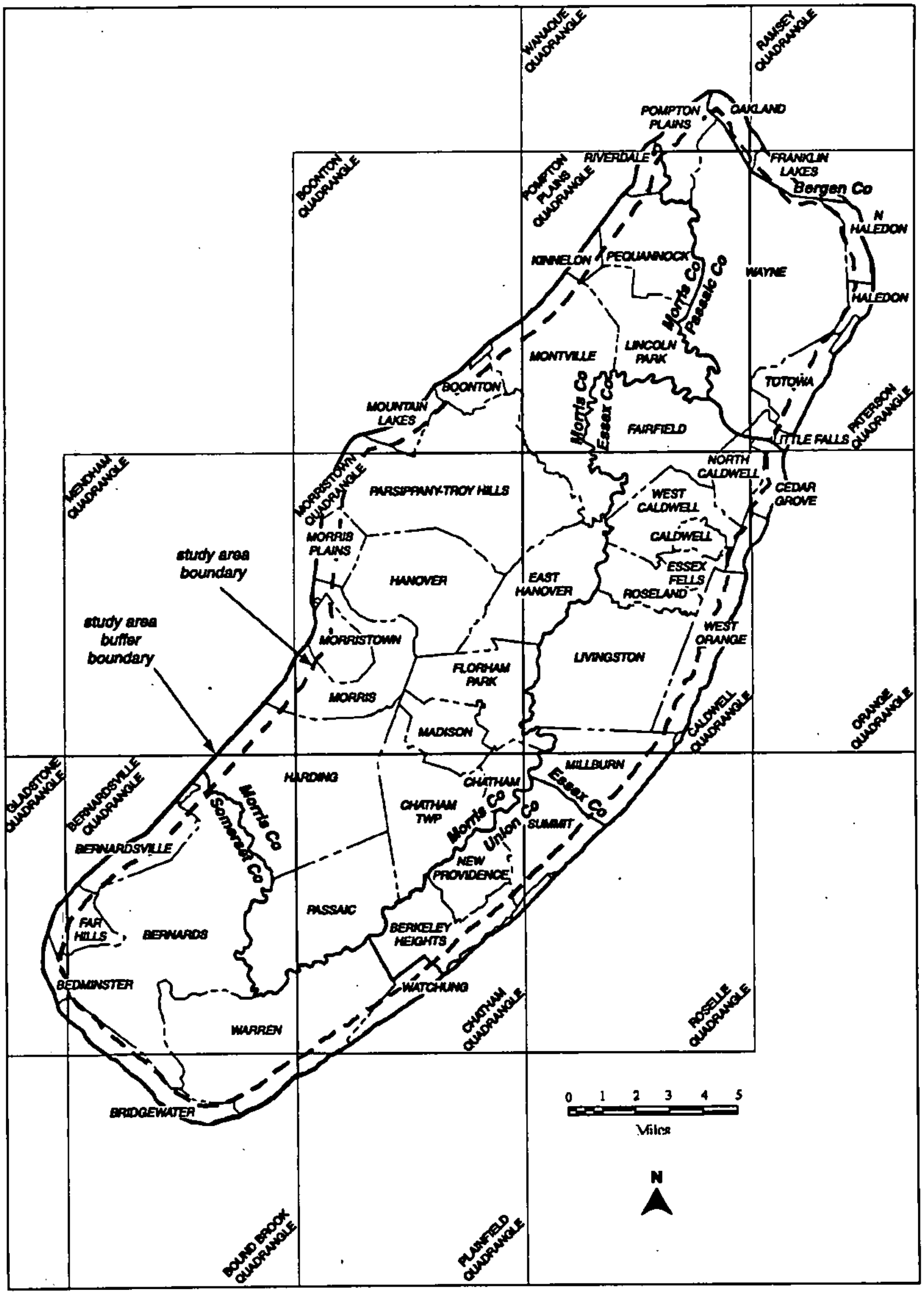


Figure 2. Study area of the Central Passaic River Basin Hydrogeologic Investigation.

coding of the individual grid cells. The manual creation of digital arrays for the model would have entailed data entry using a spreadsheet application. For the Passaic Study, this would have meant the individual coding of thousands of individual cells in four data layers.

The GIS provided an automated way to achieve the same result in less time with more accuracy. Digital maps or coverages which incorporated the pertinent data in attached INFO data tables were developed. A model grid was digitally generated to the precise specifications of the hydrogeologist. The grid was then overlaid on each data coverage. This enabled each cell to capture the pertinent data parameters from the INFO data tables. These data were then converted to digital arrays and downloaded for use in the model.

The resultant GIS database was large, inherently complex and dynamic. Several coverages were created specifically to provide input data for the model, as indicated above, while other coverages were developed for preparation of maps and report illustrations. Extensive developmental work also produced many intermediate coverages.

1.2. Overview of GIS database

Data for the Passaic Study and the GIS database were collected and assembled using various methods. More than 300 wells were located and their locations digitized. Data from 21 stream gages, 15 precipitation gages, and the evaporation gage at Canoe Brook were used in water budget calculations. Unpublished maps showing the elevations of the tops and bottoms of glacial till were obtained from the United States Geological Survey (USGS). Bedrock topography was determined from geophysical investigations, well records and other documents. Other data were obtained from Digital Elevation Models (DEM), Digital Line Graph (DLG) files, field investigations, and record and document searches.

The database includes more than 100 coverages stored in fifteen subdirectories. The coverages ranged in complexity from a single-point coverage representing an evaporation gage to highly complex polygon and line coverages. These included raw data, intermediate stages of combination and development, and final coverages. Final data coverages were used to supply data to the model grid and to create graphics for presentations and publications. Coverages included point, line, and polygon data with INFO point, line, and polygon attribute tables (PAT, AAT, and PAT, respectively); Triangulated Irregular Networks (TINs); lattice files; and other digital files.

Map features included political boundaries, roads, hydrography, topography and geology, all with attached INFO data tables. During development of the database, coverages were used to supply intermediate data to the Passaic Study. Owing to storage-space restrictions, only final coverages remain on the system. Most raw data coverages were archived to tape, and most intermediate or developmental coverages were deleted.

1.3. Acknowledgments

This pioneering integration of the NJDEP Geographic Information System with a ground-water modeling effort was made possible through the contributions and hard work of the following people and organizations: Jeffrey Hoffman, project coordinator for the Central Passaic River Basin Hydrogeologic Investigation, and David Hall, geophysicist, both of the NJGS, Bureau of Ground Water Resources Evaluation (BGWRE). The efforts of the following former NJGS employees must also be acknowledged: Gail Carter and Jack Schooley (BGWRE), and Mark Fiorentino (Bureau of Geology and Topography), and David LaShell (student assistant). The GENGRID.AML, ISECT.CPL, MATRIX.CPL and other programs were written by Randy Ulery and Mary M. Chepiga, USGS, and modified by Curtis Price, USGS, and Jack Schooley and Mark Fiorentino, NJGS. The database was created on the NJDEP Geographic Information System with support provided by the Bureau of Geographic Information and Analysis, Office of Resources Management.

2. DATA REQUIREMENTS AND USAGE

2.1. Overview of ground-water flow and the ground-water flow model--MODFLOW

More than a century ago, Henry Darcy, a French hydraulic engineer, derived a mathematical formula describing water flow through porous media. This formula and the concepts it embodies came to be known as Darcy's Law. It states that the flow rate through porous media is proportional to the head loss in the water flowing through the media, and inversely proportional to the length of the flow path. He derived this statement through experimental observations of water flowing through a cylinder of sand.

Darcy's Law is represented mathematically by the following equation:

$$Q = K A \frac{\delta h}{\delta l} \quad \text{or} \quad v = \frac{Q}{A} = K \frac{\delta h}{\delta t} \qquad (1)$$

where

Q = flow rate of the water or other liquid

K = hydraulic conductivity

A = cross-sectional area of porous media through which the water or liquid flows

$\frac{\delta h}{\delta l}$ = hydraulic gradient (the difference in head over the length of the flow path).

v = Darcy's velocity, or the flow velocity

Darcy's Law is the basis of present day knowledge and analysis of ground-water flow. The constant Darcy used, hydraulic conductivity (K), is a basic descriptor of the ability of a porous medium to transmit ground water. Many other terms and equations have been derived from Darcy's Law. One that is of the most importance for this report is conductance, which can be represented by an equation of the form:

$$C = \frac{K A}{L} \qquad (2)$$

where

C = conductance

K = hydraulic conductivity

A = cross-sectional area of porous medium through which the water or liquid flows

L = length of the flow path

Darcy's Law only describes unidirectional flow through isotropic (homogeneous) porous media under steady state conditions. Most ground-water flow is not wholly unidirectional and occurs under unsteady flow conditions. The aquifers through which ground-water flows are, for the most part, anisotropic or heterogeneous materials such as mixed sands, clays, gravels, and even fractured rock. In order to analyze ground-water flow under real world conditions, it is often necessary to consider ground-water flow in three dimensions under unsteady flow conditions. To do this, the hydraulic conductivity of the aquifer material is analyzed along three principal axes of flow using the following partial differential flow equation derived from Darcy's Law:

$$K_x \frac{\delta^2 h}{\delta x^2} + K_y \frac{\delta^2 h}{\delta y^2} + K_z \frac{\delta^2 h}{\delta z^2} = S_s \frac{\delta^2 h}{\delta t^2} \qquad (3)$$

where

K_x, K_y, K_z = hydraulic conductivity values along the principal axes of hydraulic conductivity

h = potentiometric head of water in the porous medium

t = time

S_s = specific storage of the porous aquifer material

This is the equation on which ground-water flow models are based. This equation, along with specifications of ground-water flow and/or head conditions at the aquifer system boundaries and the initial head conditions in the aquifer constitute the mathematical representation of a ground-water flow system.

The ground-water model used in the Passaic Study was "A Modular, Three Dimensional Finite Difference Ground-Water Flow Model (MODFLOW)," by MacDonald and Harbaugh (1988) of the USGS. This model is used extensively by the USGS, and by other agencies. A finite difference model, such as MODFLOW, is "... a computational procedure based on dividing an aquifer into a grid and analyzing the flows associated within a single zone [or grid cell] of the aquifer" (Todd, 1980). The object of this discretization of the aquifer is to approximate a solution to the partial differential flow equation (eq. 3) for the aquifer or aquifer system being modeled.

Because this equation cannot be solved directly, a numeric method is used by MODFLOW to obtain an approximate solution. This method, the finite difference method, replaces the continuous system represented in equation 3 with a finite set of discrete points in space and time. The partial derivatives of head are replaced by terms calculated from the differences in head at these points. This approach leads to a system of simultaneous linear algebraic equations which yields head values for the simulated aquifer or flow system at these specific heads and times.

In order to create these discrete points, the aquifer is divided into a mesh or grid of r rows by c columns by v layers, with each layer corresponding to a geohydrologic unit. This grid is set up so that the hydraulic properties are uniform within each cell. A set of internal points or nodes is produced centered either between the grid lines (block-centered) or at the intersection of the grid lines (point-centered), as shown in figure 3. These nodes make up the finite set of discrete points.

MODFLOW uses the block-centered grid cell, and flows are calculated between each cell and its six neighbors sharing common faces (fig. 4). This flow calculation is based upon a form of continuity equation in which "The sum of all flows into and out of a single grid cell must be equal to the rate of change of the affected cell" (MacDonald and Harbaugh, 1988). This flow is based upon the hydraulic conductivity of the regions between cell nodes expressed as one-dimensional, steady-state flow through a block of aquifer of a specified cross-sectional area. Each cell face constitutes the cross-sectional area and the distance between nodes constitutes the length of the flow path. This calculation of flow produces a value of head for the affected cell (fig. 5). All of the internal flows are handled as conductance in order to simplify the resultant finite-difference flow equation.

External flows which affect head in the cell must also be analyzed. These flows can be either independent of the head in the affected cell or dependent upon head. Flow from independent sources, such as wells and drains, are set equivalent to the flow rate of those independent flow sources. Dependent flows, such as streambed seepage, are treated as simple conductance. Streambed seepage is calculated as vertical hydraulic conductivity of the streambed multiplied by the cross-sectional area and the thickness of the streambed for the reach of the stream across the cell. These external flows are then summed to create one flow term which is then added to the six internal flow terms.

Analysis of all of the flows affecting the head in a single cell thus produces an equation of the form:

$$C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + QS = S_s \frac{\Delta h}{\Delta t} (\Delta r \Delta c \Delta v) \qquad (4)$$

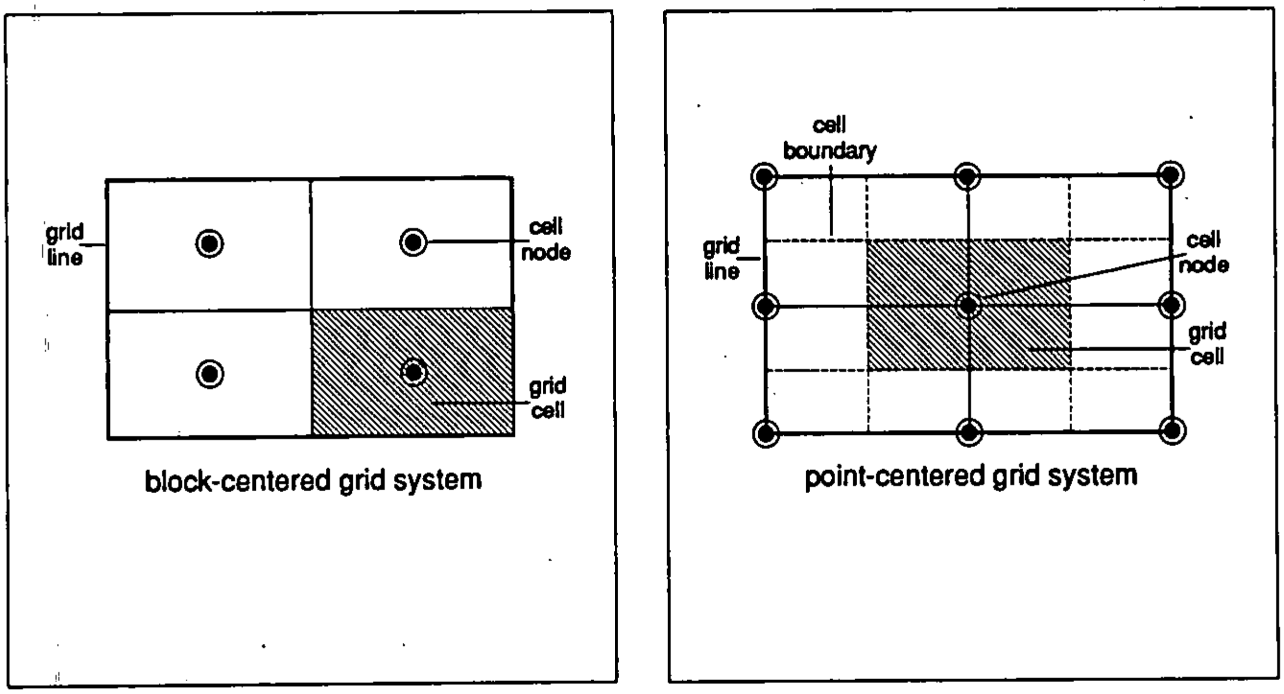


Figure 3. Grid cell and node approaches for ground-water flow models.

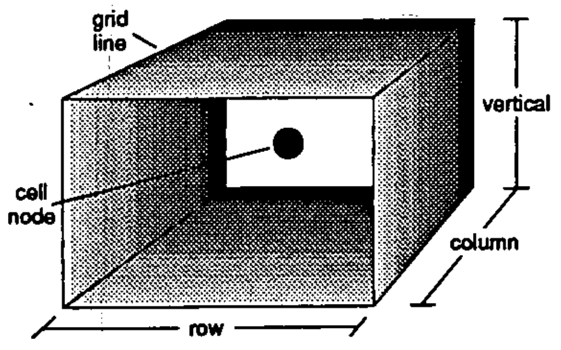


Figure 4. Three-dimensional grid cell for a block-centered grid finite-difference model such as MODFLOW.

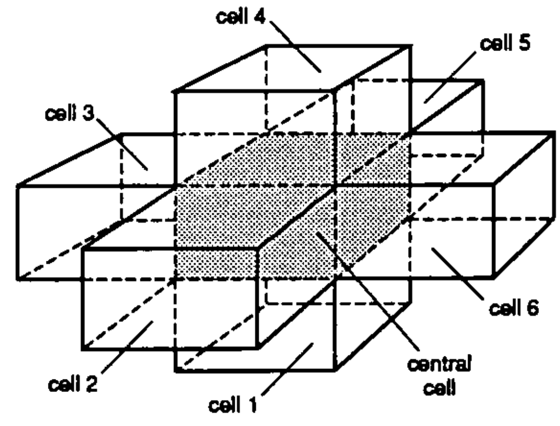


Figure 5. Flow analysis unit used in MODFLOW.

where

C_n = conductance from each adjacent cell

QS = combined flow from external sources

$\Delta h/\Delta t$ = change in head over change in time

rcv = volume of the cell

S_s = specific storage

The term $\Delta h/\Delta t$ represents the finite-difference approximation of the partial derivative of head with respect to time.

This approximation simulates flow in one cell for a discrete time only. MODFLOW must also account for the passage of time. This is done by discretization of the time to be analyzed into steps. The time derivative is approximated using the change in head during an interval preceding and including the time at which the current flow values are to be evaluated. This approach, termed backward difference, is used because it is always numerically stable, and errors introduced at any time diminish progressively at succeeding times. However, it leads to large systems of equations and to the following modification of equation 4:

$$C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + QS = \frac{h_2 - h_1}{t_2 - t_1}(\Delta r \Delta c \Delta v) \qquad (5)$$

where

$h_2 - h_1$ = change in head over the time interval being considered

$t_2 - t_1$ = change in time (the time interval)

An equation of this type is created by MODFLOW for each cell whose head varies with time. Solution of all of these produces an approximation of the partial differential flow equation. These equations, however, cannot be solved independently. Each one consists of seven unknown values of head with seven known coefficients. They can therefore be used to create a system of many equations in many variables. These equations are then solved simultaneously.

The objective of this transient solution is to predict the distribution of head values any time after the initial values, boundary conditions and other hydraulic parameters have been specified. These are the initial values at the first of the discrete time steps into which the time axis of the finite-difference equation is divided. The head values are calculated for the first time step using the system of equations previously described. Once these values are obtained this process is repeated using the solutions from the preceding time step as the initial values for the next time step. This process is repeated until that stress period is complete. A series of stress periods is used during each simulation.

In order to solve the system of finite difference equations for each stress period, MODFLOW uses an iterative process. At the beginning of each time step, an estimate is made of what the head value may be in each cell at the end of that time step. These estimates are then tested and altered. This creates interim head values which are used as estimated head values for the next step in the iteration. As the calculations progress through each iteration, the interim values more closely satisfy the system of equations, and the changes in head values between iterations become smaller.

In order to determine a stopping point for the iterative process, a closure or convergence criterion must be set. This criterion is set by the modeler and consists of an absolute head-value change. This value is compared to the largest absolute head-value change for each iteration. If this change value is larger than the specified criterion, the iterations continue; if not, the iterative process stops and these last head values are taken as the solutions for that time step. MODFLOW will only allow a limited number of iterations, and will cease to operate if the convergence criterion is not adequate or is unrealistic.

In summary, MODFLOW analyzes ground-water flow in an aquifer or aquifer system in which both time and space are approximated as discrete steps. This makes it possible to analyze the flows affecting each cell of a grid as a change in cell head over time. MODFLOW handles this analysis by creating a system of equations consisting of finite-difference flow equations for each cell. The changes in this system are then approximated through an iterative process which continues until the estimated interim head values are within a preset convergence value. The iterative process is carried out successively for each time step of each stress period of a simulation.

2.2. Overview of the ARC/INFO Geographic Information System

The digital handling of georelational data is the core concept of a Geographic Information System (GIS). A GIS integrates the handling of two types of geographic data: spatial information (location and shape) and descriptive information.

Two types of GIS are in use -- raster and vector. They differ in the manner in which they record, store, and represent spatial and descriptive information. A raster-based GIS represents data based on cells defined by rows and columns. Each cell has data that represent geographic features. This approach is similar to the way in which MODFLOW handles data. The grid is georeferenced by a Cartesian coordinate system by means of cell dimension and registration of a corner (usually upper left or right).

A vector GIS maps geographic features as points, lines and areas. Spatial information is represented using a Cartesian coordinate system. Geographic features are recorded as sets of xy coordinate values. A point is stored as a single xy coordinate value. Lines are recorded as a series of ordered xy values, and areas as sets of coordinates defining lines segments that enclose the area. The relationships between features are defined by topology. Topology defines areas, describes how linear features connect, and identifies contiguous areas.

ARC/INFO is a vector GIS. Information is maintained as linked spatial- and descriptive-information file sets called coverages. Coverages are the basic georelational unit maintained by ARC/INFO. A coverage is a digital version of a single map sheet layer. It generally describes one class of geographic feature, but may include many feature classes. The feature classes are arcs (lines), nodes, polygons (areas), points, tics, and annotation.

Arcs are linear map features. Nodes represent the endpoints and intersection points of a line. Polygons are defined as sets of arcs instead of as xy coordinate lists. In ARC/INFO a map or coverage is made up of basic elements consisting of points, arcs, and nodes. Other features, such as polygons, are made up of these basic elements. These other, larger features are composed of other smaller features, meaning that they are topologically defined.

Topology defines spatial relationships. Three major concepts of topology are connectivity-- (expressed by the statement that arcs connect to each other at nodes); area definition-- (stating that arcs that connect to surround an area define a polygon); and contiguity-- (which states that arcs have direction, and left and right sides).

Arcs are made up of xy pairs called vertices, which define the shape of each arc. Each arc has two endpoints or nodes defined as the "from" (starting) node and the "to" (ending) node. Arcs are only joined at nodes. Connectivity is maintained by ARC/INFO as it keeps track of all the arcs connected by each node.

Polygons (defined areas) are stored as lists of arcs instead of as lists of coordinates. The polygon is therefore topologically represented by the arcs that define it. Arcs defining polygons may be stored in multiple polygon-arc lists, but the coordinates for each arc are only stored once in each arc's coordinate list.

Contiguity is maintained by defining the left and right side of arcs based on the direction of the arc itself (as indicated by the location of the "from" and "to" nodes). Thus polygons that share arcs are defined as being adjacent by the arcs that make up their boundaries. ARC/INFO also defines the external or universe polygon to maintain the left-right topology of arcs.

ARC/INFO also has the ability to handle surfaces which represent continuous, three-dimensional features with an infinite number of points defining them. On a map they are represented by contour lines. A GIS models these features to facilitate analysis. Many systems for modeling surfaces have been devised. One is a Digital Elevation Model (DEM). This is a systematically spaced sample of elevations producing a lattice of points with xyz values. ARC/INFO models surfaces by creating a Triangular Irregular Network (TIN) of the surface. This is a network of linked triangles in which the apexes are points with xyz values.

Descriptive data on geographic features is stored in attribute tables. These tables are created in INFO, ARC/INFO's native relational database, or in any other common relational database such as ORACLE, SYBASE, or INFORMIX. Attribute tables store data in a record-and-item or row-and-column data structure. Each item or column represents different data-attribute types, and each record or row holds the data that describe a single geographic feature. Attribute tables are linked to each feature through the feature-identity number maintained by ARC/INFO.

This ability of a GIS to handle diverse sets of georelational data made it an ideal tool for creating much of the information needed for a ground-water flow simulation in MODFLOW. A GIS can efficiently store and manipulate the many coverages created for use in modeling. It can also facilitate quick and accurate loading of these data to the model's grid. This action effectively translates ARC/INFO's vector-based data to the raster-based data that MODFLOW requires.

2.3. GIS data requirements for MODFLOW

Many data sets were produced for the Passaic Study completely or partially using the GIS. The data sets required to run an aquifer simulation in MODFLOW, were provided in whole or part as shown in table 1.

The data in these coverages were converted to a form which could be utilized by the model and downloaded, or downloaded and then modified by the hydrogeologist for use in the model.

2.4. Treatment of model results

MODFLOW is being run on an IBM-PC compatible computer. The modeling process requires several historical calibration runs before the final predictive model runs. The predictive runs produce sets of ground-water-head values corresponding to different water use scenarios at each cell node.

Historical calibration runs are now being run to compare output of the model at well locations in the grid with real ground-water records from these wells to assess the accuracy of the data, particularly the accuracy of hydrologic variable data. The input data are being adjusted to create a better match with the historical records. These adjustments are made to the PC-based data matrices, but not to any of the GIS coverages. However, some coverages, such as the bedrock topography, were modified and new values captured and/or calculated, and then downloaded to the model. Each calibration run therefore requires a comparison with the historical data. These comparisons are made by the hydrogeologist. At the time this report was written, historical calibration was still proceeding. Once the model has been calibrated to the satisfaction of the hydrogeologist, the final predictive runs will be made.

The final runs will produce a matrix of ground-water-head values at each grid cell. This matrix will then be uploaded to GIS as an ASCII file. This file will be used as a lattice file to produce a TIN representation of the potentiometric surface of the aquifers in the Passaic Study area. This TIN surface will then be contoured by the GIS. These contours will then be inspected and approved by the hydrogeologist. After an acceptable set of potentiometric surface contours is produced, the ground-water flow paths will be delineated by the hydrogeologist.

3. DEVELOPMENT OF THE PASSAIC-STUDY DATABASE

3.1. Model grid development

In order to transform the data from the coverages into a form that can be used by the model, a variable-cell-size grid was developed. This grid was initially hand-delineated by the hydrogeologist. It was laid out parallel to the long axis of the study area, which trends northeast-southwest. The resultant grid (fig. 6) was 60 cells long by 32 cells wide. Cell dimensions varied from 2,000 by 2,000 feet to 2,000 by 4,000 feet. The smaller cells, producing a finer grid, were used in the central section of the study area. This was an area of relatively higher interest, data availability, and importance. Progressively larger cell sizes, producing a progressively coarser grid, were used in areas of relatively less interest in the northern and southern sections of the study area.

This grid was reproduced digitally using an AML program, GENGRID.AML (Ulery and others, unpub.), produced by the USGS to generate the grids used in its ground-water models, and modified for use by the NJGS. GENERID.AML produces user-specific, variable-cell grids in the desired geographic location and orientation. The output is a polygon grid and lattice for data loading and a line-grid coverage for display. The lattice and line grids are combined in one coverage, and the polygon grid is another.

3.2. Data capture and edit

The development of the database began with a listing of topics for which ARC/INFO GIS coverages might be needed. The list served as an outline for development of the database.

Simple coverages were produced while data were gathered for development of the more complicated coverages. Data were obtained from many sources in both digital and analog form. They were placed into the database using diverse methods including the manipulation of existing, on-system data using ARC/INFO commands, digitizing from paper or mylar bases, and upload from ouside sources such as the USGS.

About 30 percent of the coverages were created from existing coverages using ARC/INFO commands. The commands included reselect, union, identity, intersect, and append. Several coverages were created by clipping larger, completed data coverages with the study area buffer coverage, BUFSTAREA. This was itself created

Table 1. MODFLOW data requirements

Data required by MODFLOW	GIS data supplied
1. Model grid - discretization of study area	1. Digital grid (PASLAB and PASGRD)
2. Volumetric budget	2. Rain and stream-gage coverages, rainfall Thiessen polygons and stream-gage contribution area (GAGERAIN, GAGESTRM, and GAGETHIES)
3. Layer elevations - aquifer elevations	3. Bedrock and surface topography, glacial till elevations (BEDROCON, ROCKPILE, CON123, BURDCONLIN, SEDCOVER, and till coverages)
4. Hydrologic variables: transmissivity, hydraulic conductivity, specific yield, storage coefficient, vertical leakance	4. Bedrock and glacial (surficial) geology (GEOLOGY, and GLACSEDS)
5. External flow conditions: flows to and from streams, water bodies, wetlands and wells	5. Production-well locations, streams, water bodies, wetlands and percentage of cell covered by water (GWPWELLS, WATER, WETLANDS)
6. Initial head values	6. (Not provided by GIS)
7. Boundary conditions	7. Boundary cell matrix

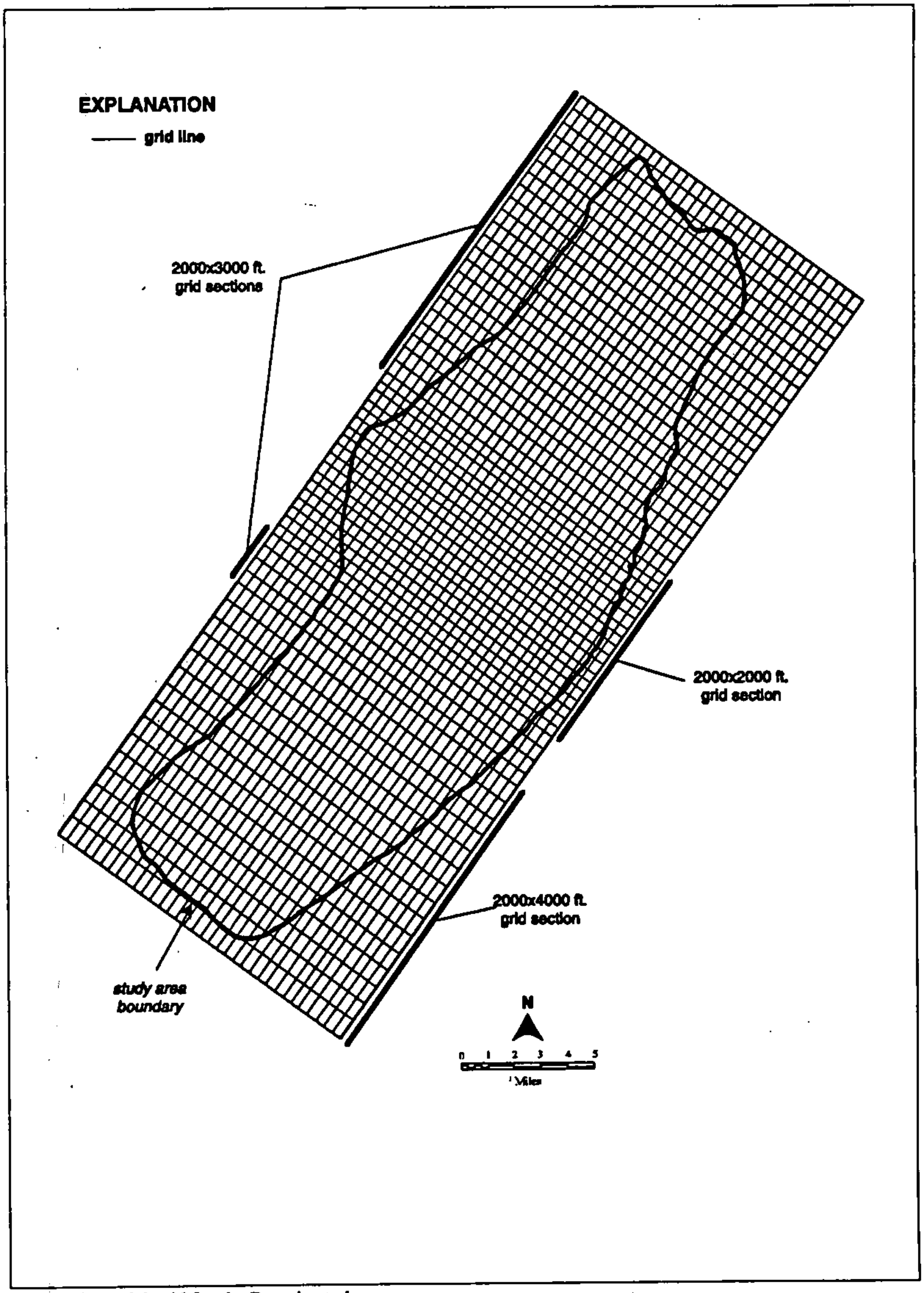


Figure 6. Model grid for the Passaic study.

by buffering one-half mile around the study area boundary coverage, STAREA.

Manual digitizing was used for about 50 percent of the coverages. It was also used to modify and edit previously created coverages through both the ARC Digitizing System (ADS) and ARCEDIT. New coverages were digitized from both paper and mylar-based data using ADS. All the newly digitized coverages were initialized using the featureless Basemap coverage as the tic and boundary template. The RMS error was variable due to the condition of the base material and differing levels of experience of the digitizing staff. RMS errors ranged from a poor 0. 010 to a very good 0.0001.

About 20 percent of the coverages were purchased and uploaded. Purchased data included DLGs, DEMs and ARC/INFO coverages. Uploaded data usually required processing to create acceptable coverages. Data sets usually consisted of longitude-and-latitude coordinate lists. These were projected to state plane coordinates prior to generation of coverages. The DEM files were first projected to state plane coordinates, then converted to lattice files for processing in ARCTIN.

All resultant coverages were edited in ARCEDIT. Attribute tables were modified in INFO, ARC, and ARCEDIT. Nonsystem editing was done on paper, and the changes transferred to the database through ARCEDIT. Editing thus served two purposes: to produce system-acceptable coverages and to create coverages that fit the needs of the ground-water model. The production of system-acceptable coverages followed established system requirements. Editing for the needs of the model was based on input from the hydrogeologist, and resulted in a particularly dynamic database.

3.3. Coverages

Coverages are grouped in 14 subdirectories according to the major features they represent. These are:

1. Basins: Drainage subdivisions of the Passaic River basin

2. Bufcovs: (Buffered coverages): Coverages clipped using the BUFSTAREA coverage

3. Buried-Valleys: Buried valleys and sand-and-gravel aquifers

4. DEM: DEM files, associated TIN directories, coverages, and ASCII files

5. Gages: Stream-gage, precipitation-gage, and evaporation-gage coverages

6. General: Coverages which do not fit under any other subdirectory; these include STAREA, BASEMAP, and wetlands coverages

7. Geology: Geology of Central Passaic River Basin, Ramapo Fault coverage

8. Grid: Model grid

9. Isolines: Bedrock-topography contours

10. Overburden: Overburden isopachs (equal thickness contours), TINs (subdirectory of Isolines subdirectory)

11. Till: Topographic contours on glacial till. (Provided by the USGS, subdirectory of Isolines subdirectory)

12. Roads: Reselected road coverages

13. Streams: Surface-hydrography coverages from USGS DLG files

14. Wells: Observation well, production well, and ground-water quality wells

3.3.1. Coverage documentation

A. Basins

Table 2 contains a complete listing of all subdirectories and the coverages that they include. The coverages are as follows:

1. PASBAS: Stream gage contribution basins reselected from the statewide drainage basins coverage (1:24,000) on the basis of the INFO attribute PRI.NAME containing "PASSAIC" (fig.7). It includes all New Jersey basins that make up the Passaic River Basin, including secondary, tertiary, and lower

Table 2. Passaic study GIS coverages

Directory	Coverages	Description	Input Scale
A. Basins	ALLCONBAS	Stream-gage contribution basins for the Central Passaic Basin	1:24,000
	CON5BAS	Stream-gage contribution basins for 5 major gages	1:24,000
	PASBAS	All of the Passaic River subbasins excluding those in New York	1:24,000
B. Buried Valleys	BURVALLEYS	Buried valleys within the study area	1:24,000
	QSDCIRCLES	Graphic representation of bedrock well pumpages	generated
	TRBCIRCLES	Graphic representation of overburden well pumpages	1:24,000
	SGAQUIFERS	Sand and gravel aquifers within the study area	1:24,000
C. Bufcovs	BUFBASINS	Clipped Passaic River subbasins	1:24,000
	BUFCOLINE	Clipped NJ county - line option	1:24,000
	BUDCOPOLY	Clipped NJ county - polygon option	1:24,000
	BUFGEO	Clipped 1:250,000 geology	1:250,000
	BUFGLACSEDS	Clipped glacial sediments	1:100,000
	BUFMUN	Clipped NJ municipality	1:24,000
	BUFQUADS	Clipped NJ USGS quadrangle	generated
	BUFBEDROCON	Clipped bedrock elevation contours	1:24,000
	BUFSTAREA	The 1/2 mile study area buffer	1:24,000
	BUFSTRMJLH	Selected streams from BUFSTRMS	1:100,000
	BUFSTRMS	Clipped Passaic hydrography	1:100,000
	BUFSWAMP	Clipped swamps	1:63,360
	BUFWTLNDS	Clipped USGS-delineated wetlands	1:100,000
D. Dem	CON123	Contours developed from TIN123 surface	1:250,000
	PTS123	Lattice from 1:250,000 DEM files	1:250,000
	TIN123	TIN surface developed from PTS123	generated
E. Gages	GAGEEVAP	Evaporation gage	1:100,000
	GAGERAIN	Rainfall gage	1:100,000
	GAGESTRM	Stream gage	1:100,000
	GAGETHIES	Rainfall gage Thiessen polygon	1:100,000
F. General	BASEMAP	Featureless, TIC/BND	1:24,000
	STAREA	Study area boundary	1:24,000
	SWAMP	Inferred swamps (wetlands) from 1894 map	1:63,360
	WETLANDS.USGS	USGS delineated wetlands	1:100,000
G. Geology	GEOLOGY	1:250,000 Geology	1:250,000
	FAULT	Ramapo Fault - main fault only	1:250,000
H. Grid	PASGRD	Grid polygon	generated
	PASLAB	Lattice and grid line	generated
I. Isolines	BEDROCON	Bedrock contours, 50-foot intervals	1:24,000
	ROCKPILE	Shaded bedrock contours	1:24,000
	ROCKTIN3	Bedrock TIN surface	generated
J. Overburden	BURDCONLIN	Overburden isopachs 50-foot contour interval	1:250,000
	BURDTIN	Overburden TIN surface	generated
	SEDCOVER	Shaded isopach	1:250,000
K. Till	BCALD	Contours for bottom of till in Caldwell quadrangle	1:24,000
	BCHAT	Contours for bottom of till in Chatham quadrangle	1:24,000
	BMORR	Contours for bottom of till in Morristown quadrangle	1:24,000
	GLACSEDS	Glacial sediments	1:100,000
	TCHAT	Contours for top of till in Chatham quadrangle	1:24,000
	TMORR	Contours for top of till in Morristown quadrangle	1:24,000
L. Roads	MAJRDS	Selected interstates, major highways, and railroads	1:24,000
M. Streams	LAKES	Water bodies reselected from BUFSTRMJLH	
	PASSTRMS	Clipped Passaic River hydrography	1:100,000
	WATER	Passaic Basin watercourses buffered and appended to LAKES	1:100,000
N. Wells	GWPWELLS	Ground water production wells	1:24,000
	GWQWELLS	Ground water quality monitoring wells	1:24,000
	OBSWELLS	Ground water observation wells	1:24,000

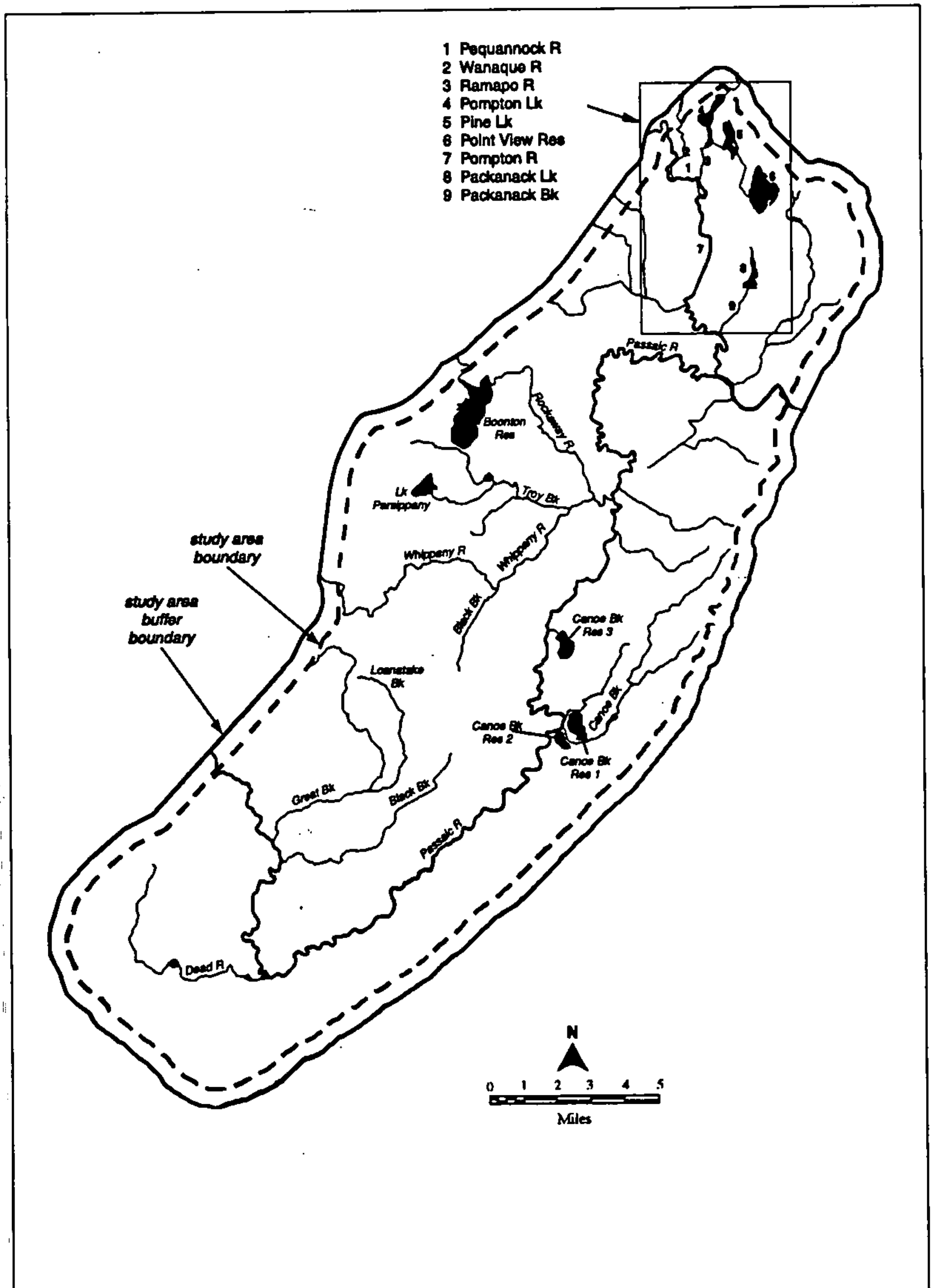


Figure 7. Selected streams, lakes, and reservoirs in the Passaic study area.

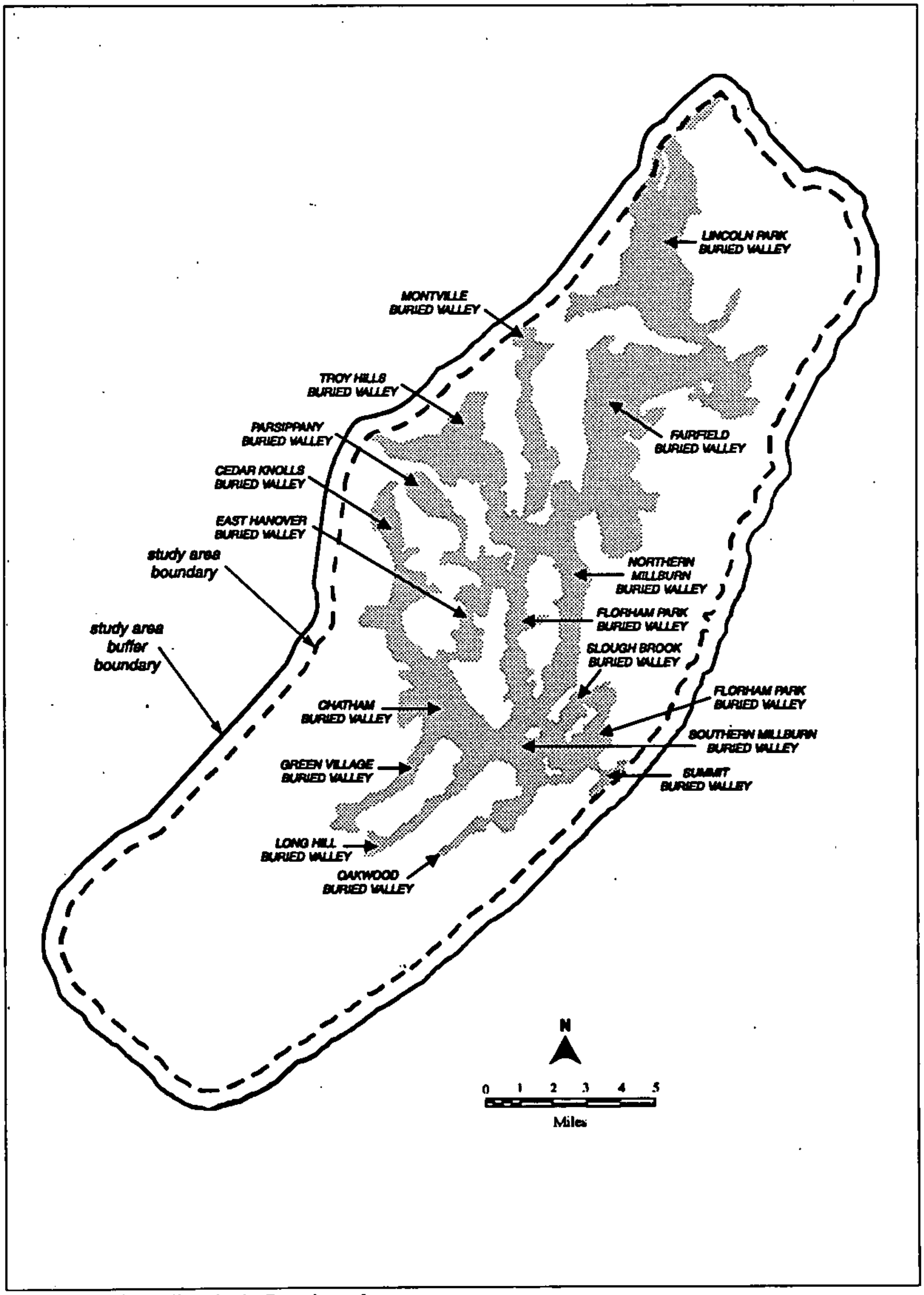


Figure 8. Buried valleys in the Passaic study area.

level drainage basins. The following two coverages were created from PASBAS.

2. ALLCONBAS: Stream gage contribution-basin coverage for all gages within the study area. Basins were delineated graphically. Contributing basins or parts of contributing basins which were outside the study area were delineated using both graphical means and INFO attributes. Interior basin lines were then deleted in ARCEDIT. This produced a group of coverages which were combined to create the ALLCONBAS coverage.

3. CON5BAS: A subset of the ALLCONBAS coverage. It shows only the basins that contribute to four selected inflow gages and the one outflow gage of the study area. The inflow gages were located on the Passaic River at Millington, Whippany River at Morristown, Rockaway River below Boonton Reservoir, and Pompton River at Pompton Plains. The outflow gage is on the Passaic River at Little Falls. A special attribute was added to ALLCONBAS.PAT. This attribute was updated on the basis of the location of the basin with respect to each of the listed gages. Interior lines were then dissolved based on the value of this attribute.

B. Bufcovs (Buffered coverages)

Coverages in this subdirectory were all clipped using the half-mile study area buffer coverage, Bufstarea, with a fuzzy tolerance of 0.001. The unclipped versions are fully described under the appropriate subdirectory headings. These coverages also represent most of the coverages which were used to load the model grid with data (see table 1, section 2.2). Their input scales are the same as those of coverages from which they were clipped.

1. BUFBASINS: Clipped directly from the Passaic subbasins coverage, PASBAS, using the polygon option of the clip command.

2. BUFCOLINE: County boundary coverage clipped from the statewide digital county boundary coverage using the line option.

3. BUFCOPOLY: Another county boundary coverage, clipped from the same coverage as BUFCOLINE, but using the polygon coverage.

4. BUFGEO: Bedrock geology clipped from the GEOLOGY coverage using the polygon option.

5. BUFGLACSEDS: Surficial/glacial geology clipped from the GLACSEDS coverage with the polygon option.

6. BUFMUN: Municipal boundaries clipped from digital statewide coverage using the polygon option.

7. BUFQUADS: Quadrangle grid clipped from the statewide digital USGS 7.5-minute quadrangle neatline coverage using the polygon option.

8. BUFBEDROCON: Bedrock elevation contours clipped from BEDROCON using the line option.

9. BUFSTAREA: Study area buffer. Created by buffering the study area coverage, STAREA, one-half mile and cleaning as a polygon with a dangle length of 100 and a fuzzy tolerance of 0.001.

10. BUFSTRMJLH: Hydrography coverage consisting of features selected by the hydrogeologist (fig. 7). These features were selected from the BUFSTRMS coverage. The unselected features were deleted in ARCEDIT.

11. BUFSTRMS: Stream coverage clipped from converted 1:100,000 DLG files. It is basically the same as the PASSTRMS coverage.

12. BUFSWAMP: Clipped from the SWAMP coverage, digitized from a map from the 1894 Annual Report of the State Geologist (Salisbury and Peet, 1894). Clipped using the polygon option.

13. BUFWTLNDS: Converted 1:100,000 DLG wetlands coverage clipped with the polygon option.

C. Buried valleys

1. BURVALLEYS: Buried valley delineations (fig. 8). Created by reselection of the 100-foot contour line from the bedrock surface contours. Corrections were then made on a 1:24,000-scale draft plot. These cor-

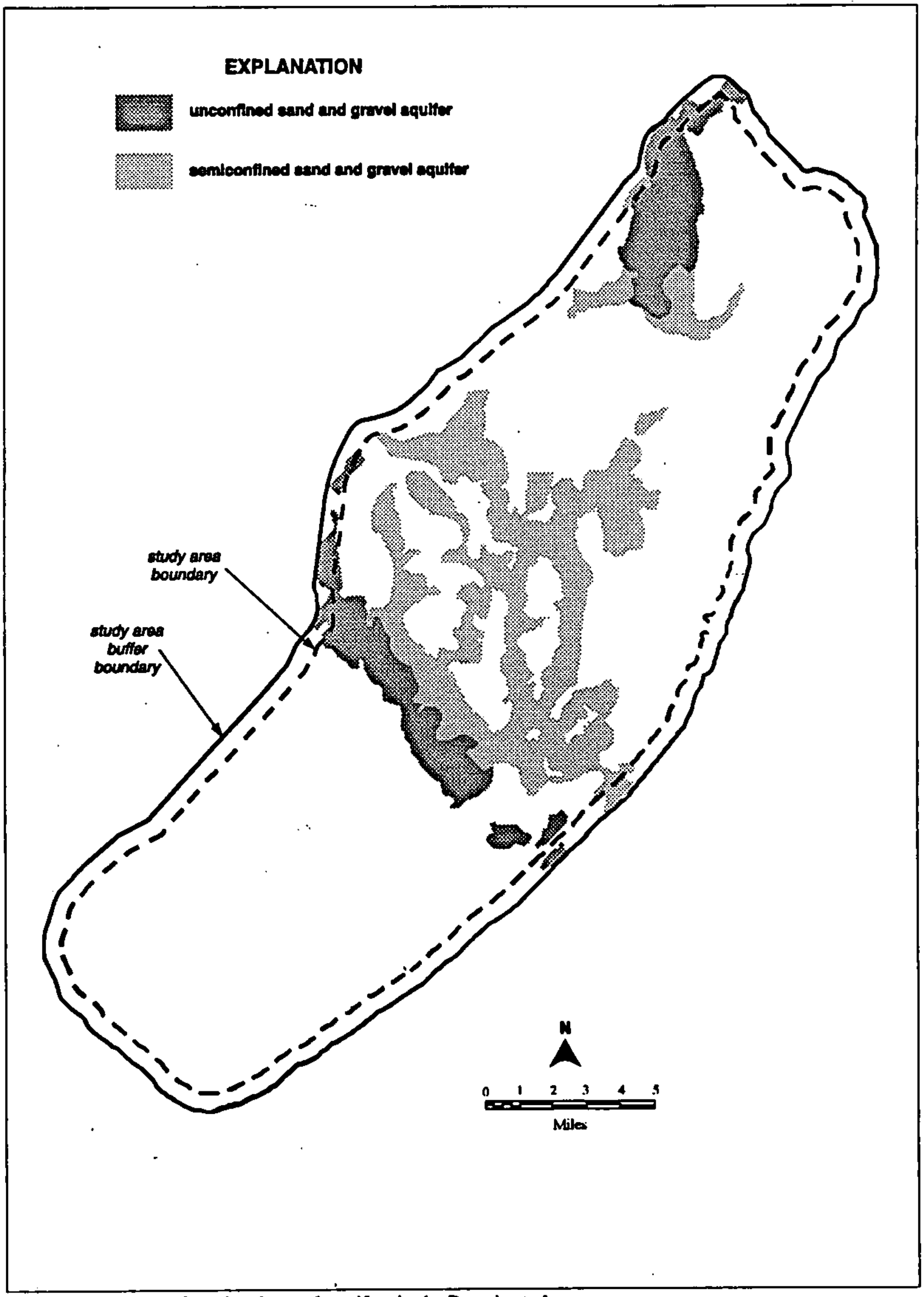


Figure 9. Location of sand and gravel aquifers in the Passaic study area.

rections were added to the 100-foot contour line in ARCEDIT. The RMS for the correction ranged from 0.002 to 0.005.

2. QSDCIRCLES: Generated circles that represent pumpage from wells at the center of the circle. The wells were grouped into pumpage ranges and each range given a unique circle radius. The circles were created using the generate command with the radius value from the PAT of the selected wells. This coverage represents pumpage from wells in the buried valleys and sand-and-gravel aquifers outside the buried valleys.

3. TRBCIRCLES: Generated in the same manner as the QSDCIRCLES. It represents pumpage from wells completed in Triassic/Jurassic bedrock.

4. SGAQUIFERS: Type and extent of sand and gravel aquifers (fig. 9). Created by combining the terminal moraine (reselected from the surficial geology coverage, BUFGLACSEDS) and the buried valley (BURVALLEYS) coverage. Editing was done in ARCEDIT, and shading attributes were added to the INFO PAT in ARC and INFO.

D. DEM

1. CON123: Ground surface topographic contours from the TIN123 TIN. Created using the tincontour command with the default fuzzy tolerance. The input scale was 1:250,000, and the contour interval 50 feet.

2. PTS123: Irregular sample-point coverage developed from three of the statewide 1:250,000 DEM lattice files. These files were first run through the VIP filtering command of ARCTIN to reduce the number of points needed to create TIN123. Three irregular sample point coverages were created in this manner. These coverage were then combined using the append command to form PTS123.

3. TIN123: TIN developed from the PTS123 coverage described above using the arctin command. Weed tolerance was set to 100 feet, and the proximal tolerance was set to 300 feet. This TIN is a representation of the ground surface topography in the study area (fig. 10).

E. Gages

1. GAGEEVAP: Location of the Canoe Brook Atmospheric Gaging Station. This is the only station in or near the study area that monitors evaporation. It was reselected from the GAGERAIN coverage (fig. 11).

2. GAGERAIN: Rain-gage point coverage. 15 gage locations in or near the study area (fig. 11). About half of the gages were digitized in ADS from a 1:100,000-scale map. The rest were from geographic coordinates of the gages.

3. GAGESTRM: Locations of the 21 stream gages in the study area (fig. 11). Created exclusively from geographic coordinates. The location of each gage was then checked and corrected if necessary. Corrections were made in ARCEDIT, and each gage was then snapped to the BUFSTRMJLH streams coverage. Snapping distances were held under 50 feet and the changes were checked.

4. GAGETHIES: Rainfall Thiessen polygon coverage (fig. 11) created using the rain gage coverage, GAGERAIN, and the Thiessen command. This coverage was used to calculate a water budget for the study area.

F. General

1. BASEMAP: A featureless coverage used as the TIC/BND template upon which all digitized coverages were based. It holds the tics used to register those coverages. These included quadrangle corners from the statewide quadrangle coverage and other tics digitized from the 1:24,000-scale base map.

2. STAREA: Study area boundary (fig. 2). Digitized in ADS at 1:24,000 and subsequently adjusted in ARCEDIT along the line of the Ramapo/Border Fault and digital drainage basin divides. The final coverage was cleaned with a dangle length of 50 and a fuzzy tolerance of 0.001.

3. SWAMP: Digitized in ADS from a map by Salisbury and Peet (1894). Areas defined as peat, humus, alluvium or swamps were treated as wetlands. The input scale was 1:63,360 (1 inch equals 1 mile). Latitude and longitude lines were plotted on the map

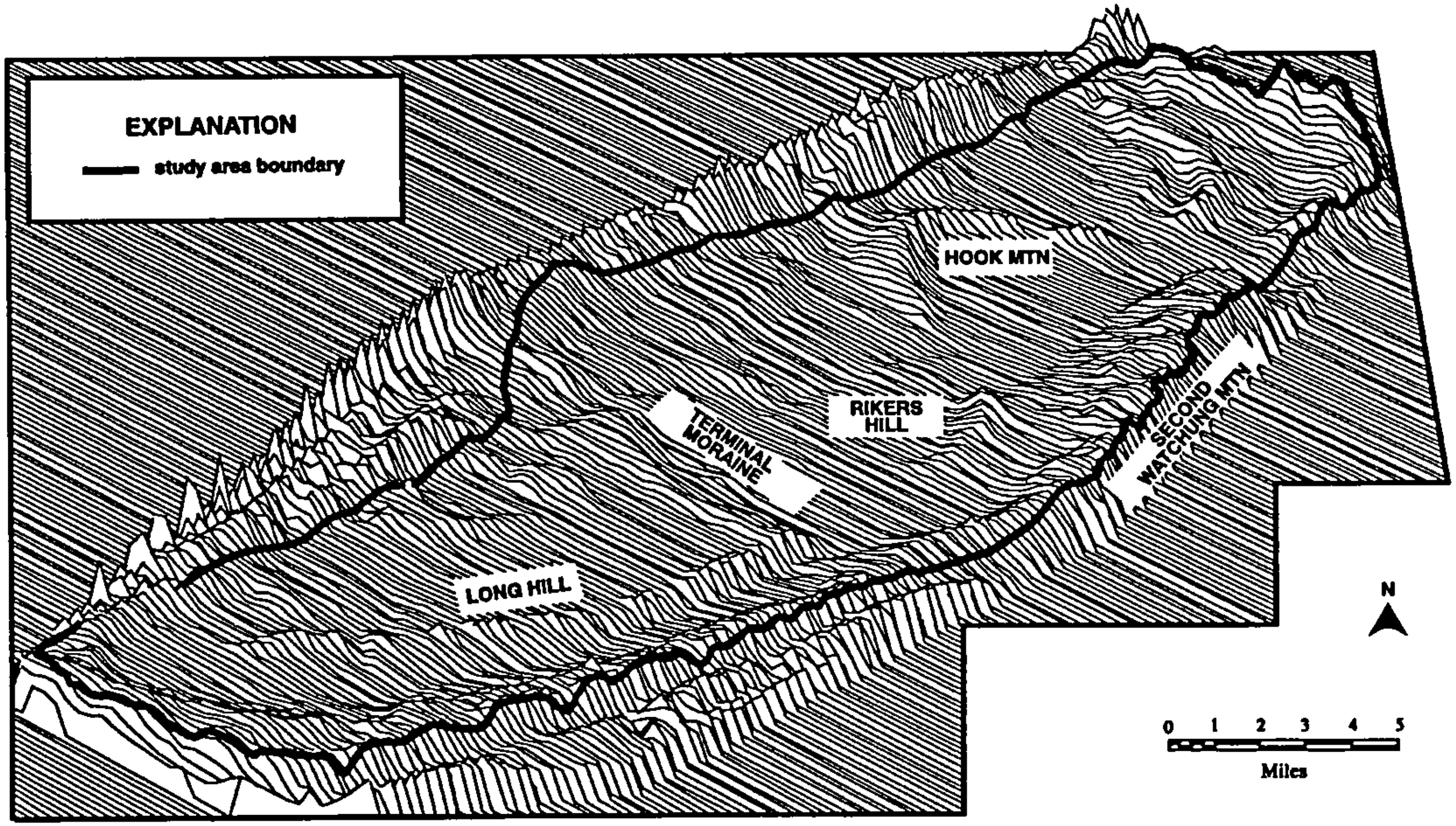


Figure 10. A computer-generated representation of the surface topography in the Passaic study area.

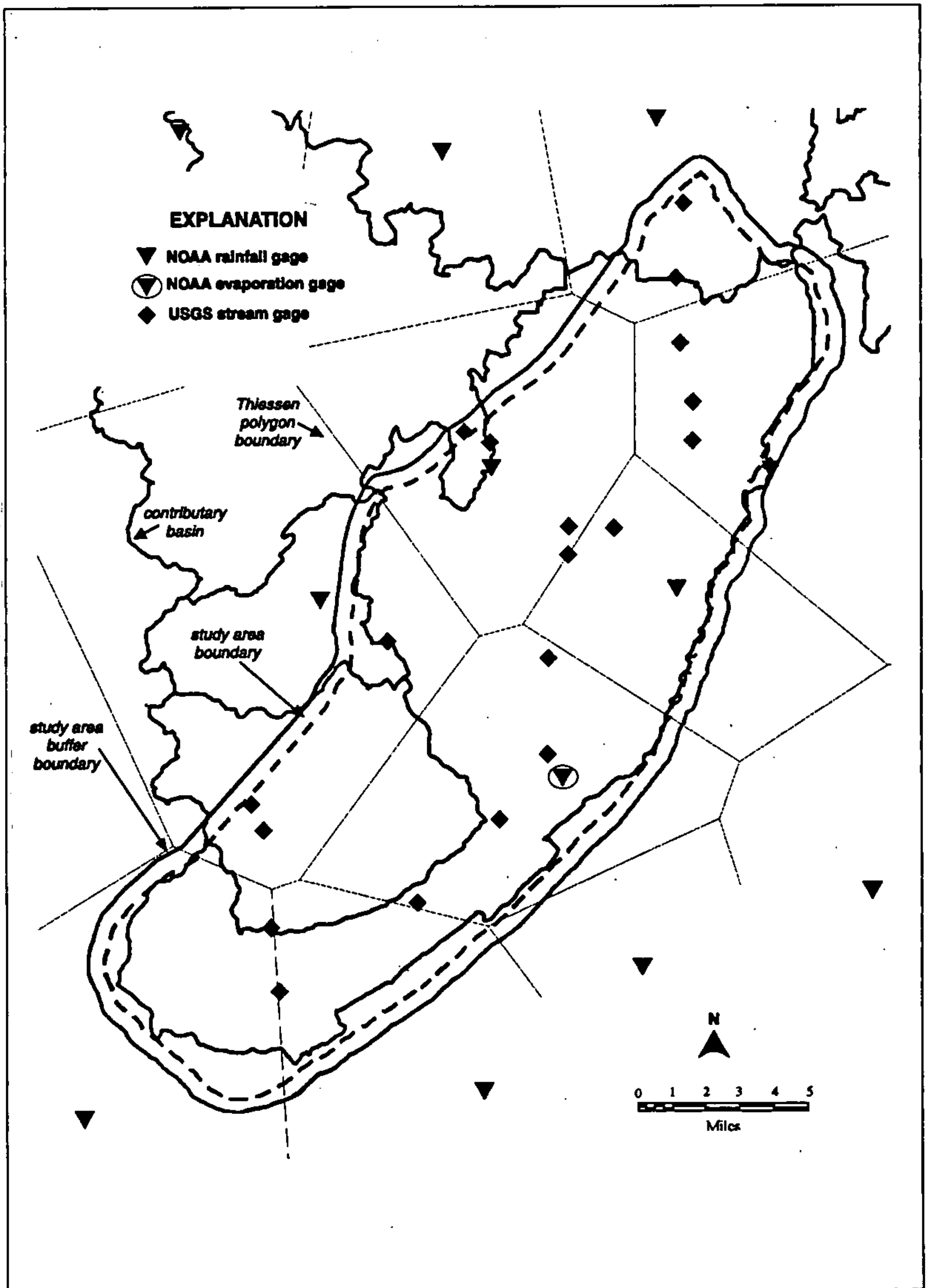


Figure 11. Distribution of rain, stream, and evaporation gages with rainfall, Thiessen polygons and stream flow contributary basins used in the Passaic study.

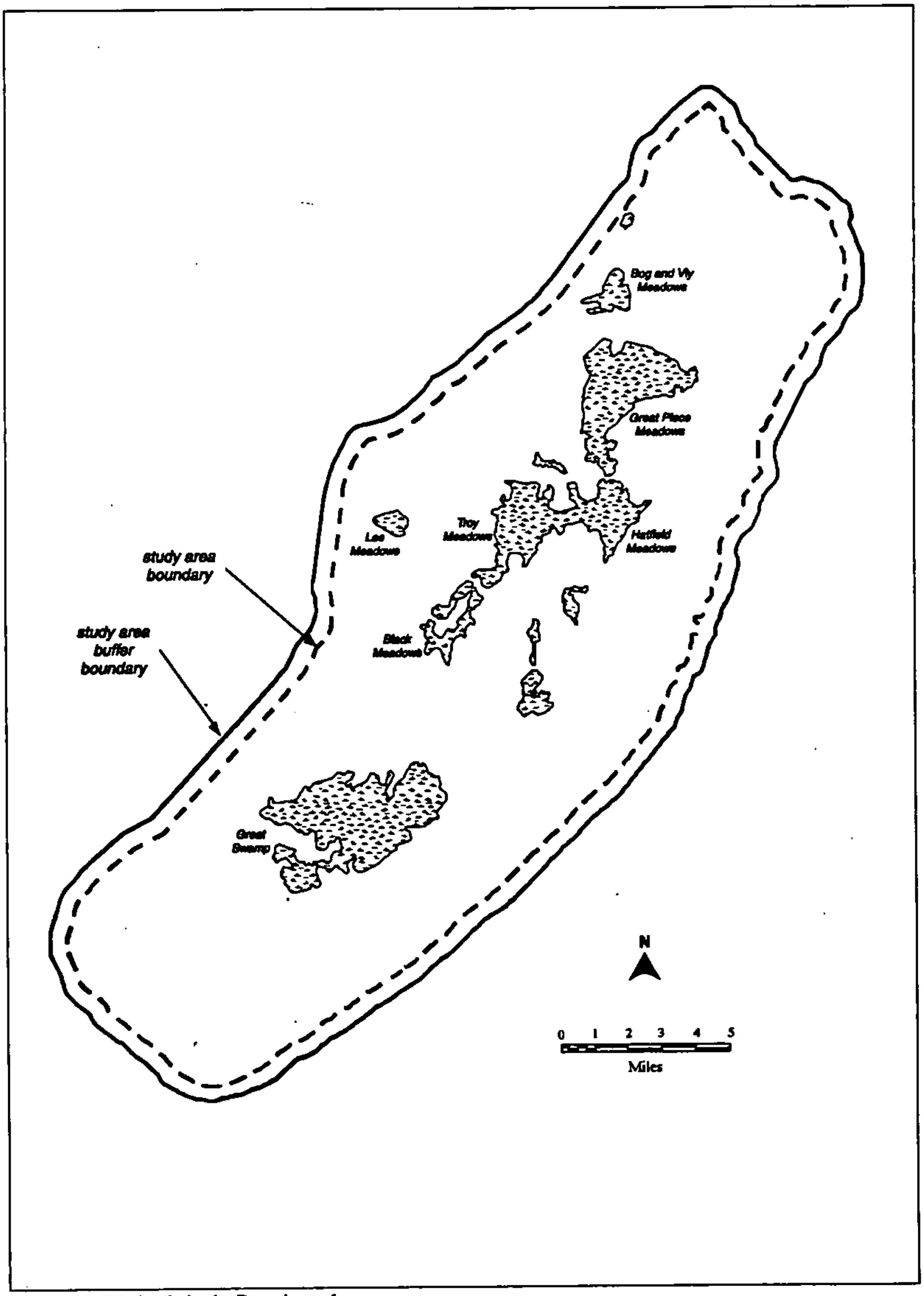


Figure 12. Wetlands in the Passaic study area.

allowing it to be registered to corresponding tics added to the BASEMAP coverage. RMS was not very good at 0.010. This coverage was used only for comparison with the recent wetlands coverage from the USGS (WETLANDS. USGS).

4. WETLANDS.USGS: Wetlands coverage developed from converted 1:100,000-scale DLG files (fig. 12). The arcs and polygons which represented the wetlands were reselected using the DLG code 111, which identified them.

G. Geology

1. GEOLOGY: Bedrock geology (fig. 13). Digitized in ADS from a 1:250,000-scale geologic map by Lyttle and Epstein (1987) and geologic data from the files of the NJGS. It was initially cleaned with a dangle length of 250, and a fuzzy tolerance of 10; subsequently a fuzzy tolerance of 0.001 was established in the TOL file. RMS was 0.001.

2. FAULT: Line coverage representing the main fault of the Ramapo/Border Fault system. Digitized from the geologic map of Lyttle and Epstein (1987) and cleaned with a dangle length of 100 and a fuzzy tolerance of 0.001. This fault is the western boundary of the study area.

H. Grid

1, 2. PASGRD & PASLAB: Two coverages generated by the ARC Macro Language (AML) program GENGRID.AML. This program produces two grid coverages, a polygon coverage (PASGRD) and a label-and-line coverage (PASLAB). It is set up to generate an unequally spaced grid from a digital data file. The grid specifications for the Passaic study were as follows:

1995799, 665440 324 (starting x,y location in state plane coordinates and orientation angle in degrees)

34 60 (number of rows and columns)

34*(times)2000 (row dimension in feet)

END

15*4000
4*3000
19*2000
22*3000 } column dimensions in feet

END

END

This produced a grid that was 34 columns by 60 rows with cells of 2,000 by 2, 000 ft., 2,000 by 3,000 ft., and 2,000 by 4,000 ft. (fig. 6). The GENGRID program and grid are described in more detail in section 3.1. These coverages were then "loaded" with data, and the result was downloaded as a matrix of values for use in the model.

I. Isolines

1. BEDROCON: Bedrock surface contours. Contours were prepared for eight USGS 7.5-minute quadrangles from geophysical data, well-log data, and document searches, then digitized in ADS as a group of separate quadrangles. These quadrangles were then combined using the append command and their edges matched using ARCEDIT. Contour values were checked and corrected. The contour arcs were also unsplit. This was done with care (only a few arcs at a time, hand selected in ARCEDIT) to see that the arcs retained correct contour values. The resultant coverage was still incomplete and was added to from various sources. Although this coverage includes many areas outside the study area, the final editing effort concentrated on filling in the blanks inside the area.

2. ROCKPILE: A special version of the final clipped bedrock contour coverage, BUFROCK (fig. 14) . It was combined, using the append command, with the study area buffer coverage, BUFSTAREA, and then cleaned and edited to create a polygon coverage. The labels in each of the polygons were updated by hand with a SHADE attribute so that each of the contour intervals could be uniquely shaded.

3. ROCKTIN3: TIN developed from the BUFROCK coverage using the arctin command. Developed in

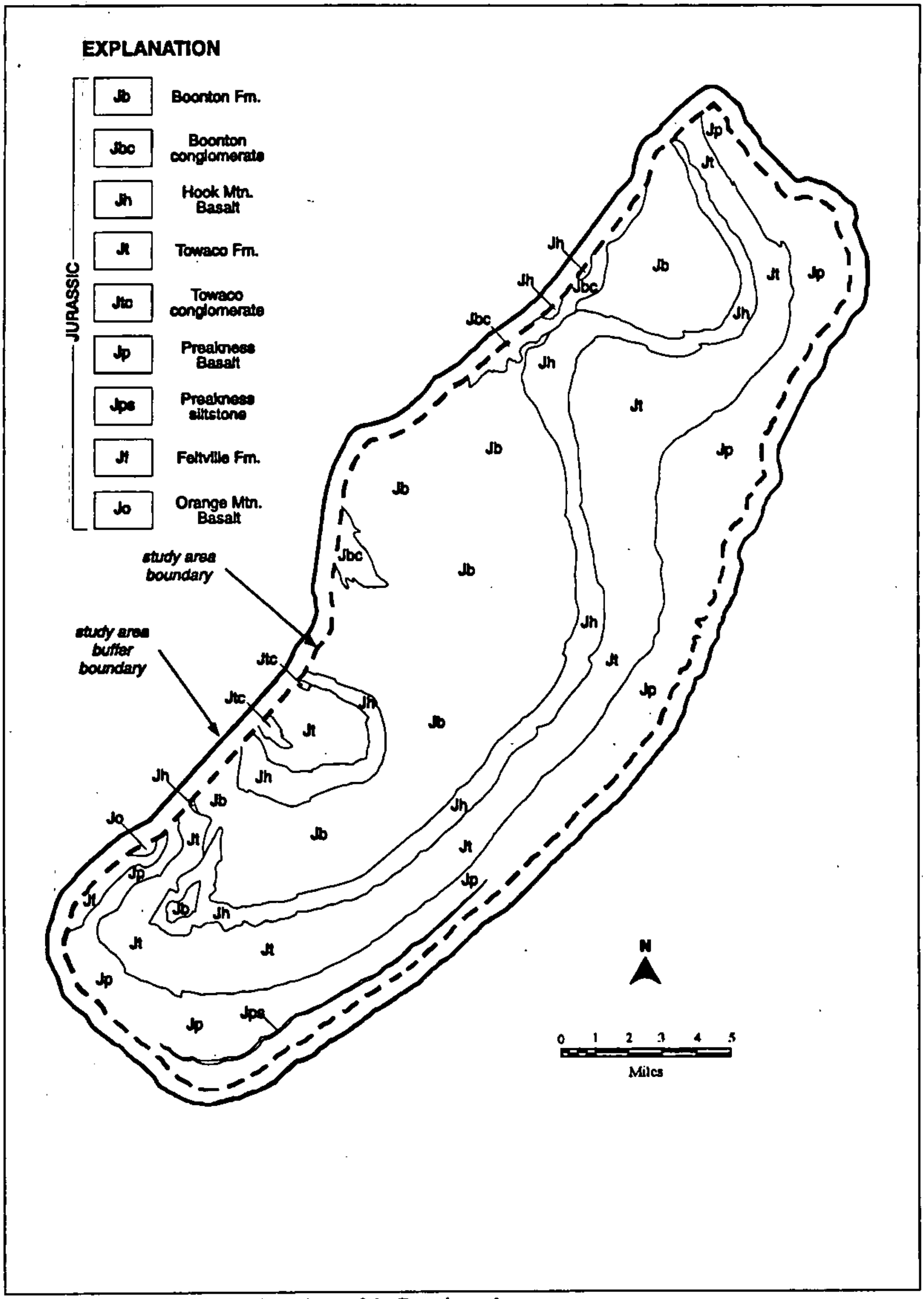


Figure 13. Generalized bedrock geology of the Passaic study area.

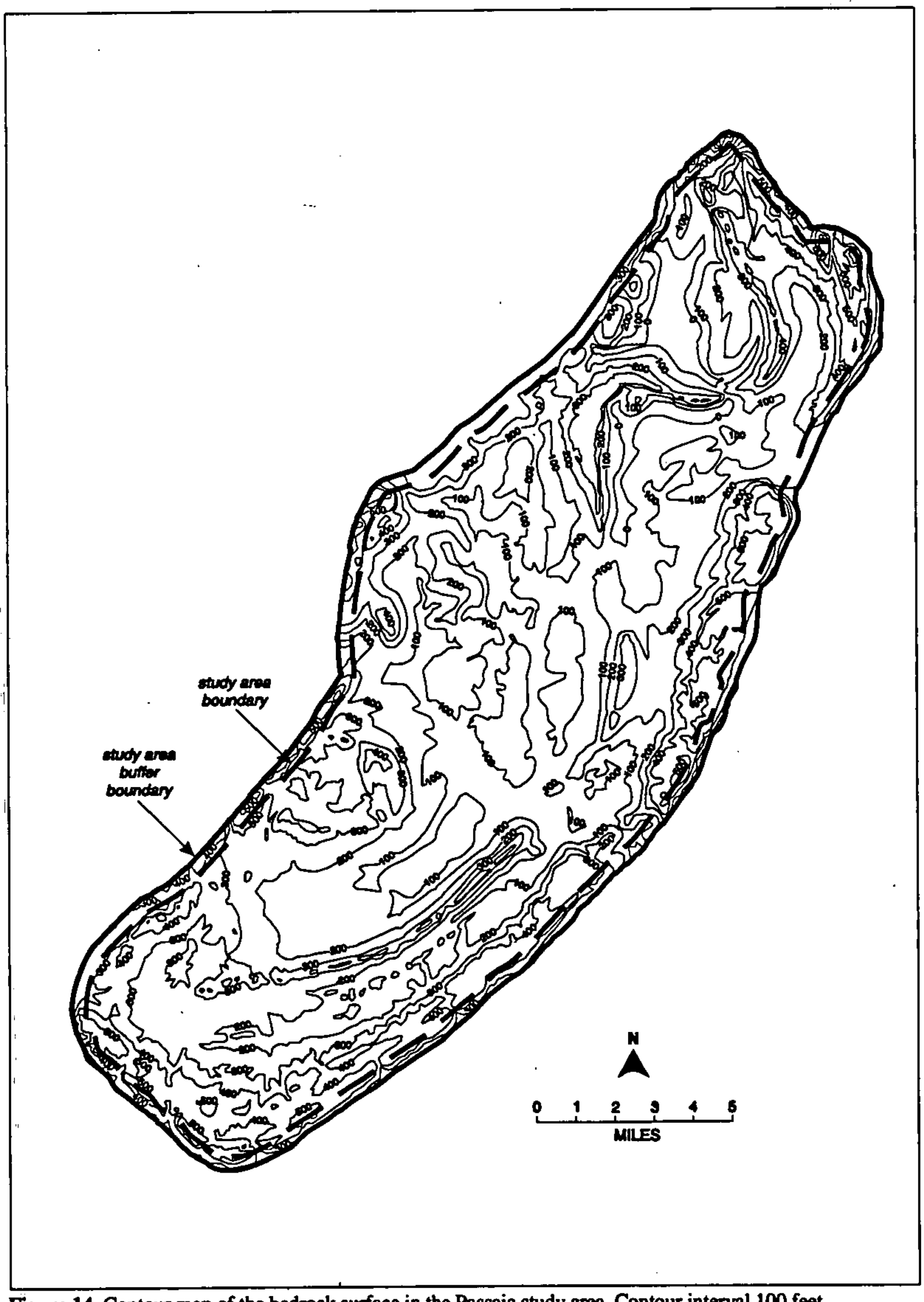


Figure 14. Contour map of the bedrock surface in the Passaic study area. Contour interval 100 feet.

order to calculate the thickness of the overburden in the study area.

J. Overburden:

These coverages were created as indicated in sections 2.3 and 4.1 from data at input scales of 1:24,000 (bedrock topography) or 1:250,000 (surface topography).

1. BURDCONLIN: 50-foot-interval isopach map of the BURDTIN overburden TIN using the tincontour command.

2. BURDTIN: Overburden TIN.

3. SEDCOVER: This is special version of the BURDCONLIN coverage which has been combined using the append command with the BUFSTAREA coverage so that a shaded contour map could be developed as shown in figure 15.

K. TILL (subdirectory of the Isolines subdirectory).

Coverages were to be combined into two coverages (elevation of top of till and elevation of bottom of till) for the whole study area, but an entire set of contour-line sheets was not completed in time for use in the model. Another method of determining thickness of the confining unit, of which the glacial till is a substantial component, was developed (see section 3. 4). These coverages were all digitized using ADS from the original mylar 1:24,000-quadrangle worksheets, except where otherwise noted.

1. BCALD: Elevation of the base of the till in part of the Caldwell quadrangle.

2. BCHAT: Elevation of the base of the till in part of the Chatham quadrangle.

3. BMORR: Elevation of the base of the till in most of the Morristown quadrangle.

4. GLACSEDS: Glacial geology of the study area (fig. 16). Digitized from a paper copy of a 1:100,000-scale map of the glacial geology of northern New Jersey prepared for Stanford, Stone, and Witte (1989).

5. TCHAT: Elevation of the top of the till in part of the Chatham quadrangle.

6. TMORR: Elevation of the top of the till for most of the Morristown quadrangle.

L. Roads

1. MAJRDS: Created by clipping the statewide ROADS coverage with a 0.001 fuzzy tolerance. Some roads and other features were reselected using the "CLASS" INFO attribute. The classes were 1 or A (interstate or equivalent highways), 2 or B (other major highways), and 8 (railroads). These arcs were then unsplit and annotated.

M. Streams

1. LAKES: Lakes and Reservoirs. Created by building the BUFSTRMJLH coverage using the polygon option, and then copying the coverage to LAKES. The lines representing the streams were deleted in ARCEDIT, leaving only the lakes. These were subsequently rebuilt as polygons.

2. PASSTRMS: Developed by combining two converted 1:100,000 DLG hydrography coverages using the append command. The streams that drain the Central Passaic River Basin were then clipped from the coverage using the BUFSTAREA coverage and the line option. The first coverage was cleaned with a dangle length of 250 and a fuzzy tolerance of 0.001, and subsequently cleaned as a polygon coverage.

3. WATER: Created by buffering along the streams in the BUFSTRMJLH coverage and combining the resultant coverage with the LAKES and BUFWTLNDS coverages. Interior lines were then dissolved to create a coverage which represented areas in the study area which were either "wet" or under water. The streams were buffered 12.5 feet on either side to create a 25-foot-wide buffer zone determined by the hydrogeologist to be the average width of streams and rivers in the BUFSTRMJLH coverage.

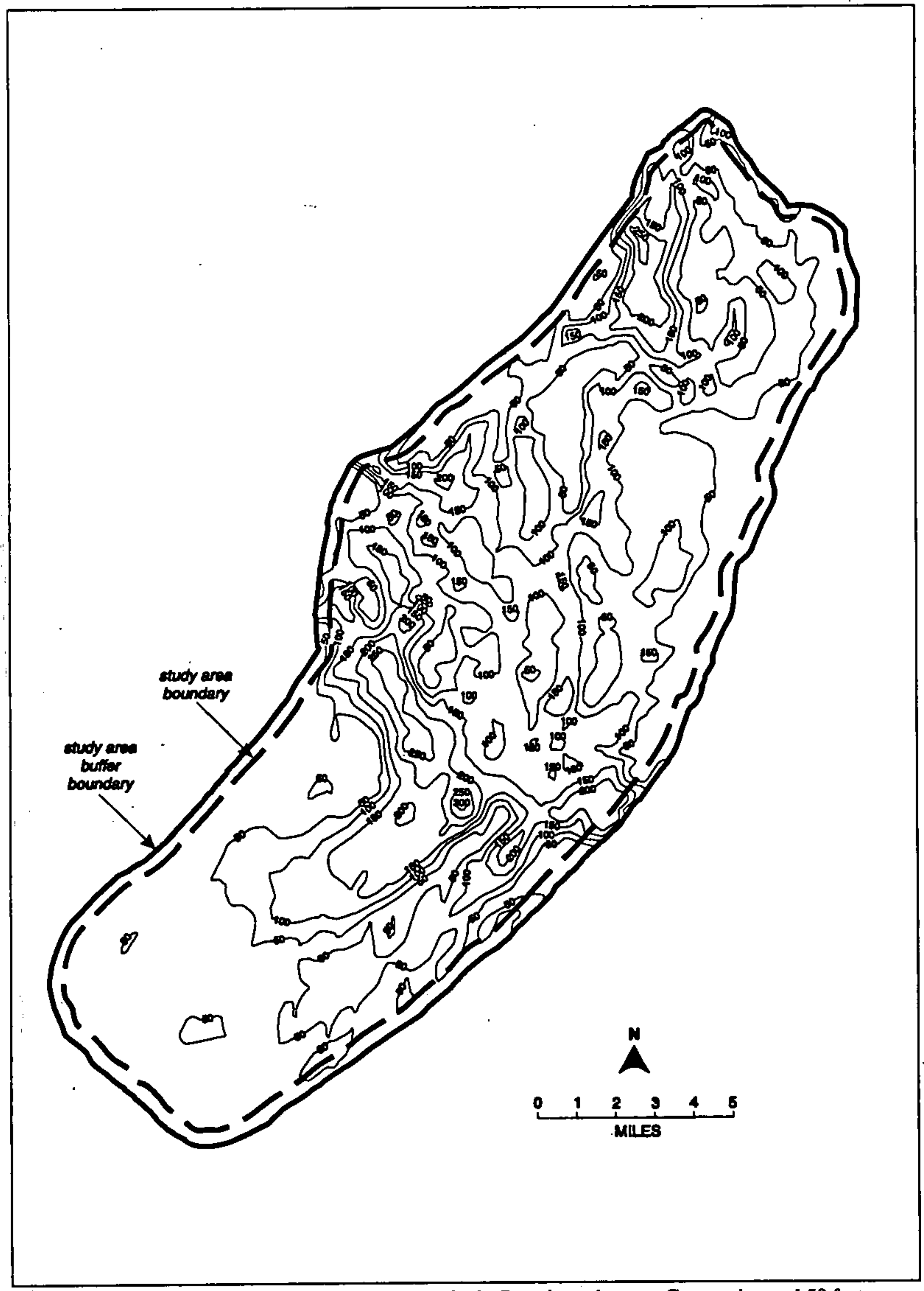


Figure 15. Contour map of the overburden thickness in the Passaic study area. Contour interval 50 feet.

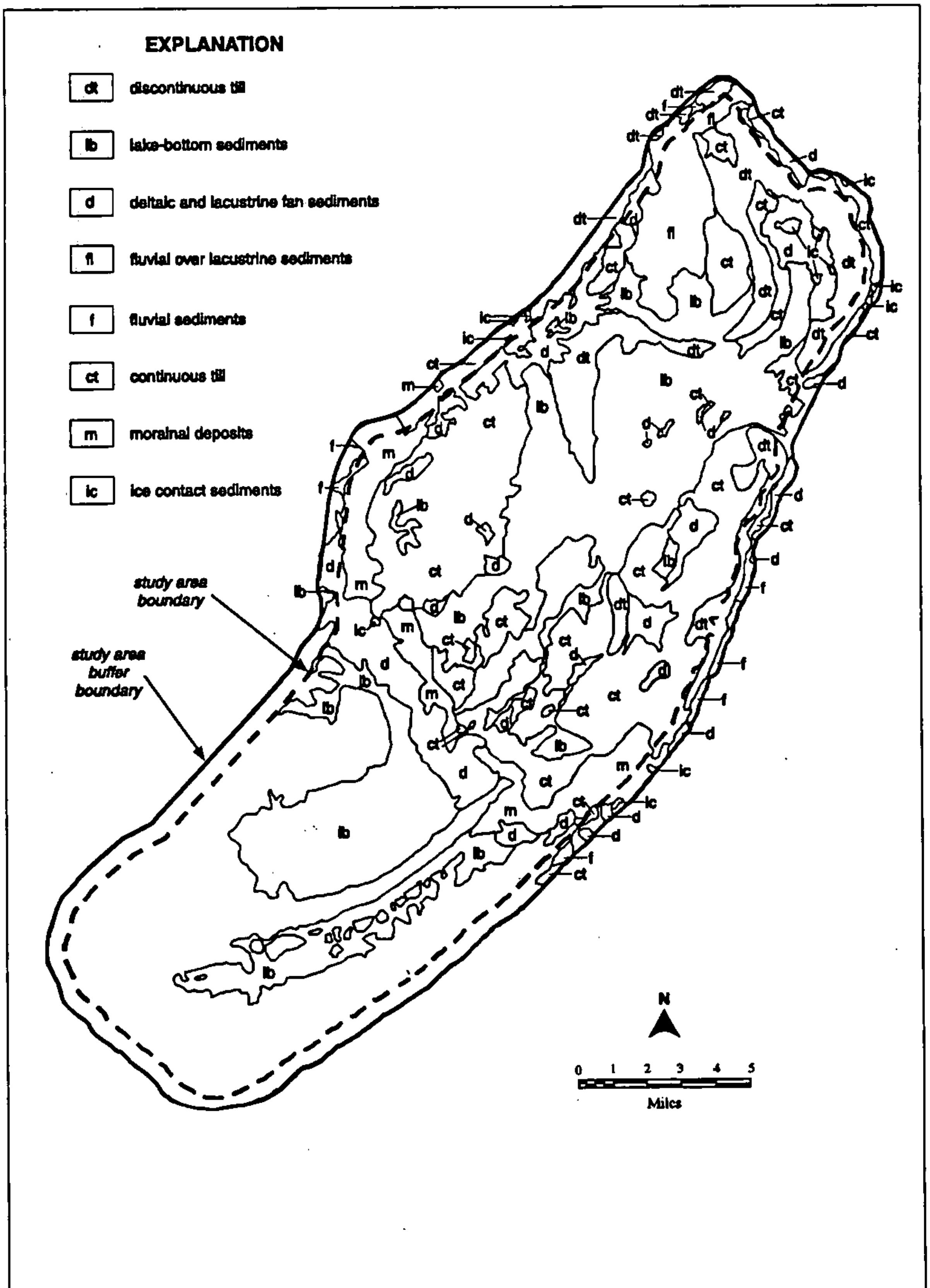


Figure 16. Glacial geology in the Passaic study area (after Stanford, Witte and Harper, 1990).

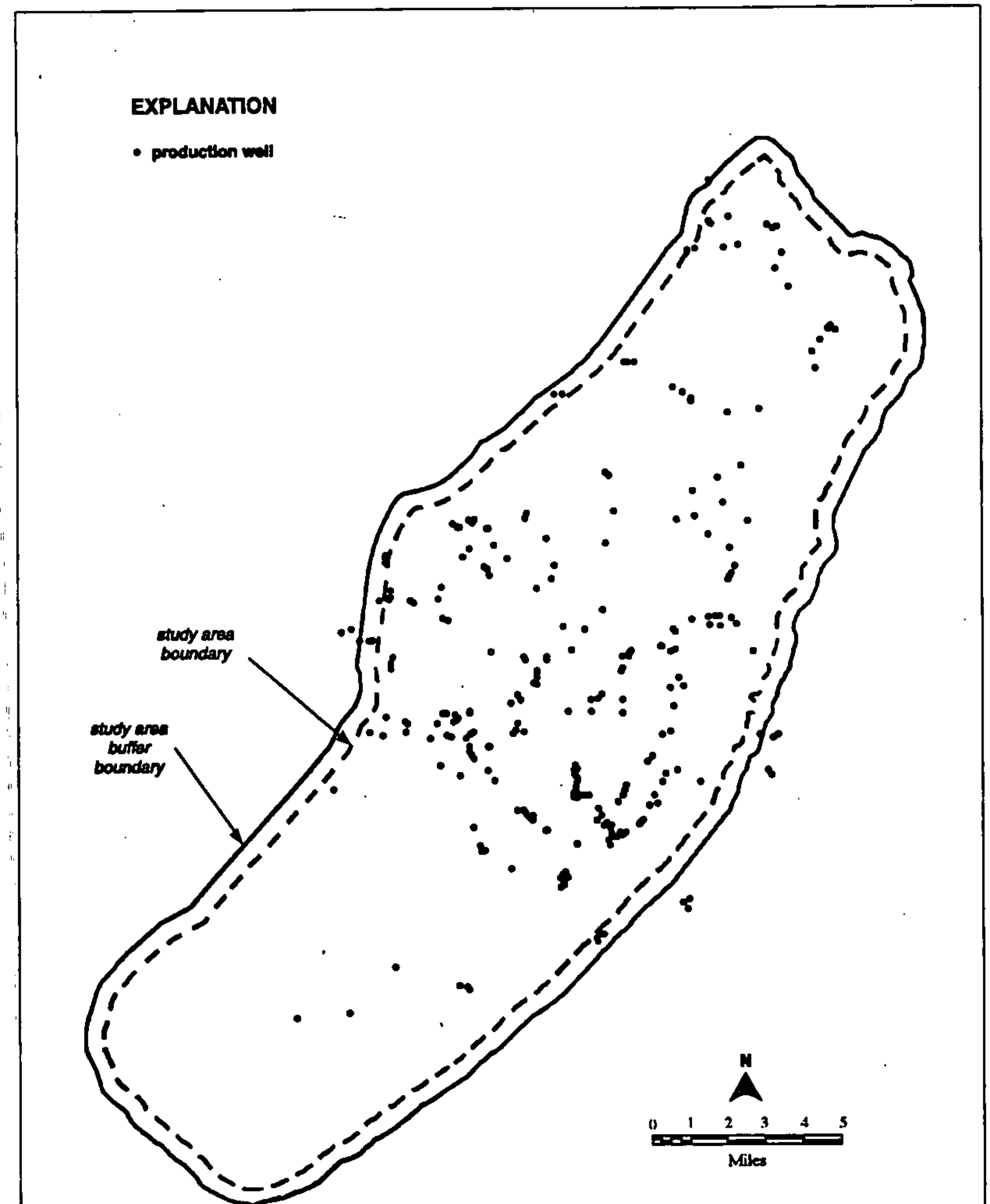


Figure 17. Distribution of production wells in the Passaic study area.

N. Wells

1. GWPWELLS: Ground-water production wells in and near the study area selected by the hydrogeologist for use in the Passaic Study (fig. 17). Well locations were digitized in ADS from 1:24,000-scale mylar maps using the BASEMAP coverage as the initial TIC/BND file. RMS ranged from 0.004 to 0.002.

2. GWQWELLS: Locations of ground-water-quality monitoring wells in the study area. Selected, located, plotted, and digitized in the same manner as production wells. RMS's, however, were not as good, ranging from a poor 0.010 to a good 0.002.

3. OBSWELLS: Ground-water observation wells in the study area. Coverage was created in same manner as for production wells and water-quality monitoring wells, and had the same RMS range as the GWQWELLS coverage.

3.4. Data loading of the model grid

Data are loaded into the grid cells by overlaying a data coverage with the grid or lattice coverage. A CPL called Isect.CPL " ... performs an intersection of the grid coverage with the [data] layer coverage and transfers the data to the [grid_name].PAT. In addition, the user can specify the method for assigning data to the grid. In the new version (8/4/89), the data are also moved to the corresponding lattice coverage representing the grid [cell] centroid (Ulery and others, 1989). This program was obtained from the USGS. Data were downloaded using another USGS program--MATRIX.CPL. This program gets input from the user [about] the ARC/INFO coverage name and the item name [to be output]. The [resultant] output file name is that item name with '. DAT' appended to it. The number of [output] rows has been set equal to 32. This CPL runs in ARC/INFO and creates an output file in matrix form which is readable by the [ground-water] model" (Ulery and others, 1989).

The numerical data for the model were maintained in and provided by INFO attribute tables. The most important attribute tables were the polygon and lattice PATs, PASLAB.PAT and PASGRD.PAT. Only a few other coverage attribute tables provided data to the model or preliminary data to the hydrogeologist.

The data in the attribute tables were produced in one of two ways. Some of the data items were created through the use of ARC/INFO commands; others were produced through direct manipulation of attributes and items in the PAT files using INFO commands. One example showing both types of data development demonstrates the capture of the surface and bedrock elevations and the subsequent development of the overburden-thickness and confining-layer elevations. This example also highlights one of the problems encountered in the development of the database.

The bedrock and surface elevations were produced by overlaying the PASLAB lattice with the tinspot command on ROCKTIN3 and TIN123. This command allowed the capture of the elevations at each lattice point. These values were then compared. Wherever the elevation of the bedrock from ROCKTIN3 was higher than the surface elevation from TIN123, the surface elevation was calculated to be equivalent to the bedrock elevation. These corrections were authorized by the hydrogeologist.

The next task was to calculate overburden thickness. At first this was to be done through the development of a related set of coverages and an overburden TIN. First a 392 row by 150 column lattice was created and tinspotted on the ground surface and bedrock TINs. The next step was subtracting the resultant bedrock elevation from the corrected surface elevation. The result was then rectified by zeroing those thicknesses that were less than 5 feet. A TIN was then produced which represented the overburden thickness. An isopach map was created using the TIN and the tincontour command set to a 20-foot interval. The PASLAB lattice coverage was then tinspotted onto this overburden TIN to capture the thicknesses. This methodology, however, introduced an unacceptable error which was discovered when confining-unit elevations and thicknesses were calculated. This problem is discussed in section 4.1.

It was therefore decided to use direct subtraction of the bedrock-surface and ground-surface elevations to calculate both the overburden thickness and the elevations and thicknesses of the confining units in the PAT. Along with these calculations, the discontinuous-till coverages were being analyzed. An elevation value was entered for each layer of the till in those cells into which the till coverages extended. All of these values--top elevations and base elevations of the till and confining units--were combined in the PAT to create a set of elevations at each grid label defining the layers for the model.

Once the data attributes were ready, they were downloaded individually through the MATRIX.CPL program, or alternatively as ASCII row and column files. Those INFO attributes that maintained downloadable data are listed and explained below:

1. ROCK.EL: Bedrock elevation at each grid label point as tinspotted from the ROCKTIN3 TIN using the PASLAB lattice coverage as the sampling lattice.

2. NEWSURF.EL: "Corrected" surface elevation for each lattice point which was first tinspotted from the ground surface TIN, TIN123. The PASLAB lattice coverage was used as the sampling lattice. The resultant tinspotted values, SURF.EL, were then compared with the ROCK.EL values. This comparison was done by a straight subtraction of the ROCK.EL bedrock-elevation values from the original surface-elevation values in SURF.EL. This comparison gave some bedrock elevations that were higher than surface elevation. Inasmuch as the bedrock elevation can only be less than or equal to the surface elevation, surface elevation for these areas was corrected to match bedrock elevation.

3. WET.AREA: The area of each grid cell covered by flowing water, freestanding water, or wetlands. To produce this coverage, the WATER coverage was combined with the PASGRD polygon grid coverage. The resultant intersection of the "wet" coverage and the grid cell polygons generated many new polygons. The total area associated with wet or water-covered poloygons was then reselected using the AREA attribute.

4. WET.PCT: The percentage of each grid cell covered by flowing water, freestanding water, or wetlands. This was calculated by dividing water and wetland areas (from WET.AREA) by the total area of each grid cell.

5. INOROUT: An attribute using -1, 0, and 1 codes to locate grid cells with respect to the study area. Any grid cell coded with a -1 was entirely outside of the study area. Any grid cell coded with a 1 was entirely within the study area. Any grid cell though which any part of the study area boundary passed was coded with a 0.

6, 7. NEWBCONTILL and NEWTCONTILL: The bottom and top elevations of the glacial till and the calculated bottom and top elevations of the various confining units. The till elevations were written into each grid cell on a draft map, and the values added to the BTILL and TTILL grid attributes using the update command in ARCEDIT. Because the till coverages were incomplete, the rest of the study area's confining units (of which the glacial till represented a significant component) and elevations had to be determined through a set of calculations developed by the hydrogeologist using the overburden thickness attribute OVCALC and the bedrock (ROCK.EL) and surface (NEWSURF.EL) elevation values. The grid cells were then encoded with a number (1, 2, or 3) in the CONUNIT attribute. These numbers identified which formula or calculation would be used to produce the confining-unit bottom and top elevations, NEWBCON and NEWTCON. Once these elevations were produced, they were combined with the correct glacial till elevation attributes. This was done by reselecting the grid cells where the till elevation was zero and copying the new confining-unit elevations into those locations using the INFO calculate command.

8, 9. Ground-water production and observation well locations: Row and column locations of the production and observation wells used in the study. These values came from the GWPWELLS and OBSWELLS coverages. They were captured using the identity command with the well coverages (GWPWELLS, GWQWELLS, and OBSWELLS) and the grid coverage (PASGRD).

10. Area values for use in water-budget calculations: Produced by overlaying GAGETHIES (the rainfall Thiessen polygon coverage), CON5BAS (the stream-gage contributing basin coverage for the four inflow stream gages and the single outflow gage) ALLCONBAS (the stream gage contributing basin coverage for all the stream gages), and the STAREA, (the study-area- boundary coverage).

Once downloaded as described above, this GIS-generated data, now represented by matrix and ASCII files, was ready to be used by the hydrogeologist.

4. CONCLUSIONS AND OBSERVATIONS

4.1. Problems

The foremost problem encountered was resolution of the data. This occurred because of the mixing and matching of the different data scales used to generate coverages. This problem was mitigated by the fact that the grid was relatively coarse, which led to a generalization of the data. The grid was necessarily coarse because of the size of the study area and the data-handling limitations of the model and the computer on which the model was run.

Even so, this scale mixing led to erroneous results when data coverages of grossly differing scale were used for calculations and subsequent analysis, as, for example, when the PASLAB coverage was tinspotted on the overburden TIN and overburden thicknesses were captured. Using this captured data and a set of calculations provided by the hydrogeologist, the elevations of the tops and bottoms of the various confining units were calculated. The data and subsequent manipulations were found to be invalid, as noted in Section 3.4. Many of the elevations and subsequently calculated thicknesses were erroneous. In some locations, tops of confining units were calculated to be above the ground surface or below the bedrock surface; at other locations, a unit's calculated thickness exceeded that of the entire overburden. This problem was probably created by the differences in scale of the two contour-line coverages. Whereas the land surface contour coverage, CON123, was created from 1:250,000 scale DEM files, the bedrock-contour coverage, BUFROCK, was digitized from 1:24,000 scale sources. This scale difference was further accentuated by manipulation of both the surface elevations and the calculated overburden thicknesses, and was probably affected by the sampling of the surface and bedrock TINs with a very high-density lattice in order to produce the overburden TIN.

Although this is perhaps the only example of this problem, other, more subtle inaccuracies may exist. This leads to some uncertainty about the overall accuracy of the data produced both for display and for the model. However, it must be pointed out that the accuracy requirements for the data were relatively low due to both the coarse overall resolution of the grid and the limitations of the model's and the microprocessor's ability to handle the amounts of data required. Therefore, some of the data provided for the model were developed at smaller scales and subsequently enlarged to 1:24,000. This was necessary because there were no other data available which would meet the needs of the model within the required time frame.

The size of the database contributed to some minor problems. Some of these problems were caused by imperfect coordination between participants in the development process, due mostly to a lack of communication about changes to the database, changes to the requirements, and progress toward the mutual goal. Other problems were caused by a relaxed attitude toward documentation of the development of the data. Data and data-coverage development were not thoroughly integrated with ongoing paper or on-line systems of documentation defining the coverages or data. The surprising size of the database also created computer storage-space problems.

4.2. Suggested improvements and solutions

The solution to the data-resolution problem is to maintain, as much as possible, rigorous scale resolution within the various data categories and create data which are of a resolution compatible with that of the modeling process itself. This should maintain the integrity of the data, help insure that the best possible data are provided, and produce the most accurate results. This solution is often precluded by time and fiscal constraints. Databases may have to be developed using readily available data of a compatible resolution, regardless of scale.

4.3. Final observations

In conclusion, there is room for improvement as noted above. Data gathering, rather than data entry, proved to be the bottleneck for the development of the database. Several of the necessary data sets were either not available as GIS products, not available in the correct scale as GIS products, or had not been completed in any format, digital or analog. Most of the background or regional data were already available and most of the specific data for this study were provided in a format easily convertible to a GIS format. Future data collection strategies could be more appropriately designed to meet the requirements of the GIS as well as other needs. Current and contiuing database development may alleviate this data gap for future projects.

Data entry for this study went smoothly and continued apace with only minor inconveniences created by the storage-space crunch and by developments in the output stage. The database was, for the most part, prepared and ready within the original time frame. However, the development of purchased coverages, and some coverages produced by data from sources outside the NJGS was delayed.

As more projects of this types are done, processes will become fine-tuned. This integration of a ground-water model with the NJDEP GIS is a very important first step in the utilization of a powerful new tool for the management of ground-water supplies.

References

ESRI, 1991, ARC/INFO user's guide, ARC/INFO data model, concepts and key terms: Environmental Systems Research Institute, Inc., Redlands, CA, p. 1-1 - 1-24.

Hoffman, Jeffrey L., 1989, Plan of study for the Central Passaic River Basin hydrogeologic investigation: New Jersey Geological Survey Report OFR 88-4, 22 p.

Lyttle, P.T., and Epstein, J.B., 1987, Geologic map of the Newark 1° x 2° quadrangle, New Jersey, Pennsylvania and New York: U.S. Geological Survey Miscellaneous Investigations Series I-1715, scale 1:250,000.

MacDonald, M.G., and Harbaough, A.W., 1988, A modular, three-dimensional, finite difference ground-water model: U.S. Geological Survey, Techniques of Water-Resources Investigations, TWI 06-A1, 576 pages (Supersedes Open-file Report 83-875).

Salisbury, R.D., and Peet, C.E., 1894, A preliminary map of the surface formations of the valley of the Passaic with the country east to Newark and southward to the Raritan: *in* New Jersey Geological Survey Annual Report of the State Geologist, 1894, scale 1:63,360.

Stanford, S.D., Witte, R.W., Harper, D.P., 1990, Hydrogeologic Character and Thickness of the Glacial Sediment of New Jersey: New Jersey Geological Survey, scale 1:100,000.

Todd, D.K., 1980, Ground-water hydrology, 2nd edition: J. Wiley & Sons, New York, 535 p.

Ulery, Randy, Chepiga, M.M., Price, Curtis, Schooley, Jack, and Fiorentino, Mark, (unpub.), GENGRID.AML: U.S. Geological Survey and New Jersey Geological Survey Arc Macro Language program.

-----, ISECT.CPL: U.S. Geological Survey and New Jersey Geological Survey, Command Programming Language program.

-----, MATRIX.CPL: U.S. Geological Survey and New Jersey Geological Survey, Command Programming Language program.

GEOGRAPHIC INFORMATION SYSTEM (GIS) DATABASE DEVELOPMENT FOR THE CENTRAL PASSAIC RIVER BASIN HYDROGEOLOGIC INVESTIGATION (New Jersey Geological Survey Open-File Report OFR 95-2)